DETERMINING THE OPTIMAL METHOD OF TRIAGE FOR THE UK MAJOR INCIDENT SETTING

By

NABEELA SHABBIR MALIK

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Inflammation and Ageing

School of Medical and Dental Sciences

University of Birmingham

November 2023

ABSTRACT

Major incidents are characterised by a mismatch between the immediate needs of patients and the resources available to treat them. In the UK, man-made major incidents comprising blast and penetrating trauma have occurred with increasing frequency in recent years and continue to constitute the greatest threat to national security (Chapter 1). Triage tools have a crucial role in ensuring that finite healthcare resources are allocated to maximise overall survival. Existing tools have largely been developed based on expert opinion and implemented without formal validation. A major challenge in furthering the science of major incident management is the lack of consensus and standardisation of end-points which best define Priority 1 (P1) status and other triage categories.

In this thesis, the literature surrounding major incident triage was reviewed and modern, evidence-based definitions of triage categories were validated amongst adults and children from the UK national trauma registry (Chapters 1 and 2). This system developed by Lerner et al defined P1 patients as those requiring any one of eight time-critical life-saving surgical and major resuscitative interventions, directly reflecting patient resource requirements, and correlates well with casualty distribution within UK regionalised systems of trauma care. Hence it is recommended that Lerner's criteria should become the gold standard in reporting triage tool performance in future UK major incidents.

These intervention-based triage category definitions were compared with the triage category assignments of ten existing international triage tools. This comparative analysis demonstrated that the current UK national triage tool, the NARU Triage Sieve, is poorly sensitive in identifying P1 patients who require time-critical, life-saving interventions, and was surpassed in sensitivity by several tools, the best of which was the UK military's BCD Triage Sieve (Chapter 2, publication number 1). Similarly, in children from the national trauma registry, the current UK tools, the Paediatric Triage Sieve and JumpSTART, demonstrated suboptimal performance and were

substantially outperformed by several tools, with the BCD Triage Sieve again demonstrating the greatest sensitivity (Chapter 2, publication number 2). This has directly informed national policy, resulting in the removal of previous existing tools and a recommendation for the NHS Major Incident Triage Tool (MITT), the civilian adaptation of the BCD Triage Sieve, to become a single tool for NHS first responders to apply to all adults and children in future UK major incidents.

Subsequently, using tree-based machine learning methodology, a novel primary triage tool, Quick Triage, as well as a secondary triage tool that is applicable via a portable device, have been developed (Chapter 3). Both tools outperform the BCD Triage Sieve amongst TARN patients, with reduced associated over-triage rates. The app-based secondary triage tool has withstood external validation amongst injured patients from the UK military's Joint Theatre Trauma Registry (JTTR); however, a paucity of pre-hospital data precluded external validation of Quick Triage amongst JTTR patients, necessitating further work. Quick Triage has balanced, favourable performance amongst adults, children and elders with blunt and penetrating trauma from the UK national trauma registry, with lower over-triage rates than the NHS MITT (Chapter 4). This offers the possibility of a single tool for use amongst patients of all ages by all NHS and non-NHS first responders, allowing complete interchangeability, simplification of the prehospital triage process and likely more rapid and accurate triage of major incident casualties in future UK and international major incidents.

DEDICATION

This thesis is dedicated to my mother, who recognised in me a thirst for knowledge from an early age, and created every opportunity to quench this thirst. Your voice often echoes in my head, Mum, "Go Nabeela, go!"

And to my daughter Sofia, who arrived in the midst of this work and in the heart of the pandemic, I hope the world we leave you in is a safer place than the one we entered.

ACKNOWLEDGEMENTS

I would like to thank wholeheartedly all those who made this work possible. Firstly, I would like to extend my gratitude to my supervisors. Prof Georgios Gkoutos, thank you for allowing me the freedom to shape and lead this work, your calm and patient approach has allowed for much personal growth. Thank you to Prof Janet Lord and Prof Tony Belli for your guidance and leadership. And to my adopted supervisor, Colonel Douglas Bowley, who has helped me navigate hurdles of an assorted nature during this period, I have certainly "grown up" a little as a result!

Thank you to the NIHR Surgical Reconstruction and Microbiological Centre (SRMRC) for funding my work and for welcoming me to the team, in particular Dr Karen Piper, who has a solution for everything.

For your support with computer coding, statistical advice and application of machine learning to my database, I'd like to thank Dr Saisakul Chernbumroong, Dr Jon Bishop and Dr Yuanwei Xu; our synergy has resulted in some impactful publications and policy changes that will certainly improve the lives of patients.

Thank you to my wonderful family; Mum, Abu and Zishaan, your support through this and my postgraduate career have been the foundation on which I stand, I wouldn't be who I am without you.

And lastly, with all my heart, thank you to my husband Hayat Khan. Your gentle love and patience through the ups and downs of this work have brought it to the finish line.

PUBLICATIONS AND PRESENTATIONS

During my registered period of doctoral research at the University of Birmingham, the following articles and conference abstracts were accepted for publication and/or presentation at conferences.

PUBLICATIONS

Book chapter

Co-editor/author of the chapter "Bladed, blast and ballistic injuries"
ABC of Major Trauma: Rescue, Resuscitation with Imaging, and Rehabilitation, 5th Edition edited by Peter Driscoll, David Skinner and Peter Goode. BMJ Books. Publication Nov 2022

Papers

Y Xu**, NS Malik (joint first author)**, S Chernbumroong, J Vassallo, D Keene, M Foster, JM Lord, A Belli, T Hodgetts, DM Bowley, GV Gkoutos. Triage in major incidents: development and external validation of novel machine-learning derived primary and secondary triage tools. Accepted by Emergency Medicine Journal, 12 August 2023.

Malik NS, Chernbumroong S, Xu Y, Vassallo J, Lee J, Bowley DM, Hodgetts T, Moran CG, Lord JM, Belli A, Keene D, Foster M, Gkoutos GV. Paediatric major incident triage: UK military tool offers best performance in predicting the need for time-critical major surgical and resuscitative intervention. Lancet EClinicalMedicine 2021;40 doi: 10.1016/j.eclinm.2021.101100

Malik NS, Chernbumroong S, Xu Y, Vassallo J, Lee J, Bowley DM, Hodgetts T, Moran CG, Newton T, Arul GS, Lord JM, Belli A, Keene D, Foster M, Gkoutos GV. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. Lancet eClinicalMedicine 2021;36 doi: 10.1016/j.eclinm.2021.100888

Malik NS, Munoz B, de Courcey C, Imran R, Lee KC, Chernbumroong S, Bishop J, Lord JM, Gkoutos G, Bowley DM, Foster MA. Violence-related knife injuries in a UK city; epidemiology and impact on secondary care resources. Lancet EClinicalMedicine 2020;20 doi: 10.1016/j.eclinm.2020.100296

Vassallo J, Chernbumroong S, **Malik N**, Xu Y, Keene D, Gkoutos G, Lyttle MD, Smith J. in collaboration with PERUKI (Paediatric Emergency Research in the UK and Ireland). Comparative analysis of major incident triage tools in children: a UK population-based analysis. Emergency Medicine Journal Published Online First: 27 October 2021. doi: 10.1136/emermed-2021-211706

PRESENTATIONS

Invited speaker

1. "Knife injuries in the West Midlands"
   Trauma Care Conference, Staffordshire, 18 November 2022

2. "Major Incident Triage Tools for the NHS"

Trauma Care Conference, Staffordshire, 18 November 2022

3. "Research in Major Incident Triage"
   University of Birmingham/West Midlands Ambulance Service Research Symposium, Birmingham, July 2022

4. "Knife injuries: Who, where, how and when"
   British Orthopaedic Association Annual Congress, Liverpool, 12 Sept 2019

5. "Epidemiology of knife injuries and their impact on secondary care resources"
   Royal College of Surgeons: Surgeons at the sharp end - Knife crime in the West Midlands, Birmingham, 9 Sept 2019

Conference presentations (Oral)

1. A retrospective, single centre, observational study, to compare the performance of different mass casualty triage tools in predicting Priority 1 status
   Shimri L, Xu Y, Lee J, Z Ahmed, **Malik NS**
   Accepted for presentation at the British Trauma Society Annual Scientific Meeting, 21-23 Nov, Cardiff, Wales.

2. Death on the urban battlefield: Preliminary analysis of 101 penetrating trauma deaths across the West Midlands (2012-2019)
   Malik NS, Bowley DM, Porter K, Battaloglu E
   Wiseman Prize session, Military Surgery Symposium at the London Trauma Conference, London, 7 Dec 2022

3. Machine learning development of novel triage tools for major incidents
   Malik NS, Xu Y, Chernbumroong S, Bowley DM, Hodgetts T, Gkoutos GV
   Association of Surgeons of Great Britain and Ireland (ASGBI) 2022 International Surgical Congress, Liverpool, 4 May 2022

4. Triage in terror-related major incidents
   Malik NS, Xu Y, Chernbumroong S, Bowley DM, Hodgetts T, Gkoutos GV
   Association of Surgeons of Great Britain and Ireland (ASGBI) 2022 International Surgical Congress, Liverpool, 4 May 2022

5. Characteristics and outcomes of children attending hospital with knife injuries: A UK paediatric major trauma centre experience
   Malik NS, Watson L, Bishop J, Gkoutos G, Belli T, Bowley D, Newton T, Arul S
   British Trauma Society Annual Meeting (virtual), 25 Nov 2021

6. Comparative Analysis of Major Incident Triage Tools in Children – a UK population-based analysis
   Vassallo J, Chernbumroong S, **Malik NS**, Keene D, Xu Y, Gkoutos GV et al.
   RCEM Scientific Conference, 6 Oct 2021

7. Comparative Analysis of Major Incident Triage Tools in Children – a UK population-based analysis
   Vassallo J, Chernbumroong S, **Malik NS**, Keene D, Xu Y, Gkoutos GV et al.
   EUSEM International Conference, 28 Oct 2021

8. The impact of knife crime on the blood bank: A retrospective review in an urban UK Major Trauma Centre
   Malik NS, Munoz B, de Courcey C, Gkoutos G, Foster MA, Bowley DM, Doughty H
   Accepted to European Congress of Trauma & Emergency Surgery, 26-28 April 2020, Oslo, Norway, however subsequently cancelled due to COVID-19

Conference presentations (Poster)

1. Major incident triage: Machine learning development and external validation of models predicting the need for life-saving intervention amongst injured patients from the UK national trauma registry
   Malik NS, , Xu Y, Chernbumroong S, Vassallo J, Bowley DM, Hodgetts T, Lord JM, Belli A, Keene D, Foster M, Gkoutos GV
   European Congress of Trauma & Emergency Surgery, Oslo, 24-26 April 2022

2. Characteristics and outcomes of children attending hospital with knife injuries: A UK paediatric major trauma centre experience
   Malik NS, Watson L, Bishop J, Gkoutos G, Belli T, Bowley D, Newton T, Arul S
   European Congress of Trauma & Emergency Surgery, Oslo, 24-26 April 2022

TABLE OF CONTENTS

CHAPTER ONE: GENERAL INTRODUCTION

What is a major incident?

A major incident can be defined as "an event or situation with a range of serious consequences which requires special arrangements to be implemented by one or more emergency responder agency[1]." The characteristic feature of a major incident is that resources are outstripped by the immediate needs of the community[2]. The challenge to responders lies in how best to allocate finite resources. From a healthcare perspective, the National Health Service (NHS) defines a major incident as any emergency that requires the implementation of special arrangements by one or more of the Emergency Services, the NHS or local Authority for the initial treatment, rescue and transport of a large number of casualties[3]. Whilst the term "major incident" is commonly used in the UK, from a healthcare perspective, it is synonymous with the term "mass casualty," which describes any situation in which emergency medical services resources, such as personnel and equipment, are overwhelmed by the number and/or severity of casualties[2,4.]

The management of casualties is one important aspect of major incident management[5]. According to the Major incident Medical Management and Support (MIMMS) paradigm, which has guided major incident practices across the UK and much of the Commonwealth for the last three decades, the management of a major incident involves several steps including detection of a major incident, incident command, ensuring scene security and safety, assessing hazards, support, triage and treatment of casualties, evacuation and recovery from the major incident[5]. Prior preparation, recognition of risks at a community level and the coordination and interoperability of multiple agencies are crucial to the successful management of a major incident[1,2,5].

Triage is one aspect of major incident management. The term triage refers to the assessment of casualties and the assignment of priority on the basis of clinical need. This assignment of priority facilitates onward transfer and treatment of casualties in a manner that will provide the "greatest good to the greatest number[5]." This differs from the treatment of casualties outside of a major incident, in which resources are not constrained and the best possible treatment and timely transfer can be offered to individual patients to maximise their chances of a favourable outcome following injury. There are no simple means by which critically ill or injured patients can be identified at the scene of a disaster and this is where triage can make its greatest impact when faced with a large the number of victims[6]. In fact, the outcome of victims assigned the highest triage category is considered to be a good indicator of the effectiveness of medical care following a major incident[7].

First responders are required to perform triage under challenging circumstances, including harsh environmental conditions and often, an ongoing threat, for example the risk of building collapse following fires or explosions, or the possibility of multiple or co-ordinated terrorist attacks[8,9]. Practitioners need to retain situational awareness to ensure their own safety as well as the safety of patients[8,10]. As such, practical factors such as ease and simplicity, time taken to apply the tool and the degree to which the results are reproducible are important factors to consider in the design of triage tools[11].

The history of triage and contemporary major incident triage categories

The concept of major incident triage, as with many aspects of trauma care, has largely evolved through wars[12]. Historically, when multiple individuals were injured in a Western military setting, casualties were treated in the order of military rank and the order in which they arrived at a treatment facility, irrespective of the nature or severity of injuries. During the Napoleonic Wars in the late 18th century, Baron Dominique Jean Larrey (1776-1852) served as French surgeon in Napoleon's Grande Armée[12]. Larrey made some important observations which gave rise to the concept of triage as we know it now. Larrey observed that military casualties could be divided into three broad categories, as summarised in the table below[12].

Table 1: Larrey's classification of combat casualties

Larrey's division of combat casualties into three categories (Napoleonic wars, 18th century)
1. Likely to survive, regardless of care
2. Likely to die, regardless of care
3. Immediate care would make a positive difference

On this basis, he recommended that casualties should no longer be treated in the order of rank but instead, by clinical need. Larrey ordered the adaptation of horse-drawn artillery carriages into ambulance volantes or "flying ambulances" for the immediate transfer of those patients in whom immediate care was likely to make a positive difference. These "flying ambulances" were the predecessor of the modern field ambulance. Prior to transfer, casualties were sorted to identify those in need of immediate care – hence the term "triage" was applied, from the French word "trier" (to sort). Larrey is also famous for advocating early amputation following battlefield limb injuries, in order to prevent death from bleeding or sepsis, which occurred hours or weeks later[12]. He also pressed for hospitals to be located as close to the battlefield as possible, indeed in contemporary military practice this has translated into Field Hospitals with the delivery of Damage Control Resuscitation and Surgery prior to rearward evacuation for definitive care[13,14].

In the contemporary major incident setting, Larrey's three triage categories have been adapted over time into five categories, to ensure that constrained medical resources are directed at achieving the greatest good for the most number of people[5]. There are differences in the terminology used across different countries and organisation, however approximate equivalence is noted in line with Larrey's observations, as summarised in the table below[11]:

Table 2: Contemporary classification of casualties in a major incidents

Terminology for each triage category (approximate equivalence)	Description
Priority 1, Immediate/red, T1	Immediate transfer, immediate medical care
Priority 2, Delayed/yellow, T2	Urgent or delayed transfer, urgent medical care
Priority 3, Minor/green, T3	Non-urgent, "walking wounded"
Priority 4/Expectant (blue) T-hold	Casualty who is expected to die
Dead, (white/black)	Casualty with no signs of life

Ledger: Common UK, US-based (including coloured tags) and NATO terminology is shown, respectively. The Australian CareFlight system assigns Minor, Immediate, Delayed and Unsalvageable categories.

In the context of a major incident, health care providers cannot deliver the highest standard of care to every injured individual and must instead use available resources to deliver the greatest good to the greatest number[2]. As such, casualties are assigned triage categories which reflect the urgency with which they should be transferred to a medical treatment facility and receive treatment[1]. According to the Major Incident Medical Management and Support (MIMMS) employed in the UK civil and military environment, a Priority 1 (P1) casualty is one who requires "immediate life saving treatment[5]." Differentiation between P1 and P2 casualties remains the greatest challenge of major incident triage[10]. The P3 or Minor category is often assigned to patients who are ambulatory (the so-called "walking wounded" category[11], however, evidence suggests that as many as one in five ambulatory patients in the major incident setting have significant underlying injuries[15]. Assignment of the Priority 4 (P4), also known as Expectant, category is the most contentious and ethically challenging category for healthcare practitioners[1,5]. This denotes a casualty who cannot survive despite treatment or for whom the degree of intervention required is such that in the current circumstances, their treatment would seriously compromise the provision of treatment for others. Several triage systems use the assignment "Dead" for those casualties who are not breathing despite manoeuvres to open their airway[11].

In measuring the accuracy of triage systems, studies may look at the overall accuracy of triage category assignments[16,17]. However, more commonly, studies have focussed on measuring the ability of triage tools to detect the Priority 1 category, as it is these patients who are at greatest risk of adverse outcome if assigned an incorrect triage category[18-21]. It is widely accepted that correct identification of Priority 1 casualties is the most critical function of triage tools, in order to maximise overall survival[9]. Incorrect triage may fail to identify patients in need of urgent intervention (under-triage): in the case of critically ill individuals, there is an absolute harm associated with delays in care or onward transfer to an inappropriate facility for definitive care[2,9,22,23]. However, its inverse (over-triage) risks overwhelming healthcare facilities with patients who do not require time-critical treatment. Frykberg demonstrated a direct link between overtriage and mortality[24]. Currently, no national or international standards have been defined to guide or govern the performance of triage tools when applied in the major incident setting. By comparison, the American College of Surgeons' Committee on Trauma sets a standard for the pre-hospital triage of patients outside of a major incident setting, considering a 5% under-triage rate and 20% over-triage rate acceptable[25].

Recent UK major incidents: blast and penetrating trauma

Planning for major incidents, including choice of triage tool, should cater for any eventuality i.e. have an "all-hazards" approach, however for any given region or nation, specific types of major incident occur with greater frequency[2,5]. Elements of disaster planning for any region or nation, including choice of triage tool, must consider the prevalent mechanism of major incident during planning[2].

Historically in the UK, explosions have occurred accidentally, resulting from mining accidents and the storage of munitions or agricultural chemicals (e.g. ammonium nitrate)[26]. In recent times, they have resulted from attacks using explosives[27-29]. During the Troubles in Northern Ireland (1968-1998), violence spread to parts of mainland Britain from 1972 onwards, resulting in bomb attacks targeting British military personnel as well as civilian urban centres[30]. Vehicle-borne bombs detonated by a timing device could carry a relatively large amount of explosives as a means to kill and injure people or incur significant economic and infrastructural damage, as seen in the 1973 Old Bailey bombing and in central Manchester in June 1996[31]. Bombs were also conveyed by parcel, or left in buildings, as seen in the 1974 Birmingham pub bombings, killing 21 people and leaving 182 injured[30].

However, more recently, the UK has encountered suicide bombings associated with extremist religious ideology[29]. The first of these was the 7 July 2005 London Bombings, involving improvised explosive devices containing triacetone triperoxide (TATP) packed into backpacks[28]. Four co-ordinated attacks occurred targeting commuters on three underground trains, and later a bus. This left 56 dead, including the four perpetrators and injured 784 people. The effects of these explosions were amplified by the enclosed nature of the underground tunnels. The 2017 Manchester Arena bombing involved a single suicide bomber, who detonated a TATP-based improvised explosive device packed with almost 2,000 nuts as 14,000 people exited a concert venue[27]. This left 23 dead (largely children) and over 400 injured, 116 of whom required hospitalisation. Such events have the ability to rapidly overwhelm the entire trauma system and distribution of casualties must take into account the varying capabilities of hospitals (Major Trauma Centres, Trauma Units and Local Emergency Hospitals) within regionalised trauma networks[29]. Emergency care providers remained vigilant to the threat from secondary devices and the possibility of multiple co-ordinated attacks. In fact, the two blast events described account for the largest loss of life on UK soil since World War 2.

Injuries involving firearms and mass shootings are fortunately infrequent in the UK, where access to firearms is strictly controlled, compared with countries such as the United States where private possession of automatic and semi-automatic assault weapons is permissible[32]. Following the Hungerford massacre of 1987, in which 16 people were killed by a lone gunman armed with legally owned semi-automatic rifles, and the 1996 Dunblane primary school massacre in which 16 children and one teacher were killed, the Firearms (Amendment) Acts of 1988 and 1997 have banned the private possession of all cartridge ammunition handguns in the UK[33]. Interpersonal violence involving firearms in the UK is thus fortunately infrequent, with 110 offences per 1 million population in 2019[34]. The most commonly implicated firearms are illegally owned handguns, imitation firearms (e.g. BB guns) and shotguns34. Members of the public may own shotguns and sporting rifles, subject to licensing.

Strong gun controls in the UK mean that bladed weapons are more commonly used than firearms in terror-related incidences and cases of interpersonal violence[29,35]. Bladed instruments include anything that can be used to cut or puncture. In recent major incidents, the use of bladed instruments have been combined with vehicular attacks on civilians in crowded, urban settings, as seen in the 2017 Westminster attack and 2017 and 2019 London Bridge attacks[29]. Attacks involving bladed weapons have formed a high proportion of UK major incidents in the last 15 years, however they are still relatively infrequent occurrences overall[29]. In terms of major incident planning, particularly from a secondary care perspective, much can be learnt from the management of patients suffering knife injuries outside of a major incident. For example, it is known that victims of penetrating trauma experience higher early mortality and a higher proportion of victims require operative intervention compared with blunt-injured patients, which has implications for major incident planning[36]. A London-based study has demonstrated that terror-related bladed attacks tend to involve relatively older victims, affecting both genders equally, and victims demonstrate a higher injury severity with a greater likelihood of requiring surgical intervention owing to the lethal intent of perpetrators[35].

Less frequent mechanisms amongst UK major incidents have included burns and the use of a chemical nerve agent[29,37]. A large residential fire in London involving the Grenfell tower block in 2017 was responsible for claiming 72 lives (70 died at scene) and injuring a further 70 victims[38]. The rapid spread of fire has been linked to the cladding material used, which has led to a thorough review of the use of building materials in new and existing residential buildings, to mitigate the risk of such ocurrences in future. A major incident was declared following use of the nerve agent "Novichok" in Salisbury in 2018[37]. Major incidents relating to the use of Chemical, Biological, Radiological and Nuclear (CBRN) agents are fortunately a very rare occurrence in the UK and uncommon internationally, especially since such offences are illegal according to international humanitarian law, some of which are described in the Chemical Weapons Convention[37]. As such, no official contemporary databases of healthcare data exist with which one may test the performance of existing triage tools amongst CBRN victims. A recent example of the illegal use of chemical weapons includes the use of chlorine for hostile purposes in Syria, with an incident in 2017 estimated to have affected several hundred individuals[37].

Overall, in the last 15 years, man-made blast and bladed mechanisms have predominated amongst UK major incidents, accounting for significant mortality and morbidity[29]. Hence, whilst major incident planning (including choice of triage tools) in the UK should cater for any eventuality, it is particularly important that any plans put in place are well suited to dealing with incidences involving blast and penetrating trauma. As major incidents occur infrequently overall, several lessons may be learnt from penetrating trauma occurring outside of a major incident.

Major incident triage in the UK: Current practice in the context of regionalised trauma systems of care

With regards to major incident management in the UK, the National Ambulance Resilience Unit (NARU) is centrally responsible for providing a co-ordinated approach to emergency preparedness, resilience and enables NHS Ambulance Trusts to work together to provide a safe and reliable response to major, complex and potentially protracted incidents[39]. NARU recommends the use of the NARU Triage Sieve for the primary triage of adults and the Triage Sort as the secondary triage tool of choice across the UK[39]. The Paediatric Triage Tape (PTT) has been recommended for use amongst children (<12 years) for many decades, however hospital-based guidelines have recently recommended the use of JumpSTART for children arriving following a major incident[3]. Each tool is described in depth, alongside other international tools, in the next section. Overall, the Joint Emergency Services Interoperability Programme (JESIP) has a national role in ensuring interoperability between the practices and procedures of multiple agencies that may be involved in responding to major incidents[1].

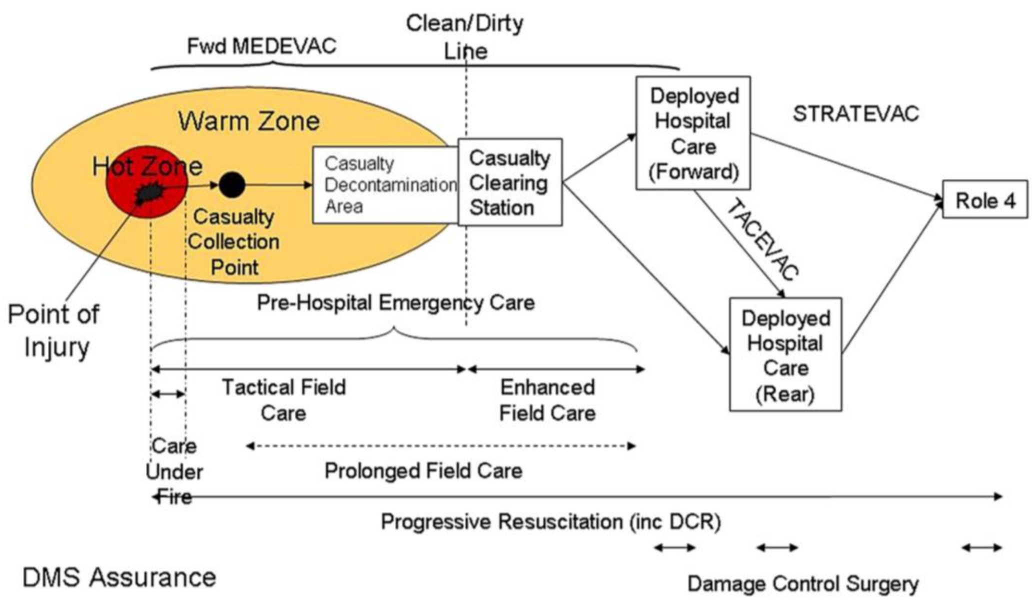
In England, trauma remains the commonest cause of death in those under 40 years[40]. Since the creation of a National Health Service (NHS) in 1948, emergency care has been based upon ambulances transporting patients to the nearest Emergency Department, regardless of the hospital's capability to provide trauma resuscitation and definitive care. There was growing awareness that outcomes following trauma in England were lagging behind those achieved in other countries in which organised trauma systems including designated trauma centres existed[23,40-42]. The NCEPOD 2007 report "Trauma, who cares?" identified significant shortcomings in the organisation of trauma care in England[40]. In response to this, the National Audit Office recommended the development of regional trauma networks in England. In April 2012, a series of Regional Trauma Networks became operational, with 27 designated Major Trauma Centres (MTC) including dedicated paediatric MTCs. The London network started earlier in 2010. Ambulance services are now trained and equipped with a trauma triage algorithm which facilitates the bypassing of local hospitals in favour of the regional MTC in the case of individual patients suffering severe injuries[25]. MTCs receive additional government funding per patient by means of "Best Practice Tariff," based on the severity of injury of patients[43]. Other hospitals within these regions were designated Trauma Units (TU).

In 1988, a working party at the Royal College of Surgeons recommended several changes in trauma management, including the need to audit and research injury and systems of care[43]. This gave rise to the Trauma Audit and Research Network (TARN) registry, which has a central role in the quality assurance of care provided across regionalised trauma systems of care. Based at the University of Manchester, TARN now receives clinical data from the point of injury until the rehabilitation phase of care from all 169 trauma receiving hospitals in the UK[43]. TARN focusses heavily on patients with major trauma and as such, inclusion criteria are: length of stay of 72 hours or greater, the need for intensive care and/or death following injury[43]. Notably, TARN does not include patients who die in the pre-hospital environment. TARN does not include burns victims unless they have concurrent traumatic injuries. TARN also excludes those with isolated pubic ramus and neck of femur fractures. Each trauma receiving hospital has a trained TARN co-ordinator who is responsible for the prospective identification of eligible patients and submission of specified data fields covering the incident, pre-hospital care, emergency department care, imaging, critical care, ward, discharge and rehabilitation phases of care. Analysis of TARN data has demonstrated that the regionalised, tiered system of trauma care, which has been in place for over ten years, has resulted in a 0.08% excess of survivors per quarter since inception and a 19% increase in the adjusted odds of survival from severe injury between 2008 to 2017[23].

The major incident triage category assigned to a patient determines not only the order of treatment or transfer away from the scene, but also the location for the delivery of definitive care[1,39]. Within these regionalised systems of trauma care, regional major incident casualty distribution plans recommend the transfer for Priority 1 casualties to Major Trauma Centres, Priority 2 casualties to Trauma Units and those with minor injuries (Priority 3) to minor injury units or emergency departments some distance from the scene of the major incident[3,39]. The appropriateness of this pattern of casualty distribution, in terms of resource requirements, has not been formally assessed.

The UK Defence Medical Services (DMS) delivers healthcare to deployed military personnel via the Operational Patient Care Pathway (OPCP), which is a coherent end-to-end healthcare system, from point of injury to the home base hospital[14].

Figure 1: The operational care pathway (UK military)



Ledger: DMS=Defence Medical Services. MEDEVAC=medical evacuation to a place where medical care can be delivered. Strategic evacuation (STRATEVAC) is the movement of patients from the operational theatre to a first world-standard medical facility, designated a role four (R4) facility, usually in the UK. That is – out of theater to definitive care). Tactical evacuation (TACEVAC) encompasses both medical evacuation (MEDEVAC) and casualty evacuation within the operatational theatre. DCR=Damage Control Resuscitation

The diagram above illustrates this rearward chain of evacuation from a non-permissive environment to a more permissive environment where care delivered increases in complexity as a casualty move further away from the point of injury[14]. Initial care is provided by the "buddy-buddy" system where a casualty is moved out of the line of fire by a colleague. Following this, the casualty is tended to by a Combat Medical Technician, who can administer life saving interventions as per the Battlefield Advanced Trauma Life Support algorithm (BATLS) in a semi-permissive environment[14]. The casualty is then extracted rearwards via a chain of facilities or including a casualty clearing station (Role 1), a forward deployed hospital (Role 2), a rearward deployed hospital (Role 3) and finally the Home base hospital, which is currently the Queen Elizabeth Hospital, Birmingham (Role 4).

The Battlefield Casualty Drills (BCD) handbook is provided to all service personnel within the UK Armed Forces. Included within this is the BCD Triage Sieve, an algorithm which ordinary soldiers are trained to use when faced with multiple casualties, including but not limited to situations in which a major incident is declared[44]. The UK military's Clinical Guidelines for Operations support the use of the Triage Sort as a secondary triage tool, however this appears to have fallen out of favour with no reports of recent use in practice45. Evidence suggests that the Triage Sort performs poorly with a sensitivity of 15.7% in predicting Priority 1 status (defined as the need for life-saving interventions) amongst injured personnel[45].

A high-quality, prospective military trauma registry exists to capture clinical data from injured servicemen and women, as well as other patients eligible for care at deployed British military treatment facilities. The Joint Theatre Trauma Registry (JTTR) includes all injured patients who triggered trauma team activation at a deployed Medical Treatment Facility, largely comprising combat casualties injured during military operations in Iraq and Afghanistan. Inclusion criteria were expanded in 2007 to include all patients evacuated to the Royal Centre for Defence Medicine, Birmingham following injury. Patients from all services and nationalities are included, as well as children, pregnant women and detainees. Studies have highlighted the challenges of collecting comprehensive pre-hospital data during combat, whilst at the same time highlighting its importance, since it is in this environment that most trauma deaths occur[46]. A previous study seeking to model a novel major incident triage tool found that only 17% of JTTR patients had a complete set of core pre-hospital physiological data with which to model a novel tool[47]. Nonetheless, the JTTR has captured high-quality data that has formed the basis of important clinical research on military subjects, may findings of which have shaped aspects of contemporary trauma care, particularly with regard to transfusion practices and the role of damage control resuscitation[48-50].

Existing major incident triage tools and their performance

Existing major incident triage tools

The purpose of a major incident triage tool is to categorize patients for transfer and treatment such that finite resources are directed towards maximising overall survival. Several such tools exist internationally, all of which are paper-based and many of which have been derived using expert opinion. Major incident triage tools can be divided into primary and secondary triage tools.

Primary triage occurs at the scene of a major incident, which may be a “semi-permissive” or physically dangerous environment. Individuals with basic medical training such as a paramedic typically perform primary triage[5]. In the United States, members of the fire service rather than paramedics routinely perform primary triage[8]. Existing primary triage tools consist of simple, paper-based algorithms, which can be applied rapidly, in a stepwise pattern[11]. The purpose of primary triage is to stratify patients by how urgently they need treatment, how urgently they need to be transferred from the scene, and importantly, triage category assignments also form the basis of the choice of medical treatment facility for definitive care (casualty distribution)[6,10,39]. Primary triage tools may incorporate basic life-saving interventions such as airway manoeuvres, application of tourniquets and/or haemostatic dressings for catastrophic haemorrhage and the placement of unconscious casualties into the recovery position[51]. Primary triage tools for use in adults include the US-based Simple Triage and Rapid Treatment (START)[6] and modified version, MSTART[11], the UK military's Battlefield Casualty Drills (BCD) Triage Sieve[44] and the National Ambulance and Resilience Unit (NARU) Triage Sieve[39], which was adapted from the UK military's former MIMMS Triage Sieve[11]. Novel tools include the Modified Physiological Triage Tool (MPTT)[47], it's derivative the MPTT-24[52,53], and US-based Rapid Assessment of Mentation and Pulse (RAMP)[54]. The Australian CareFlight, renowned for it's simplicity, is applicable to both adults and children[55]. Existing paediatric-specific primary triage tools include the UK-based Paediatric Triage Tape (PTT) applicable to children under 12 years[56], and the US-based JumpSTART, developed for use in children under eight years, alongside the adult START tool[57]. Whilst elders constitute a growing proportion of the population of most countries including 17% of the UK population, there are no major incident triage tools specifically targeted at elders[58]. Only one triage tool (Sacco Triage Method) considers advanced age as part of the triage process, despite recognition that they suffer worse outcomes following injury[59,60,61]. The US-based Sacco Triage Method (STM) was designed to be applied at scene (primary triage), however its complexity including the calculation of points based on physiological measurements bears similarity to the UK-based secondary triage tool, the MIMMS Triage Sort[45,60].

Secondary triage often, but not always, follows primary triage[10]. According to the MIMMS approach, secondary triage is performed within a permissive setting, such as a casualty clearing station or a hospital reception area[39]. Here, triage is typically performed by a doctor in a place of relative safely, hence a more detailed evaluation of the casualty is often possible, including blood pressure measurement and calculation of Glasgow Coma Score (GCS)[62]. Thus, it is expected that the secondary triage process will offer more accurate triage compared to primary triage tools, however this is not always the case[45]. The most widely described secondary triage tool is the points-based MIMMS Triage Sort[11].

Several studies highlight the important point that triage is not a static process[51,63,64]. Casualties can improve following an intervention (e.g. application of a tourniquet for catastrophic haemorrhage) or deteriorate over time (e.g. splenic rupture, haemopneumothorax, transient response to fluid resuscitation)[46,65]. Hence, the assignment of a triage category at one point in time is not definitive and casualties must be reassessed at various points in time as they transit through the system of care. On reaching a hospital setting, it is common practice for a senior clinician to perform triage, such as a surgeon or emergency department doctor, in order to determine the order of treatment within the hospital setting. The evidence for triage by senior clinicians has been anecdotal[10]. Triage beyond the hospital reception stage has not been considered in this thesis.

Below, adult and paediatric-specific triage tools are each described in more detail in chronological order of introduction into practice, including how the tool was developed (where this information is available), current use and geographical distribution, and tool performance as described in the existing body of scientific literature. A diagram of each tool is included in Appendix 1 (Existing major incident triage tools).

1. Simple triage and Rapid Treatment (START) Triage

The START triage tool was formulated by Newport Beach Fire and Marine Department and Hoag Hospital in Newport Beach, California in 1983[51]. It is the earliest major incident triage tool described and the most extensively studied amongst existing triage tools. It is a primary triage tool applicable to individuals aged over 8 years, used widely used in the USA as well as Canada and parts of Australia[11]. The algorithm quickly dismisses ambulatory patients into the Minor (green) category, followed by further categorization into Deceased (black), Immediate (red) and Delayed (yellow) categories on the basis of airway status, the presence and rate of breathing, perfusion status (namely, capillary refill) as well as the ability to follow commands. No anatomical criteria are used. The algorithm allows two interventions to be performed: airway positioning and control of bleeding.

A post-event evaluation sought to determine the effectiveness of START when prospectively applied to 148 victims of a train crash in California in 2003 (blunt injury mechanism)[6]. This is the only prospective study evaluating triage tool performance in actual victims of a major incident. The end-points measured included modified Baxt criteria (see later) to define Red (Priority 1) status and the need for hospital admission[66]. It was found that START ensured acceptable levels of under-triage (100% red/Priority 1 sensitivity and 89% green/Priority 3 specificity) but incorporated a substantial amount of over-triage whereby 54.4% of patients across all triage categories were assigned a higher category than their outcomes-based designation[6]. Notably, START proved useful in prioritizing transport of the most critical patients, with red category patients arriving in hospital a median of one hour before patients assigned other categories[6]. This finding reiterates the value of triage tools in ensuring the prioritisation of the most critical patients.

START has been criticised for lacking an evidence-base[59]. Additional criticisms of the START system are that respiratory rate is difficult to measure accurately and that a high respiratory rate can be caused by anxiety rather than pathology following injury in a major incident, which may account for the high degree of over-triage seen with this system[6]. Further criticism includes the use of capillary refill for field triage which may be inaccurate in a cold or dark environment, or in patients with dark skin. Others have described START as being too complicated overall to apply under the "fear effect" that arises following man-made major incidents, which have become more prevalent in recent years[10,67]. Indeed, its use during the 9/11 co-ordinated terrorist attacks on New York's World Trade Centre illustrated this weakness[8]. Measurement of capillary refill has been eliminated in the modified START tool, described below.

2. The Major Incident Medical Management and Support (MIMMS) Triage Sieve and Sort

The MIMMS approach was developed by two British Army doctors following the bombing of their base hospital during the Irish troubles in the 1980s[5,68]. As well as extensive guidance on how to approach and manage a multi-agency major incident, MIMMS included a primary triage tool (the MIMMS Triage Sieve) for adults (age 12 and above) that could be readily applied by minimally trained personnel without the need for specialist equipment[5]. This tool involves four assessments in total: ambulatory status, ability to breathe spontaneously, calculation of respiratory rate and measurement of capillary refill time. The only permitted intervention is to open the airway of a non-breathing patient, following which a patient may be declared Priority 1 (if begins to breath) or Dead (non-breathing following airway manoeuvre). This tool is the precursor for several UK-based tools including the BCD Triage Sieve, the NARU Triage Sieve and to some degree, the MPTT and MPTT-24[11,47]. The MIMMS "Sieve and Sort" approach, involving rapid primary triage at scene followed by a more complex secondary assessment involving medical equipment in a more permissive setting, dominated UK military and civilian practice for several decades.

The MIMMS Triage Sort is a points-based secondary triage tool. Three assessments are conducted for every patient (respiratory rate, systolic blood pressure and Glasgow Coma Score), and a score of zero to four is assigned for each parameter. Scores are summated and those with a score of less than 10 are assigned Priority 1 status. This tool was calibrated to correspond to the Revised Trauma Score, which predicts mortality[11]. Hence, it works on the rationale that once patients arrive in a medical treatment facility alive, those that at greatest risk of mortality should be treated first, irrespective of whether their injuries are compatible with life.

3. Battlefield Casualty Drills (BCD) Triage Sieve

The Battlefield Casualty Drills (BCD) Triage Sieve is the UK Armed Forces' current triage tool for use by ordinary soldiers when faced with multiple casualties, including but not limited to situations in which a major incident is declared[44,68]. It appears within the BCD handbook, provided to all service personnel within the UK Armed Forces[68]. It was first introduced in 1998, undergoing serial updates by experts, in line with emerging evidence and changes in clinical practice[68]. The most recent update in 2018 built upon the 2007 version by incorporating a revised respiratory rate threshold (adopted from MPTT-24)[44], an assessment of mental status, a revised heart rate threshold and the rolling of unresponsive patients into the three-quarter prone position. At the time of writing, there were no published evaluations of the BCD Triage Sieve present within the scientific body of literature.

4. CareFlight

The CareFlight triage algorithm was introduced into use in Australia in 2001 and is known to be one of the simplest and fastest triage methods to apply. The authors claim that it takes only 15 seconds to triage a patient using this method[11]. CareFlight results in the assignment of "immediate," "delayed," "minor" and "unsalvageable" categories, using background colours consistent with other common triage tools. Like the SALT triage tool, it does not distinguish between expectant and dead casualties, labelling any casualties who do not breathe after an airway manoeuvre as "unsalvageable." This tool does not involve any counting or arithmetic calculations.

CareFlight has been successfully applied to both adults and children, and demonstrated utility in the 2003 Bali bombings, in which burns and shrapnel injuries predominated[61]. After START, CareFlight is perhaps the most widely validated tool, having been applied to Australian, South African and UK civilian and military populations (see later)[17-20,55,69,70].

5. Sacco Triage Method (STM)

The Sacco Triage Method (STM) is a points-based triage method first described in 2005 by Sacco et al66. It was commercially developed in response to the 9/11 co-ordinated terrorist attacks on New York's World Trade Centre. STM involves the calculation of an "RPM" score, based on the sum of coded values for respiratory rate, pulse rate and motor response[60]. The tool is designed to predict mortality, incorporating some adjustments to the points with age.

STM was developed using complex mathematical modelling to predict the probability of mortality in a time-dependent fashion, which the authors describe as being well suited to the resource-constrained major incident setting[60]. The authors have validated the STM's ability to predict the probability of time-dependent victim survival amongst over 75,000 blunt injured civilian patients from Pennsylvania[60], over 7,000 civilians with penetrating injuries[71] and a further 99,369 military-age victims[59], using complex modelling to demonstrate that STM's RPM score predicted survivorship equivalent to the Revised Trauma Score and exceeded that of the Injury Severity Score. The authors state that STM outperforms START in every respect[60]. Unlike START, which is open-sourced and in the public domain, STM is proprietary and licensed for commercial use only[60].

One independent evaluation of STM included a prospective, observational study using virtual reality simulation technology based on actual major incident cases that resulted from a train accident in Chatsworth, Los Angeles, California in 2008[72]. 26 advanced care paramedic students were randomised to use either the STM or START method: the mean total triage times were slightly longer amongst those individuals using STM, although the difference was not statistically significant. The difference in triage ranking order between the two tools related to one patient's ambulatory status. The absence of obvious advantages over START coupled with the proprietary status of STM have been a deterrent for more widespread use of the STM.

6. The Modified START (MSTART) Triage

Described in 2006, this is a modification of the START Algorithm with the elimination of capillary refill as a measure of circulation, instead only the presence of a radial pulse is used[11].

In a small study evaluating the prospective application of this tool by German EMS on 152 trauma patients outside of a major incident setting, the overall triage accuracy of the tool across all triage categories was 62%, with a very low over-triage rate amongst P2 and P3 casualties. However one in two P1 casualties were missed, particularly those with head injuries[73]. The authors acknowledged the need for a larger, multicentre study to examine the performance of MSTART more robustly[73].

7. SALT (Sort, Assess, lifesaving intervention, treatment/transport) Triage Algorithm

SALT is a primary triage method, developed using expert opinion, introduced in the United States in 2008. It was a product of a Centre for Diseases Control Sponsored working group, who sought to standardise triage practices across the USA[74].

It has the advantage of being applicable to adults as well as children[74]. The first step involves a global assessment of how mobile and responsive the patient is, followed by a more detailed individual assessment at the scene, which differs from the MIMMS "Sieve and Sort" approach of a single assessment at scene followed by a second in a different location[5]. Individual assessment notably involves several subjective assessments: determining whether a patient is in respiratory distress (no parameters are set to define this) and whether major haemorrhage is controlled. Another unique factor is reliance on the user to apply judgement in deciding whether the patient is likely to survive given current resources, which determines whether the patient will be evacuated immediately or not at all (expectant category). This expectation of a high level of situational awareness amongst first responders has received criticism[10], which may be particularly challenging in major incidents involving multi-site and coordinated terrorist attacks, and may result in inter-user variability in assigning triage categories. The subjective judgements involved also make it very difficult to apply this tool retrospectively in order to determine its inherent ability to discriminate between triage categories.

SALT allows several interventions as part of the triage process including controlling haemorrhage, opening the airway, chest decompression and the administration of auto injector antidotes. It is unique in permitting the last two interventions as part of a primary triage algorithm. Notably, SALT does not involve any physiological measurements that involve arithmetic, such as heart rate and respiratory rate; in fact, it was developed to avoid these in view of the recognition that such estimations are likely to be performed inaccurately under challenging circumstances[74,75].

In a simulated, independent evaluation of SALT, the accuracy of SALT in assigning all triage categories was 78.8% and practitioners required a mean duration of 15 seconds to triage each patient (range 5-57 seconds)[16,76].

8. The National Ambulance Resilience Unit (NARU) Triage Sieve

The NARU Triage Sieve is in current use in the UK as the preferred civilian major incident primary triage tool, as recommended by the National Ambulance Resilience Unit[39]. It is derived from the UK military's former MIMMS Triage Sieve11. Introduction of the "big C" ("Catastrophic haemorrhage?") assessment as the initial step was adapted from military practice into this tool in 2014, following the recognition that death from compressible haemorrhage could easily be prevented using tourniquets[77]. Following this assessment, the tool allows for the early identification of uninjured ambulatory individuals to be directed to a survivor reception centre. The NARU Sieve assesses ambulatory status, breathing, followed by measurement of the respiratory and heart rates, both of which involve manual counting and arithmetic calculation. The only permissible interventions are opening of the airway and placement of the unconscious patient into the recovery position. The colour codes of red, yellow and green are used in conjunction with the P1, P2 and P3 triage category assignments, as seen elsewhere. No "expectant" category exists and patients who do not breath after an airway manoeuvre are declared "dead." The performance of the NARU Triage Sieve has been compared with the MPTT, Careflight and START primary triage tools in UK civilian and military trauma populations (see later)[47,78].

9. The Amberg-Schwandorf Algorithm for Primary Triage (ASAV)

This tool is the German adaptation of the MSTART, first described in 2013[79,80]. It involves six qualitative assessments, avoiding the need for arithmetic estimations of physiological parameters. It begins with assessing ambulatory status followed by a judgement by the user of whether the patient appears to be "deadly injured," in which case the category "Dead" is assigned. The next step involves an assessment for "breathing difficulties" involving overall clinical judgement rather than arithmetic calculation of respiratory rate. An assessment for "spurting haemorrhage" constitutes the fourth step; this contrasts with several UK-based tools (e.g. BCD Triage Sieve, NARU Triage Sieve), which recommend assessment for external haemorrhage (referred to as "catastrophic haemorrhage") as the opening step on the basis that preventable large-volume external haemorrhage may kill faster than a compromised airway[11,65,77,79]. The final two steps are in common with CareFlight, with assignment of Red/Priority 1 status to patients with an absent radial pulse and those who are unable to follow a command (i.e. GCS Motor Score <6)[11].

ASAV permits interventions to open the airway and to stop external haemorrhage[80]. The terminology utilised is assignment of Priority I, II or III (Roman numerals) which correspond with the colours red, yellow and green, as seen in several other tools. The colour black corresponds with the "Dead" category. Like the SALT tool, the subjective judgements involved make it very difficult to apply this tool retrospectively in order to determine its ability to discriminate between triage categories. Currently, there are no published evaluations of the accuracy of this triage tool in real-life patients.

10. The Modified Physiological Triage Tool (MPTT) and its derivative, the MPTT-24

Designed by Smith and Vassallo in 2015, the MPTT was derived using a military cohort of patients from the Joint Theatre Trauma Registry[47]. The authors used logistic regression to predict the presence of any one of 24 life-saving interventions. This outcome measure to define Priority 1 status was developed by the authors via a Delphi consensus exercise of their UK and South African-based colleagues, specifically for use in modelling a novel tool[81]). The MPTT involves 5 steps including assessment of ambulatory status, ability to breathe spontaneously, respiratory and heart rates and estimation of a full GCS score[47].

In 2017, the MPTT underwent modification into the MPTT-24[53]. The modification from an upper threshold respiratory rate of 20 to 24 was felt by the authors to be more easily applicable in practice, however this claim has not been substantiated by implementation studies[82]. Three versions of the

MPTT-24 appear in the literature under the same name, with variations in the order of assessments and the choice of mental statement assessment (GCS<14 vs. Responds to voice?)[3, 52,53]. For the purpose of this thesis, the version appearing alongside the tool's validation has been considered as the definitive version[53].

The MPTT and MPTT-24 have yet to be used in a major incident setting. In contrast to other existing primary triage tools, MPTT and the initial MPTT-24 involve calculation of a full Glasgow Coma Score at the scene of a major incident. All iterations of the tools are unique in assigning the "dead" category to any non-breathing patients without permitting airway manoeuvres; this may raise ethical concerns since simple airway manoeuvres are quick to perform and may save the lives of several injured patients[25]. As these tools are relatively new, there are no published studies to verify the ease and accuracy of their application in practice[82]. The NHS Clinical Guidelines for Major Incidents state that the MPTT-24 can be performed in 30 seconds, however there are no published reports of how the authors determined this[3]. The MPTT has been validated by retrospective application to the largest civilian trauma registry in Europe, the Trauma Audit and Research Network (see later)[21]. At present, there are no independent evaluations of MPTT or any of the version of MPTT-24 in the existing body of scientific literature. Further studies are required to examine the ease and accuracy of the application of this tool, which has yet to undergo implementation studies yet is more complex than its comparators.

11. Rapid Assessment of Mentation and Pulse (RAMP)

RAMP was introduced by the Rocky Mountain Fire Department, Colorado in 2018. Described by Keating in 2016, this tool sought to simplify the triage process, recognising that the "fear effect" in responders was likely to result in a loss of critical thinking and fine motor skills[10]. The tool is based around two qualitative assessments: ability to follow a command and presence of a radial pulse (assumed to represent a systolic blood pressure of 100mmHg); if the outcome of either one of these is negative, the patient is assigned the highest triage category "Urgent" (rather than "Delayed" or

"Expectant/Deceased")[10]. RAMP permits three interventions to be performed during primary triage: control of massive haemorrhage, maneouvres to open the airway and chest decompression.

RAMP was developed by expert opinion, with the main outcome of interest being mortality; the scientific evidence in support of the tool arose from the care of combat casualties using the Field Triage Score[77],[83]). More specifically, the author noted that initial GCS correlates directly with hospital discharge in trauma, but since GCS is very difficult to score accurately under challenging circumstances, the ability to follow a command is the next best substitute in predicting survival[90]. Additionally, they observed that lack of a radial pulse and inability to follow commands is associated with a mortality rate of 92%[84].

Regarding the performance of RAMP, the authors undertook a simulation exercise involving 19 patients and compared the performance of RAMP to the widely used and extensively studied START tool[85]. They demonstrated that the time spent to triage a patient was shorter (45 vs. 60 seconds), overall triage accuracy was 84% versus 58% using START, and that all "Red" (Priority 1) patients were removed from the scene faster when RAMP was utilised (20 vs. 30 minutes, respectively)[85]. The authors describe the benefits of RAMP as facilitating the rapid identification of the most severely wounded, ease of use, easy to teach, lacking in reliance on numbers or critical thinking and being based on scientific evidence. At present, there are no publications describing the independent evaluation of RAMP in the scientific literature.

12. Paediatric Triage Tape (PTT)

The Paediatric Triage Tape is the paediatric modification of the MIMMS Triage Sieve, first introduced in 1998[86]. Since use of adult physiological parameters are likely to result in an artificially higher triage category assignment in children, PTT includes physiological parameters specific to the child's height. A "tape" is laid out next to the child, starting at their head end, and the algorithm closest to the child's feet (one of four choices) is used to triage the child. Children who are known to be 12 years or older, or those whose height exceeds that of the triage tape (>140cm), must be triaged using an adult tool[86].

In 2006, an independent validation of the PTT was performed using prospectively collected data from 3461 children who attended a South African trauma unit within 12 hours of injury[56]. Amongst these, median age was 7 and 93% of patients were male; no breakdown of injury mechanism was provided. The PTT was shown to be modestly sensitive (41.5%) in predicting the need for life-saving interventions according to Garner's criteria, particularly in children of the 50-80cm height block (25% sensitivity) and 80-100cm height block (23.6%). Specificity for Garner's criteria was remarkably high at 99.6%; PTT is excellent at identifying children who are non-P1. Additionally, PTT demonstrated poor sensitivity for predicting ISS of 16+ (37.8%) and NISS of 16+ (26.1%), when these alternative outcome measures were applied.

13. JumpSTART

JumpSTART is a US-based paediatric tool first introduced in 2001[11,57.] It is widely used amongst children aged under 8 years, alongside its parent tool, START (applicable to adults). In 2018, JumpSTART was also incorporated into the UK's NHS Clinical Guidelines for Major Incidents, where its use was recommended in the hospital triage of children[3].

JumpSTART uses identical triage category assignments to that used by START and the tools follow a very similar sequence of assessments, although the respiratory threshold is revised (<15 or >45 breaths per minute consistutes P1 vs RR>30 in START) and neurological assessment involves the AVPU scale rather than ability to follow commands (cf. START utilises the GCS Motor Component). In addition to an airway manoeuvre, it allows the administration of five rescue breaths to apnoeic patients, recognising that children may have a respiratory cause for cardiopulmonary arrest, although this observation has not been substantiated in paediatric trauma patients[57].

A summary of the characteristics of existing triage tools is presented in Table 3.

Table 3: Summary of triage tool characteristics

Tool	Description and geographical use	Tool components				Remarks
		Assessment of respiration	Assessment of circulation	Mental Status assessment	Permissible interventions	
Adult Primary Triage Tools						
Simple Triage and Rapid Treatment (START)	United States (introduced in 1983)	Spontaneous breathing; RR>30	Capillary refill>2 seconds	Ability to obey commands	Position airway	5 steps in total (1 arithmetic)
MIMMS* Triage Sieve	Former UK military adult** triage tool (introduced 1995)	Spontaneous breathing; 10<RR>29	Capillary refill>2 seconds	No	Open airway	4 steps (1 arithmetic)
Battlefield Casualty Drills (BCD) Triage Sieve	Current UK military adult** tool (introduced 1998, revised 2018)	Breathes with open airway	Presence of catastrophic haemorrhage; HR>100	Responds to voice (component of AVPU)	Apply tourniquet; Open airway; Place casualty in the ¾ prone recovery position	6 steps (2 arithmetic). Expert-led update of the previous MIMMS Triage Sieve
CareFlight	Australian tool used in adults and children (introduced 2001)	Breathes with open airway	Presence of radial pulse	Ability to obey commands	Open airway	Applicable to adults and children. 4 steps (all qualitative measures)
Sacco Triage Method (STM)	United States, commercially developed points-based tool (described in 2005)	Respiratory rate	Pulse rate	Motor component of GCS (Ability to follow commands)		Points assigned for each variable, with an additional allowance for extremes of age. Proprietary nature has limited its dissemination.
Modified Simple Triage and Rapid Treatment (MSTART)	United States, modification of START (described in 2006)	Spontaneous breathing; RR>30	Presence of a radial pulse	Ability to obey commands	Position airway	5 steps in total (1 arithmetic); Modification of START - radial pulse assessment replaces capillary refill
Sort, Assess, Life-saving interventions, Treatment/Transport (SALT)	United States (Introduced by the Centre of Disease Control, 2008)	Spontaneous breathing; presence of respiratory distress	Presence of peripheral pulse; is major haemorrhage controlled	Obeys commands or makes purposeful movements	Control major haemorrhage; open airway (2 rescue breaths in children); chest decompression, auto-injector antidotes	5 steps including "global sorting"(all qualitative measures); involve subjective judgements: "Minor injuries only?" and "Likely to survive given current resources?" limiting retrospective application
National Ambulance and Resilience Unit (NARU) Triage Sieve	Current UK civilian adult** tool (this version was introduced in 2013)	Spontaneous breathing; 10<RR>29	Presence of catastrophic haemorrhage; HR≥120; capillary refill>2 seconds	Unconscious (Yes or No)	Application of tourniquet/haemostatic dressing; open airway; place casualty in recovery position	7 steps (2 arithmetic); Assesses for presence of injury. Adapted from the MIMMS Triage Sieve
Amberg-Schwandorf Algorithm (ASAV)	German adaptation of the MSTART (described in 2013)	Presence of breathing difficulties (see remarks)	Presence of radial pulse; presence of "spurting haemorrhage"	Ability to obey commands	Open airway; stop the bleeding	6 steps (no arithmetic); includes the subjective judgement "Deadly injured?" and subjective judgement of breathing status: "airway obstructed, bradypnoea, apnoea, dyspnoea, tachypnoea (not obviously psychogenic) and cyanosis;" limiting retrospective application
Modified Physiological Triage Tool (MPTT)	Novel UK-based tool modelled in a military cohort (described in 2017)	Spontaneous breathing; 12<RR≥22	Presence of catastrophic haemorrhage; HR≥100	GCS < 14	Application of a tourniquet or haemostatic dressing (no airway intervention permissible)	6 steps (2 arithmetic; GCS calculation); has yet to undergo practical use or implementation studies
Modified Physiological Triage Tool 24 (MPTT-24)	Novel UK-based tool, modification of MPTT (described in 2017)	Spontaneous breathing; 12<RR≥24	Presence of catastrophic haemorrhage; HR≥100	GCS < 14	Application of a tourniquet or haemostatic dressing (no airway intervention permissible)	6 steps (2 arithmetic; GCS calculation); replaces MPTT's upper RR threshold with 24; has yet to undergo practical use or implementation studies
Rapid Assessment of Mentation and Pulse (RAMP)	United States, used by the Rocky Mountain Fire Department, Colorado (introduced in 2018)	None	Presence of radial pulse	Ability to obey commands	Control massive haemorrhage; open airway; chest decompression	3 steps (all qualitative measures); Begins with subjective assessment "Casualty without obvious signs of death." RAMP is the only tool which doesn't include an assessment of ambulatory status
Paediatric-specific primary triage tools						

Paediatric Triage Tape (PTT)	Current UK paediatric tool (<12 years) (adapted from MIMMS Triage Sieve in 1998)	Spontaneous breathing; Respiratory rate (height-specific threshold)	Capillary refill <2 s (use child's forehead); Pulse rate (heigh-specific threshold)	Alert and moving all limbs (children <100 cm height) or Walking	Position airway	Use tape to gauge child's length in order to determine which set of physiological values to utilise for triage
Jump Simple Triage and Rapid Treatment (JumpSTART)	United States, used in several states alongside START in children <8 years (introduced in 2001)	Spontaneous breathing; 15<RR>45	Presence of radial pulse	Inappropriate "P" or "U" from AVPU assessment	Airway positioning; 5 rescue breaths if apnoeic	5-7 steps (1 arithmetic)
Secondary triage tool						
MIMMS Triage Sort	UK-based, points-based secondary triage tool (described in 1998)	Respiratory rate	Systolic blood pressure	GCS		Points (0-4) assigned for each variable; total score <10 denotes a Priority 1 casualty

Ledger: *MIMMS= Major Incident Medical Management and Support; **Adult refers to those aged ≥12 years; RR=Respiratory rate; HR=Heart rate; GCS=Glasgow Coma Score; AVPU refers to the Alert, Voice, Pain, Unresponsive scale (Neurological assessment). All tools described are applicable at the scene of a major incident (primary triage tools).

Historically, although not strictly a triage tool, the concept of "reverse triage" has been applied by military personnel, whereby triage has been used to identify those patients with minor injuries who could be treated and then returned to battle in order to maintain fighting power[87]. This concept has also been employed to increase the capacity of hospitals during disasters in order to facilitate the rapid discharge of casualties with minor injuries[87].

Challenges in evaluating and comparing the performance of major incident triage tools

There are several challenges in evaluating the performance of MI triage tools. Firstly, major incidents are unpredictable and therefore planning prospective studies are logistically challenging and the randomization of patients in a resource-constrained major incident setting would be widely regarded as unethical. Hence, in order to inform practice, policy makers and care providers must rely on assimilating the findings of other types of study design, each of which is prone to particular biases.

Post-event evaluations of actual major incidents provide a real-world assessment of the triage method in use for a specific major incident[28]. Here, triage is performed by actual emergency medical services providers, and the human factors that affect performance in this setting are fully assimilated, as noted in the evaluation of the use of START during the 9/11 co-ordinated terrorist attacks on New York[8]. However, this means that it is difficult to differentiate between inaccurate triage resulting from a poor inherent capability of tools to differentiate between triage categories and their incorrect application by practitioners. Other limitations are that one mechanism of injury often predominates in a single event; therefore, it is necessary to assimilate information from several events to draw meaningful conclusions in developing plans for an "all hazards" approach [29]. An Utstein-style template to standardise the reporting of the acute medical response in disasters has been developed; however, this highlights that there are no common or standardised definitions of each triage categories[7]. Each post-event evaluation uses author-selected end-points to define triage categories, ranging from ISS>15 to mortality or utilisation of hospital resources[10,28]. The various end-points that have been used to measure triage tool performance are considered in greater detail in the next section.

Since a post-event evaluation will only report on the triage of patients using the tool(s) in place in that system, some researchers have overcome this by retrospectively applying multiple tools to data from the victims of actual events. For example, Chellen and Walter retrospectively applied multiple triage tools to the victims of the London 7/7 bombings[69]. A key limitation of this approach is that there is often inadequate pre-hospital data available following actual events and that an individual event may involve a relatively small number of patients: any conclusions drawn from such studies must evaluate whether the demographics of the affected population are representative of the whole, in terms of implications for planning for major incidents[69].

Some studies evaluate tool performance by applying triage tools prospectively to patients attending hospitals following injuries incurred in regular day-to-day practice. Ideally, prospective studies would facilitate accurate and contemporaneous application of the tool in question, however this is a resource-intense and time consuming approach[88,89]. In some cases consecutive patients have been assessed[20] and in others, convenience samples have been used[94,95]. The advantages of this approach are that larger numbers of patients can be harnessed and a range of injury mechanisms can be covered[20]. However, the relative proportion of patients injured day-to-day with each injury mechanism might not be reflective of major incidents (e.g. proportion with penetrating trauma may differ). Furthermore, outcomes of regular hospital patients might be much better than those injured in a major incident, and triage is often "conducted" by a senior clinician in the safe and familiar hospital environment; these factors may limit the external validity of study findings[88,89].

Another means by which to measure major incident triage tool performance is by their retrospective application to patients captured by existing trauma registries. This assumes that patient physiology in response to trauma will be similar whether they present individually or as a cohort of casualties injured in a major incident setting. Since several trauma registries collect data prospectively, this approach minimises information bias. Studies that retrospectively apply tools to entire trauma registry populations at once allow more rapid procurement of results attained from very large numbers of patients and covering a broad range of injury mechanisms; this has proven particularly useful in children[18,70,90]. The majority of recent published studies have adopted this approach[18,45,47]. Computed application of tools allows the innate discriminatory capability of triage tools to be assessed independently from human error, which invariably arises in the real-world setting[18]. However, interpretation of such studies must acknowledge that all existing trauma registries have specific inclusion criteria, which may bias results, for example, trauma registries in developed countries that use hospital length of stay as an inclusion criterion may be prone to over-representation of elders compared with the general population[45]. Additionally, submission to some trauma registries such as the US National Trauma Database (NTDB) is not mandatory, hence patients contained within this registry may not be nationally representative in terms of patient demographics and injury characteristics[18]. Notably, following the 2017 Manchester Arena bombing, the UK's Trauma Audit and Research Network (TARN) registry has introduced a data capture field to record when patients have been injured in a major incident setting[43]. This will be of assistance in facilitating future studies examining triage, outcomes, transfer times and other aspects of patient care in a major incident setting and allow these patients to be assessed separately from those injured outside of a major incident setting.

Simulation studies have a role in allowing the direct comparison of two or more triage tools with all other factors remaining equal[16,91,92]. There is the possibility of mimicking, to some degree, the human factors at play in a major incident, as well as measuring the practical applicability of tools, including time taken to perform triage, and the inter- and intra-user reproducibility of results[82]. It is important to note that it is often impractical to test more than two tools using this approach, and that most such studies utilise student paramedics rather than qualified practitioners of varying experience[92]. Importantly, simulation studies allow the evaluation of the practical applicability of tools before subjecting actual patients to the proposed triage methods, providing a safe setting in which to learn lessons and gain feedback prior to wider dissemination[16,91,92]. Notably, several emergency responder agencies may chose not to publish the results of simulation or implementation studies as this may reveal aspects of their major incident response plans, which may compromise security in the context of man-made major incidents.

Furthermore, when evaluating the performance of major incident triage tools, controversy exists on what constitutes a P1 casualty. This is considered in more detail in the next section. A lack of standardized, universally acceptable definitions for triage categories severely limits our ability to compare the findings of studies assessing triage tool performance, which limits the opportunities to improve practice in this area[7].

What endpoints have been used to measure triage tool performance?

Studies measuring triage tool performance have largely focussed on correct identification of P1 patients, as these patients are at greatest risk of adverse outcome if they are not correctly identified[45,47], however other studies have looked at overall triage accuracy across all triage categories[88,89]. When evaluating the performance of major incident triage tools, controversy exists on how best to define a P1 casualty. A P1 casualty has been broadly defined as "one who is in need of immediate medical care." However, in academic terms, there is no consensus of how this is best measured. Studies have generally defined the Priority 1 casualty using one or more of three distinct end-points:

1. Those at greatest risk of mortality
2. Those with the most severe injuries
3. Those in need of life-saving interventions

Mortality

This definition of the P1 casualty assumes that those who are most likely to die are in greatest need of care. Mortality is the simplest outcome to measure. However, this doesn't take into account that a proportion of these patients cannot be saved and may die regardless of the care they receive (those who may be assigned P4/Expectant category), thus channelling resources on this basis might not be the most efficient practice. Furthermore, those who were at significant risk of death but received timely interventions which averted death would not be readily identifiable. Notably, the MIMMS Triage Sort tool was modelled upon the Revised Trauma Score, correlating with risk of mortality[11]. The STM was also modelled to predict the probability of mortality in a time-dependent fashion[60].

That said, a thorough understanding of the timing and cause of deaths after trauma is crucial to developing systems of care which maximise survival. In particular, understanding the degree to which death is preventable is crucially important. Triage at scene involves coming into contact with patients who may have signs of life but may have unsurvivable injuries (P4, expectant) as well as all the other triage categories.

In 1974, Trunkey described three temporal peaks in death after injury, referred to the "trimodal death distribution," highlighting that the majority of deaths occur within the first 24 hours of injury [93]. The first peak in the classic trimodal model of trauma mortality includes immediate deaths at the scene within minutes or the first hour of injury. It is thought that these deaths occur due to unsurvivable injuries and can only be prevented using injury prevention strategies and not by advances in treatment[93]. The second peak of early deaths is described as occuring within a few hours to 48 hours after injury. This group of deaths are potentially preventable through high quality acute care in the pre-hospital and hospital settings, for which appropriate triage with timely transfer and treatment are most important. The third peak in deaths occur between 1 to 4 weeks, from complications of trauma including multi-organ failure and sepsis. It is felt that high quality early care coupled with critical care can help minimise deaths at this stage. Indeed, this third peak is not observed in contemporary civilian practice and may have been eliminated through advances in trauma care in the decades after it was first described[93].

Whilst variations in the proportions of deaths at each time period has evolved since with further variation by age group and setting (military vs. civilian trauma), an appreciation of the underlying concept of temporal distribution of deaths is crucial to major incident planning. In contemporary studies, over 50% of trauma deaths occur within the first 24 hours, with no clear distinction between the first and second peaks[77,93]. It is known that penetrating trauma is associated with a higher early mortality than blunt trauma, which is highly relevant to planning, particularly for terrorist-related major incidents[10,65,77,94]. Deaths within a combat environment similarly occur within minutes to hours after injury, with haemorrhage cited as the leading cause of death (90.9%). The site of lethal haemorrhage was truncal (67.3%), followed by junctional (19.2%) and peripheral-extremity (13.5%) haemorrhage in a study of nearly 5,000 deaths recorded by the American armed forces in a ten year period encompassing recent conflicts in Iraq and Afghanistan[77]. In civilian studies, immediate deaths from trauma have most commonly been attributable to severe head injuries[93]. In a German civilian trauma registry study, 76% of deaths following trauma had a head AIS score greater than 31[95]. Severe chest injuries are the second most common cause of early trauma deaths followed by abdominal injuries, as seen in several international studies[77,94,96]. Overall, several studies highlight non-compressible haemorrhage as the leading cause of potentially preventable trauma deaths[77,96].

Overall, prediction of mortality is a crude but readily available outcome by which the performance of triage tools can be measured. However, an understanding of the timing and common injury patterns associated with death following trauma is crucial to the development of trauma systems of care and in particular, in the development of triage systems for the major incident setting. The most opportune time to save lives following trauma is in the first few hours following injury, hence the correct categorization of patients into immediate (Priority 1) and urgent (Priority 2) in order to facilitate timely transfer and treatment is of great importance.

Injury Severity Score >15

A further method of defining the Priority 1 casualty is one who has an Injury Severity Score (ISS) of greater than 15, which is a widely used definition of "major trauma" outside of a major incident setting. Brohi et al utilised this retrospective definition as equivalent to Priority 1 status when analysing casualty distribution and outcomes following the London 7/7 bombings[28]. ISS is an anatomically based scoring system in which each injury in every body region is assigned a score (Abbreviated Injury Score) based on its severity using a six-point scale, ranging from minor (=1) to maximal (=6 i.e. currently untreatable); ISS is then calculated by taking the highest AIS severity code in each of the three most severely injured ISS body regions, squaring each of these AIS scores and then adding the three squared numbers[43]. ISS>15 is also used as the justification for the requirement of Major Trauma Centre care in the UK, a key outcome measure considered by the UK's Trauma Audit and Research Network (TARN) registry[43]. This has been shown to correlate particularly well with mortality, and less so with morbidity and the length of hospital stay after trauma[97]. However, other studies have demonstrated that anatomical classification of injury correlates poorly with predicting the resource requirements (i.e. need for medical interventions) of patients, which is of critical importance in the resource-constrained major incident setting[56,66].

In one study, the New Injury Severity Score (NISS) has been used as an additional outcome measure of the performance of a triage tool on the basis that it may be a more accurate measure of injury severity[56]. NISS measures the three worst injuries in any body region, rather than using the highest score in each of three different body regions, and a threshold of NISS of 16 or greater was utilised[98]. However, the authors highlighted the key limitation of this, in that NISS focusses on injury pattern and not the requirement for medical intervention, which many regard as the main priority behind triage in a major incident[66].

The need for life-saving interventions

The third outcome used to measure triage tool performance is the need for life-saving interventions. This concept has gained popularity as a definition of Priority 1 in the last two decades, as it relates directly to patients' resource requirements in the major incident setting. However, amongst researchers and practitioners in this field, there is a lack of consensus on what interventions constitute a life-saving intervention, with various definitions of these, based on a combination of indirect evidence and specialist opinion.

Baxt and Upenieks outlined in 1990 that ISS was a poor predictor of the resource requirements of patients and should not be used to justify trauma centre care[66]. They suggested that the requirement for trauma centre care should be defined by the need for any one of three criteria: a non-orthopaedic operative procedure with positive operative findings within the first 48 hours of admission, fluid resuscitation of more that 1L or transfusion of blood products to maintain a systolic blood pressure of more than 89mmHg or the need for invasive central nervous system monitoring with raised intra-cranial pressure, or a positive head CT scan. Baxt et al also included death following injury amongst the criteria to define the need for trauma centre care[66]. Subsequently, in 2001, Garner et al modified Baxt and Upeniek's criteria to make these more specific to the the major incident setting[19]. They shortened the time to necessary surgery from 48 to 6 hours, included all other criteria described by Baxt and added in a fourth criterion: the requirement of a procedure to maintain a patent airway or requirement for assisted ventilation, either out of-hospital or on arrival in the ED[19]. In 2006, Wallis and Carley used these "Garner criteria" to validate the Paediatric Triage Tool[20]. The authors highlighted the relative merits; Garner's criteria covered many of the common interventions that a patient may require following injury in a major incident and that the criteria may equally be applied to both adults and children. They also acknowledged the weaknesses in that Garner's criteria were developed using expert opinion, they have not been validated as an outcome measure for a triage tool, and importantly, these criteria could not be used to measure triage categories other than Priority 1/T1[20]. In 2007, Wallis and Carley replicated this methodology with a paediatric-specific modification (inclusion of a fluid bolus of greater than 20ml/kg, in addition to the adult criteria of 1000ml) in a prospective, hospital-based study of the performance of multiple triage tools[20]. Kahn et al further modified the criteria used by Garner et al in 2007 in their post-event evaluation of a Californian train crash disaster in 2003[6].

In 2006, Wallis et al published a set of paediatric-specific criteria against which triage tools can be measured[99]. These were derived by a Delphi consensus including 16 pre-hospital and hospital based experts from the UK and South Africa. They defined T1, T2 and T3 (Priority 1, 2 and 3 equivalent) categories. T1 was defined as patients requiring major resuscitative and haemostatic interventions and the need for torso or other surgery within one hour, a very stringent cut-off. The definition of T2 included the need for DPL or FAST ultrasound in ED, both of which are uncommon interventions in contemporary practice. Contrary to the more widely held definition of “walking wounded,” the definition of T3 included the need for theatre within 1 day of injury (not including those needing surgery within an hour, who would be assigned T1)[99]. From a casualty distribution perspective, studies demonstrated that there is often little opportunity for inter-hospital transfer by EMS within the first 24 hours following a major incident and this definition would not align well with the transport of T3/P3 patients to minor injury units, for example[10,22]. Burns and CBRN-related interventions were not included in the criteria. Although proposed, these criteria have not been validated or applied in subsequent studies, with the authors themselves favouring the use of Garner’s criteria with paediatric modification in their study comparing the performance of multiple triage tools, published in the same year[20].

In 2013, Cross and Cicero undertook a large study involving over 500,000 patients from the US National Trauma Data Bank (NTDB), reporting in-hospital mortality as the primary outcome[100]. There were also a variety of author-selected secondary outcomes including time spent in ED, death in ED and overall hospital length of stay, which may not seem directly relevant to the choice of triage tool in the major incident setting (particularly, time spent in ED). The authors included the need for a ventilator as a secondary outcome: this reflects a particularly resource-intense subset of patients injured in the major incident setting and is therefore of value.

Adopting a similar approach to that of Wallis et al in 2006, Vassallo et al conducted a Delphi consensus of South African and British specialists and formulated a list of 32 life-saving interventions, including the need for airway interventions not previously described by Baxt, which formed the basis of modelling a new tool, the Modified Physiological Triage Tool (MPTT) and its derivative, the MPTT-24[47,81,97]. Relative strengths of this system include the use of commonly described medical interventions that are routinely recorded in patient medical records. This system considers CBRN mechanisms but fails to consider burns-specific definitions of P1. Potential disadvantages of this system are that it includes the administration of tranexamic acid (TXA) (regardless of haemodynamic status) and the application of a pelvic binder (regardless of the presence of underlying pelvic injury); practice around TXA administration and the application of pelvic binders are highly variable in practice, making these subjective surrogates for P1. The system appears to have been designed for adults with no mention of children. Vassallo's system proposes that patients with cardiac arrest (including at-scene cardiac arrest) should be considered Priority 1 patients, whereas many would consider these to be "Expectant" or "Dead[11]." Like the so-called Garner criteria, this system only defined the Priority 1 category and the criteria have not been validated as an outcome measure for a triage tool. Crucially, the timing of interventions were not considered in defining P1 status, such that patients requiring delayed surgery or those intubated or undergoing intercostal drainage weeks following injury due to respiratory complications were also considered P1.

In 2016, Lerner et al developed an objective system to define each mass casualty triage (Dead, Expectant, Immediate, Minimal, Delayed), incorporating the need for commonly described surgical and resuscitative interventions which would be routinely recorded in patient medical records, as well as the timing within which those interventions needed to be administered[101]. This system was developed based on an extensive review of the literature followed by a Delphi consensus of named US-based experts in adult and paediatric pre-hospital and hospital medicine with experience in disaster management. Strengths of this system include definitions of all possible triage categories using a logical flowsheet, consideration of all mechanisms of injury that may be implicated in a major incident including CBRN and burns, and the definitions are applicable to adults as well as children. The authors used their system to conduct comparative analyses of the performance of multiple tools in small convenience samples of adults[17] and children[88] presenting to ED in a US Level 1 trauma centre (see later), however they did not comment on whether the patients assigned each category differed with regard to clinically important characteristics such as mortality, requirement for intensive care, injury severity or length of stay.

None of the systems described have been internally or externally validated to determine whether P1 patients differ clinically from non-P1 patients in terms of key characteristics such as mortality, hospital length of stay, need for intensive care and injury severity. Several studies measuring triage tool performance use author-determined definitions and are therefore prone to bias, whereby definitions may have been based on the data available to investigators rather than scientifically objective criteria[47]. Amongst the various end-points used to define the Priority 1 patient, Lerner's criteria appeared to be the most objective and scientifically robust method to define triage categories[101].

Studies comparing the performance of multiple tools

Over the last two decades, several studies have sought to compare the performance of multiple triage tools using selected patient populations, some of which have used patients actually injured in a major incident[69] whereas most have used surrogates, such as trauma registry populations[18] or patients attending ED in trauma-receiving hospitals[19]. These studies have varied in the outcome measures selected and in their geographical location, spanning multiple countries and healthcare systems.

In an Australian study published in 2001, four tools (CareFlight, START, MSTART and the MIMMS Triage Sieve) were retrospectively applied to 1,144 consecutive adult patients (≥14 years, median age 33 years, 65% male) captured in 1994 by the trauma registries of two major trauma centres in New South Wales[19]. Most patients suffered a motor vehicle collision (39.1%) or a fall <5m (21.3%), with 2.8% suffering penetrating trauma whilst 1.8% suffered burns. Here, Garner modified the original Baxt criteria, as described (referred to as "Garner criteria" in subsequent literature) as the primary end-point measure. The differences in sensitivity between CareFlight, START and MSTART were not dramatic, with sensitivities of 82% (95% confidence interval [CI] 75% to 88%), 85% (95% CI 78% to 90%), and 84% (95% CI 76% to 89%), respectively. However, CareFlight demonstrated the highest specificity of 96% (95% CI 94% to 97%), compared with START (specificity of 86%, 95% CI 84% to 88%), and MSTART (specificity of 91%, 95% CI 89% to 93%)[19]. Notably, the lead author was employed by CareFlight. The study also found that of the physiologic variables, the Motor Component of the Glasgow Coma Scale (<6) had the strongest association with predicting the need for life-saving interventions, followed by a systolic blood pressure of <80mmHg[19].

Regarding paediatric triage, in 2006 Wallis and Carley undertook an independent, prospective study comparing the performance of the PTT, CareFlight, START and JumpSTART amongst 3461 children (≤12 years) attending a South African trauma centre within 12 hours of an acute injury[20]. They defined a Priority 1 (or T1) patient using Garner's criteria, modifying this for paediatric use by including fluid resuscitation in excess of 20ml/kg[20]. Additional end-points considered were ISS>15 and NISS>15. They found that CareFlight demonstrated the highest sensitivity (46%) for detecting the need for one of more of the modified Garner criteria, as well as ISS>15 (48.4%) and for NISS>15 (31.5%). PTT performed similarly (41.5%, 37.8% and 26.1% sensitivity, respectively). Both CareFlight and and PTT demonstrated over 98% sensitivity for all three outcomes, with the authors recommending either tool for use, stating that the differences in performance were unlikely to be clinically significant. In contrast, JumpSTART and START demonstrated worrying low sensitivities (<40%) for all outcomes, with JumpSTART exhibiting sensitivities of under 4% for all outcomes. A few years later, Price et al applied JumpSTART (≤8 years), START (>8 years), CareFlight, PTT and the Triage Sort to the data of 31,292 children (67.9% male, mean age 7.9 years, mortality 3.3%) taken from the UK TARN trauma registry in 2009, with imputation of missing pre-hospital physiology[70]. The authors reported tool sensitivity and specificity for in-hospital death and 1SS>15. Contrary to the findings of Wallis and Carley, the PTT had the worst performance amongst the tools, with sensitivity of only 37.8% for mortality in all <16s (25% in those ≤8 years) and 36.4% for ISS>15 (17.6% in those ≤8 years). Overall, there was great variation in tool performance across age subsets and for each outcome measured, and the authors were unable to recommend one tool as being superior the others. Each study has utilised a different age cut-off for defining a child, different inclusion criteria (ED attenders versus those meeting trauma registry inclusion criteria) and neither study presented a breakdown of the injury mechanisms in their cohorts. As such, it is still unclear which existing tool performs best amongst children.

In an attempt to standardise triage practices and increase interoperability between emergency services across the various regions in 2006, the United States Centre for Diseases Control and Prevention (CDC) and the National Association of EMS Physicians convened a working group of national experts[74]. They undertook an in-depth literature review, including the studies outlined above, followed by a Delphi consensus exercise involving several experts, taking into account lessons learnt from several international post-event evaluations[74]. Following this extensive exercise, they acknowledged that there was insufficient evidence to recommend one method of triage over another for use in across the United States. Instead, they compiled in a national guideline including a list of 24 recommended “Model Uniform Core Criteria” which regions should apply when selecting a triage tool. They introduced the SALT tool, which is the only tool which complies with all of these recommendations[80]. Notably, although MUCC represents the most comprehensive effort undertaken in the United States to develop common uniform criteria for major incident triage systems, the Federal Interagency Committee on EMS acknowledge that there is a lack of evidence regarding the impact on patients outcomes of using a MUCC compliant triage method versus a non-MUCC compliant tool[102].

Ahead of the 2012 UK Olympic games, Challen and Walter sought to retrospectively compared the performance of START, the Manchester Sieve (MIMMS Triage Sieve) and CareFlight using 166 patients injured during the 7th July 2005 London bombings[69]. The authors found that the tools performed identically, however they openly acknowledged that the amount of missing data seriously compromised their attempts to evaluate tool performance.

The largest study to date comparing the performance of triage tools was undertaken by Cross an Cicero, using 530,695 paediatric, adult and geriatric trauma patients from the US National Trauma Data Bank (NTBD)[100]. This compared the performance of six triage tools: START, Fire Department of New York (FDNY, a modification of START), CareFlight, Sacco Triage Method Score, and Unadjusted Sacco Score, as well as the ordinary Glasgow Coma Scale)[18]. The authors compared tool-assigned outcomes to mortality on discharge, reporting area under the receiver operator curve (AUC) data. Secondary outcomes included death in the emergency department, ventilator use, and hospital length of stay. The authors undertook subgroup analyses by age, trauma mechanism (blunt, penetrating and burns) and patient sex. The patient population included 84.8% blunt-injured patients, 11% who suffered penetrating trauma and 0.9% who had burns; the bulk of patients had non-severe injury with only 24.4% having ISS>15. Results demonstrated that the Sacco Triage Method predicted mortality most accurately (AUC 0.883, 95%CI 0.880-0.885) and that CareFlight was best amongst patients suffering burns (AUC 0.870, 95% CI 0.85-0.89), although CareFlight had a tendency to mistriage “salvageable” patients to the “Unsalvageable” (Dead/black) category (41% of these survived). START performed particularly well among burn (AUC 0.86, 95%CI 0.84-0.88) and penetrating trauma patients (AUC 0.926, 95%CI 0.922-0.930) but less well among patients with blunt trauma (AUC 0.826, 95%CI 0.822-0.829). With regard to the elderly, all the triage methods were markedly less accurate than in other age groups, with the authors highlighting this as an area requiring further research. The study was limited by the exclusion of a large proportion of patients due to missing data particularly in those aged 0-15 years; overall, only 29% of the eligible patients had complete data for analysis. Additionally, the methodology used to assign “true” triage categories was subjective and based on ISS, and not resource requirements. Not all trauma-receiving hospitals submit data to the NTDB, so the dataset is not nationally representative. Whilst the need for ventilation is particularly resource-heavy making this a useful outcome to report when considering major incidents, some of the secondary outcomes such as time spent in the Emergency Department may be less useful or relevant to inform major incident planning. Despite the limitations, this study was able to highlight the strengths and weaknesses of individual tools, however it was unable to determine a clear winner amongst them.

In a series of UK-based studies starting in 2017, Vassallo et al sought to compare the performance of their newly derived MPTT against the the MIMMS Triage Sieve and the NARU Triage Sieve, as well as START and Careflight[47]. A single primary outcome measure was used: this was the need for a life-saving intervention, determined by an author-led modified Delphi consensus of experts[81]. This end-point was used by the author to model their novel tool, the MPTT, and then used to compare the performance of their tool to others[47]. Other outcome measures such as mortality and injury severity were not reported. Comparative analyses were conducted amongst military and civilian patient populations[21,47]. Amongst patients from the UK military’s Joint Trauma Theatre Registry, the MPTT and its derivative, the MPTT-24 demonstrated the highest sensitivity (66.3 and 69.5%) but lowest specificity (65.3 and 69.9%) amongst the triage tools tested[47,52]. The Australian Careflight tool demonstrated the highest specificity (98.4%) and lowest over-triage rate (5%) although sensitivity was 32.9% with regards to the chosen outcome measure. This study was limited by the application of tools to patients' first recorded hospital physiology, as the JTTR does not reliably capture pre-hospital observations, hence results may have been influenced by treatments administered in the pre-hospital setting. 40% of patients in the JTTR database were excluded from the study due to missing data. Another limitation is that the patients were injured servicemen and women cared for under combat conditions, which has limited generalisability to major incidents affecting civilians. Vassallo et al went on to conduct a similar comparative analysis in a civilian population from the UK's TARN registry including over 100,000 patients[81]. Here, again, tools were retrospectively applied to patients' first hospital (rather than pre-hospital) physiological observations. The results demonstrated that the MPTT had the highest sensitivity of 57.6% and the lowest level of under-triage (42.4%) compared with other tools. The NARU Triage Sieve and the START tool were very similar in performance, with a sensitivity of 28% and a positive predictive value above 50%. Both these tools had a much lower over-triage rate of nearly half compared with 67% seen with the MPTT; the authors described this as an acceptable rate of over-triage[47]. In 2019, the authors went on to compare the MPTT-24 and NARU Triage Sieve to the Triage Sort (MIMMS), a secondary triage tool by retrospective application to TARN registry data, using the need for life saving intervention (LSI) as the outcome measure [103]. The Triage Sort demonstrated the lowest accuracy of all triage tools at identifying the need for LSI (sensitivity 15.7% (95% CI 15.2 to 16.2) correlating with the highest rate of under-triage (84.3% (95% CI 83.8 to 84.8), but it had the greatest specificity (98.7% (95% CI 98.6 to 98.8).

In 2020, the authors who developed the consensus-based triage category definitions[101] (as well as the SALT tool) conducted a prospective study using a small convenience sample of 125 ED attenders at a US-based level 1 trauma centre, to test the triage accuracy of SALT, START, MIMMS Triage Sieve and CareFlight[88]. The primary end-point measure was the accuracy of each tool in assigning all (not just P1) triage categories. SALT was found to have the highest accuracy rate (52%; 95% CI 43–60) compared to START (36%; 95% CI 28–44), CareFlight (36%; 95% CI 28–44), and TriageSieve (37%; 95% CI 28–45). SALT also had the lowest under-triage rate (26%; 95% CI 19–34) compared to START

(57%; 95% CI 48–66), CareFlight (58%; 95% CI 49–66), and Triage Sieve (58%; 95% CI 49–66). SALT had the highest over-triage rate (22%; 95% CI 14–29) compared to START (7%; 95% CI 3–12), CareFlight (6%; 95% CI 2–11) and TriageSieve (6%; 95% CI 2–11). This study had several limitations, the most significant ones being that it was based on a very small convenience sample of patients that contained only 5 patients who fulfilled criteria for Priority 1 status, and that it was conducted by the authors that developed the SALT tool. The study highlighted that all four systems had relatively high rates of under-triage. The same authors had previously used identical methodology to compare the performance of SALT, JumpSTART, Triage Sieve, and CareFlight amongst 115 children (<18 years), however results were inconclusive because the confidence intervals for both the accuracy and under-triage rates overlapped between all triage tools, with high rates of under-triage noted amongst all tools[88].

Studies comparing the performance of multiple tools vary in their choice of metrics used to report performance of tools. Triage tools are essentially diagnostic tools which seek to assign the correct triage category to individual patients. For reasons outlined previously, diagnosis of the P1 category is widely considered the most important and this is the focus of several studies outlined in the previous section. Like any diagnostic tool, a number of measures of accuracy exist including sensitivity, specificity, under- and over-triage (1-sensitivity), area under the receiver operating curve (AUC) as well as both positive and negative predictive values (PPV, NPV). Policymakers must be clear of the implications on likely patient outcomes as well as resource considerations when selecting triage tools, taking account both sensitivity and specificity, but also the resilience of the system in question and its ability to deal with over-triage.

Sensitivity describes the ability of a diagnostic test to correctly identify patients who have a condition, also referred to as the true positive rate. Mathematically, sensitivity is calculated as the number of true positive patients divided by the sum of the number of true positives and the number of false negatives. However, sensitivity does not take into account the false positives. If a test has a high sensitivity for P1, then a negative result will be useful for "ruling out" the disease. In the case of a triage tool, one with a high sensitivity is likely to assign a P1 label to a high proportion of patients who are truly P1, however this comes at the likely expense of a degree of over-triage, or assignment of P1 category to patients who are non-P1. Specificity, on the other hand, describes a test's ability to assign a negative result to patients who do not have a condition. Mathematically, specificity is equal to the number of true negatives divided by the sum of the number of true negatives and the number of false positives. A triage tool with a high specificity for P1 status will correctly assign non-P1 status to nearly all the non-P1 patients, but this is not particularly useful for "ruling out" P1 status. Hence, a tool with high specificity might fail to assign P1 status to several true P1 patients, with a significant rate of under-triage. The ideal diagnostic tool, or indeed triage tool, has a very high sensivity and high specificity, however in reality one often comes at the expense of the other. A suitable balance must be sought between correctly identifying all high-acuity patients (under triage refers to those high-acuity patients who fail to be correctly identified) and inadvertently assigning high acuity status to less severely injured patients (over-triage). No tools have been shown to eliminate under-triage or over-triage completely and notably, there are no national or international standards governing the performance of triage tools in the MI setting. Several studies reporting the performance of triage tools have used trauma registry populations with specific inclusion criteria which give rise to a biased population with over-representation of severely injuried patients. The "denominator" of non-P1 patients might not be representative of true MI populations, and the degree of false positives in practice may be greater than that reported by such studies.

A receiver operating character curve (ROC curve) is a graphical plot in which the x axis consists of the false positive rate (1-specificity) and the y axis represents the true positive rate i.e. sensitivity. The area under the curve (AUC) provides users with an appreciation of how well a model or diagnostic test can produce relative scores to discriminate between positive or negative instances across all classification thresholds. This amalgamates the results of sensitivity and specificity at any given threshold in a visually appealing graphic and the AUC between different tools may be assessed and compared in this way. Notably, some studies such as Cicero et al report AUC values in isolation, which may limit policymakers in that individual tool sensitivity and specificity is not expressed, hence modelling and prediction of patient outcomes may prove challenging.

In selecting tools for a particular setting, policy makers must also take into account the resources available to treat patients and the number of patients that are likely to be affected in major incidents. There is a point at which healthcare facilities may be overwhelmed to the degree that individual patient care and therefore overall survival and outcomes are compromised. In an analysis of terrorist bombings between 1969 and 2004, Frykberg demonstrated a direct linear correlation between over-triage and critical mortality. The degree of over-triage that a system can deal with will also be affected by the setting, the number of victims arriving at healthcare facilities (which is also influenced by the inherent performance and correct practical application of pre-hospital triage tools) and the timing of major incidents. For example, individual hospitals and regional ambulance services may be more easily overwhelmed outside of working hours (e.g. overnight, at the weekend) and the resilience of hospitals may vary with the time of year, for example summer months versus the height of the so-called "winter pressures" experienced within the NHS[29]. Paediatric major trauma centres, which are geographically dispersed, are particularly vulnerable to being overwhelmed during major incidents, particularly when coupled with the human factors tendency to over-triage children, as observed in several post-event evaluations[10,22]. In a high-resource setting and one in which a high level of scrutiny is likely to be cast upon triage decisions and outcomes, sensitivity appears be a higher priority. The under-triage of children, for example, is likely to be heavily criticised by the general public and expectations of outcome following MI have never been so high. This was exemplified by the Manchester Arena Enquiry[27].

Therefore, in the relatively well-resourced setting of the UK, it can be proposed that the most important aspect of triage tool performance is the correct identification of Priority 1 patients, i.e. sensitivity. Inevitably, a considerable degree of over-triage is to be expected. In a low-resource setting where there is particular vulnerability to over-triage, specificity of a pre-hospital triage tool is likely to be the most important performance metric to consider, in order to maximise overall survival. Lastly, it is challenging to determine whether underperformance of a triage tool in major incidents arises from poor discriminatory capability of the tool itself or whether the tools have been applied incorrectly. Incorrect application can result from human error, or due to lack of simplicity of the tool e.g. those involving arithmetic are more likely to be abandoned or applied incorrectly[8]. It is apparent that large, registry based studies are valuable in determining the inherent physiological discriminatory capability of tools, but real-life event evaluations as well as simulation studies can give an idea of whether the tools are practically applicable by humans under challenging conditions. Policy makers must take both aspects into account when selecting tools for use, since overall success requires a tool to possess both of these essential qualities.

Summary of major incident triage tools and their performance as relevant to contemporary UK practice

Multiple major incident triage tools exist, including several adult primary triage tools START, mSTART, MIMMS Triage Sieve, NARU Triage Sieve, MPTT, MPTT-24, SALT, ASAV, CareFlight, BCD Triage Sieve, two paediatric-specific primary triage tools (JumpSTART and PTT) as well as two points-based triage tools, the MIMMS Triage Sort (a secondary triage tool) and the Sacco Triage Method[11,47,52]. There are several challenges in comparing the performance of existing tools. The unpredictable nature of major incidents makes it challenging to plan prospective studies in the major incident setting, with only a single study of this design performed to date[6]. Retrospective studies using data gathered on actual major incident patients are often severely limited by a lack of pre-hospital data[75]. Other studies have used surrogate populations, such as ED attenders in trauma receiving hospitals, in which case triage is performed by a single or few hospital-based practitioners without the challenging atmosphere of a major incident setting[56].

A key limiting factor in comparing the performance of tools is the variation in end-points used: controversies exist regarding the definition of a P1 casualty in academic terms. Studies evaluating the performance of major incident triage tools have used mortality, injury severity scores (namely, ISS>15 and NISS>15) and the need for various, sometimes subjective, life-saving interventions as the primary end-point measure. None of the end-points identified have been validated as outcome measures for triage tool performance and several are author-led[81]. Variations in the end-points have limited the ability to directly compare the results of studies. A validated, universally accepted endpoint would permit direct comparison between studies; this is an urgent research requirement. Such a standard may facilitate improvements in patient outcomes including overall survival. Garner's criteria to define P1 status appear to have gained popularity in multiple international studies, however these include some outdated medical practices: intracranial pressure measurements are now known not to correlate as well with neurological outcome as previously thought[19]. Additionally, large volume crystalloid resuscitation has given way to the preferential use of blood products and a growing tendency to hypotensive resuscitation, hence Garner's criterion of large-volume crystalloid resuscitation is less applicable in the contemporary setting19. Amongst the end-points in the existing literature, Lerner's criteria, although not currently validated for use as an outcome measure, appear to be the most scientifically objective, defining not just P1 but all triage categories, they are applicable to patients of all ages and incorporate a comprehensive range of injury mechanisms including burns and CBRN[101].

Amongst the tools identified, START has been studied the most and is the only tool to have undergone prospective validation in an actual major incident setting[6]. START has demonstrated sensitivity of 57% to 100% across various studies using a variety of primary endpoint measures, it has consistently demonstrated specificity of over 67%[6,11,17,18,52]. The MSTART is very similar to START, although their triage accuracy relative to one another has yet to be assessed. Lauded for its simplicity, CareFlight is the second most studied tool, and has demonstrated moderate performance in adults and children across a range of injury mechanisms, including blast and burns trauma[10,11,18,20,47,55]. The novel RAMP tool appears to be very similar to CareFlight and has only been evaluated by the authors, who found that it is much quicker and more accurate to apply than START[10]. Further independent evaluation of RAMP is warranted. The novel US SALT tool and German ASAV have yet to undergo independent evaluation, however the subjective nature of the assessments they include makes both tools very difficult to apply retrospectively[67,80]. MPTT and MPTT-24 are novel tools evaluated using a single author-defined end-point, which was used as a basis for modelling these tools. The authors recommend that the MPTT-24 replaces the NARU Triage Sieve in current UK civillian practice based on a modestly higher sensitivity for predicting the need for life-saving interventions[52]. However, their results suggest that the MPTT-24's rate of over-triage is higher and specificity lower than most other tools, which has implications in practice, including the potential to overwhelm medical facilities with patients labelled P1 following a major incident[24,47,52]. Independent evaluation of the performance of the MPTT-24 tools is warranted alongside the current UK NARU Triage Sieve, including the use of additional, objective outcome measures as relevant to the major incident setting. To date, no published evaluations of the performance of the UK military's Battlefield Casualty Drills Triage Sieve exist, although the tool has been used extensively during UK military operations[68].

The two point-based triage tools, Triage Sort and Sacco Triage Method, have both been modelled to predict mortality rather than resource requirements in the major incident setting[39,60]. Vassallo et al demonstrated that Triage Sort, the only secondary triage tool identified in the literature, had the lowest accuracy of all triage tools at identifying the need for life-saving interventions with sensitivity of only 15.7% (95% CI 15.2 to 16.2) correlating with the highest rate of under-triage of 84.3% (95% CI 83.8 to 84.8)[45]. As such, there is an urgent requirement to develop a novel secondary triage tool that offers refinement in performance, namely an increase in sensitivity and a decrease in overtriage of the P1 category, compared with existing primary triage tools. The proprietary nature of STM coupled with the limited improvements in performance relative to the widely utilised START tool in the US has limited its uptake into practice[11,60,71]. STM is also more complex to apply in practice with a heavy reliance on mathematical calculation by the care provider, making this a less attractive choice for the challenging major incident setting. Whilst both points-based tools are paper-based, the popularity of hand-held devices and personal mobile phones creates the possibility of developing tools that take the form of a portable device application once patients have moved on from the often dangerous and unpredictable scene of a major incident (i.e. secondary triage).

Regarding the performance of triage tools amongst children, two sizeable studies undertaken in paediatric populations have reported conflicting outcomes regarding the triage accuracy of the JumpSTART and PTT tools[20,69,70]. CareFlight has been commended for its accuracy in children, and there are practical and logistic advantages of using a tool which is applicable to both adults and children; this also simplifies the training delivered to first responders[55,74]. It is unclear which paediatric triage tool would perform best in the contemporary UK major incident setting, warranting further research in this area. At the other extreme of age, only one US-based study has evaluated the performance of triage tools amongst elders, reporting that all tools performed less accurately amongst elders compared to their younger adult counterparts[100]. Those aged 65 years and over form 18.3% of the current UK population and this proportion is likely to grow further[58]. Major incident planning must cater for the unique needs of this subset of our population, including selection of triage tools which perform well in this age group.

Several studies have been based outside of the UK and the degree to which their findings are generalizable to a UK setting are hampered by differences in population demographics and differences in the delivery of healthcare in those settings[18,20]. Previous UK-based studies have utilised patient outcome data that preceded a major change in the delivery of trauma care in 2012, when care was reorganised into inclusive, regionalised trauma systems[21,43,70]. Of the two UK (TARN) registry-based studies, one involved children cared for prior to 2009 and utilised only mortality and ISS>15 as outcome measures[70]. The other was an author-led validation of a novel tool (MPTT-24) alongside others, using patients captured by TARN from 2006 to 2014, and utilised a single author-defined end-point; a study design which may be prone to author bias[21]. Hence, the generalisability of their findings to the contemporary UK major incident setting is limited, warranting further research using a contemporary database of UK patients. Furthermore, following the regionalisation of trauma care, casualty distribution plans stipulate that P1 casualties would be conveyed to a major trauma centre, P2 casualties to a Trauma Unit whilst P3 casualties may be conveyed to minor injury units or EDs at a distance from the major incident[3,45]. However, there has been no evaluation of whether the patient resource requirements of each triage category align well with the capabilities of the corresponding tier of trauma care[21,28,69].

Given the existing evidence, it is difficult to determine superiority of one tool above the others, with reference to both UK-based and international settings[11,80]. The use of trauma registry data has facilitated the retrospective application of triage tools in large numbers of patients, sometimes including valuable subgroup analyses by age and injury mechanism, as demonstrated by Cross and Cicero's large, US-based registry study[18]. Two high-quality trauma registries exist in the UK, namely TARN and JTTR, capturing large numbers of injured patients, with the potential to form the basis of a comparative analysis of the performance of multiple triage tools[43,47]. Mandatory subscription to TARN by all trauma-receiving hospitals renders the resulting TARN population a nationally representative sample of injured patients, the largest of its kind in Europe[43]. Although JTTR largely constitutes injured young, male soldiers cared for in military medical treatment facilities, the predominance of blast and penetrating trauma in this cohort offers the opportunity to examine the performance of any novel or existing tools in patients with these mechanisms of injury. This is highly relevant since, in the last 15 years, man-made blast and bladed mechanisms have predominated amongst UK major incidents[29]. Hence, whilst the choice of triage tools for use in the UK should cater for any eventuality, it is particularly important that they perform well amongst patients suffering blast and penetrating trauma, warranting evaluation of the performance of triage tools in UK patients with these mechanisms.

Machine learning and its potential to improve triage

In the last fifty years, improvements in the power of computer processing and analytic ability has led to marked technological advances[104]. Additionally, the amount of routinely collected data that is available for analysis has grown[105]. This has been associated with the development of artificial intelligence (AI). There has yet to be a widely accepted definition of AI within data science and computing, partly due to the rapidly evolving nature of AI. However, from a clinician's perspective, AI can be described as "human-like intelligence displayed by a machine, in terms of learning from experience or observations, then using this knowledge to recognise, interpret and make autonomous actions when faced with similar situations[106]. AI is present in several aspects of everyday life, from entertainment suggestions on social media to email spam filters and even self-driving vehicles[106]. A natural expansion into healthcare is to be expected[104]. Machine learning (ML) is commonly viewed as a subset of AI. Alan Turing famously distinguished between the two by proposing that the cognitive question "Can machines think?" (AI) be replaced with the more operational question "Can machines do what we (as thinking entities) can do?"[107]. In 1959, machine learning was described as a "field of study that gives computers the ability to learn without being specifically programmed." In practice, machine learning explores the construction and study of algorithms that can "learn" or make predictions on data to make predictions, without following strictly static computer programming instructions. Machine learning can be broadly classified further based on the nature of the "feedback" available to the learning system. In supervised learning, the computer is presented with example inputs and the desired outputs with the goal of learning a general rule that maps inputs to outputs. In unsupervised learning, no labels are given to the learning algorithm, leaving it on its own to determine structure in its input. Unsupervised learning can help to reveal hidden patterns in data or a means to an end.

Machine learning has several advantages over traditional statistics, for example by using algorithms to identify previously unrecognised input variables that predict outcome, unlike the requirement for human-determined variables in traditional statistics[106]. Additionally, non-linear relationships between variables and outcomes can be harnessed, making machine learning particularly useful in identifying complex or unclear data inter-relationships. Hence, machine learning models excel in exploratory predictive (regression or classification) modelling using large or computationally weighty datasets. By comparison, traditional statistical methods are perhaps better suited to confirming specific hypotheses in smaller datasets.

In healthcare, machine learning has the potential to impact and progress patient triage, diagnosis of disease, prediction of patient outcomes and estimation of prognosis. This is made possible by the large amounts of data collected routinely by individual healthcare facilities and multicentre regional or national disease registries by regulatory bodies[34,105,106]. These large databases often include patient variables and outcome data, lending themselves to the application of AI and ML. However, despite the potential advantages outlined earlier, there have been barriers to adopting machine learning-based findings into healthcare, such as worries about unstructured reporting, unsuitable algorithm selection, concerns regarding data privacy and bias related to the use of proxy variables [105,108,109]. The majority of studies utilise retrospective data, which were collected for reasons other than the AI/ML applications, and this also raises concerns about the use of data retrospectively without prior patient consent. A further reason for hesitance amongst clinicians of embracing ML-based studies and their findings is a lack of understanding of when AI/ML based models are likely to be superior to those derived from traditional statistical methods: ML has not always been shown to be superior to traditional logistic regression[109]. Notably, it is not always the case that models undergo external validation, which is an essential step in their evaluation prior to considering their adoption into clinical practice[106]. The confusing use of terminology, for example the interchangeable use of AI and ML, can also baffle clinicians, patients and regulators alike. There has also been fear amongst clinicians of being replaced with AI and ML-based technologies, however comparison between clinicians and ML is not immediately relevant: AI/ML has the potential to reduce cognitive load, enable the efficient use of clinician time and supplement or improve clinical care by combining the best of human and machine abilities[106]. Additionally, ML clinical studies can involve deep neural networks or other complex algorithms and the way that the models arrive at their conclusion cannot easily be visualised by clinicians; this lack of transparency is a barrier to the acceptance of such models by clinicians as well as patients and regulators[105,108].

Tree-based machine learning models have demonstrated utility in clinical risk stratification, with the ability to capture non-linear interactions between input variables[106]. In fact, the binary (present/absent or yes/no) nature of decision trees (also known as Recursive Partitioning and Regression Trees, RPART) closely resembles the "flowsheet" format of existing MI triage tools, making this ideally suited to developing paper-based primary triage tools. Both random forest (RF) and gradient boosted tree (XGB) are popular machine learning algorithms with strong predictive power. RF is based on the concept of bagging, which lowers the prediction variance[104]. Furthermore, instead of growing each tree using all variables, it randomly chooses a subset of variables at each split of the node in the tree, thereby forcing it to learn through all subsets of available variables. For XGB, the prediction target is estimated by sum-of-trees, and the model is built by successively fitting each tree to the residue of previously fitted trees, while regularizing the fit through multiplication by a scaling factor known as learning rate. In short, XGB estimates the target function by a sum of trees each of which explains a small and different portion of the target and no single tree dominates the prediction. Both RF and XGB methods lend themselves to the development of an app-based MI triage tool, perhaps more suited to the secondary triage setting where reliance on a portable device is more feasible.

Although there is a paucity of databases encompassing data from major incident victims, large trauma registry databases exist, encompassing a broad range of patient demographics and injury mechanisms. There is potential for machine learning applied to these large databases to result in models which outperform existing tools in predicting the need for life-saving interventions in injured patients, particularly as previous triage tools are largely based on expert opinion. Hence, there is scope to develop machine-learning models that can be adapted into primary and secondary MI triage tools. Crucially, any such models must be externally validated using an independent population of injured patients prior to recommending these for practical use in MIs.

Hypothesis and aims of thesis

Major incidents are characterised by the mismatch between the immediate needs of patients and the resources available to treat them. Triage tools have a crucial role in ensuring that finite healthcare resources are allocated to maximise overall survival. Existing tools are are largely paper-based and have been developed based on expert opinion. A major challenge in furthering the science of major incident management is the lack of consensus and standardisation of end-points which best define Priority 1 status and other triage categories.

This thesis is focussed on the hypothesis that current methods of major incident triage in the UK civilian setting are suboptimal.

In view of this hypothesis, the specific aims of this thesis are:

1. To determine an objective, practically applicable and resource-based system to define major incident triage categories, and to validate this using a nationally representative sample of adults and children
2. To compare the performance of existing international tools in adults, older adults and children across injuries of all mechanisms using a nationally representative sample of injured patients, in order to determine the best performing existing tool
3. To identify the best performing existing triage tool for use in the most prevalent type of UK major incident, i.e. terrorist-related major incidents, which are characterised by blast and penetrating trauma

4. To use machine learning to develop novel primary and secondary triage tools for the UK civilian setting and to validate these externally using a geographically distinct population of injured patients

References

1. Joint Emergency Service Interoperability Programme. JESIP Joint Doctrine: The Interoperability Framework Edition 3. 2022. https://www.jesip.org.uk/news/joint-doctrine-edition-3/.

2. World Health Organisation. Mass Casualty Management Systems: Strategies and guidelines for building health sector capacity. In: Prevention HAiCIV, editor.; 2007.

3. NHS England. NHS Clinical Guidelines for Major Incidents and Mass Casualty Events (Version 2). 2020.

4. American College of Surgeons Committee on Trauma. Advanced Trauma Life Support: Student course manual (10th Edition). Chicago, IL; 2018.

5. Advanced Life Support Group. Major Incident Medical Management and Support: The Practical Approach at the Scene, 3rd Edition: BMJ Books; 2011.

6. Kahn CA, Schultz CH, Miller KT, Anderson CL. Does START triage work? An outcomes assessment after a disaster. Annals of emergency medicine 2009; 54(3): 424-30, 30.e1.

7. Debacker M, Hubloue I, Dhondt E, et al. Utstein-style template for uniform data reporting of acute medical response in disasters. PLoS Curr 2012; 4: e4f6cf3e8df15a-e4f6cf3e8dfa.

8. Asaeda G. The day that the START triage system came to a STOP: observations from the World Trade Center disaster. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2002; 9(3): 255-6.

9. Perkins ZB, De'Ath HD, Aylwin C, Brohi K, Walsh M, Tai NR. Epidemiology and outcome of vascular trauma at a British Major Trauma Centre. Eur J Vasc Endovasc Surg 2012; 44(2): 203-9.

10. Pepper M, Archer F, Moloney J. Triage in Complex, Coordinated Terrorist Attacks. Prehospital and disaster medicine 2019; 34(4): 442-8.

11. Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehospital and disaster medicine 2008; 23(1): 3-8.

12. Nakao H, Ukai I, Kotani J. A review of the history of the origin of triage from a disaster medicine perspective. Acute medicine & surgery 2017; 4(4): 379-84.

13. Bricknell M. For debate: the Operational Patient Care Pathway. Journal of the Royal Army Medical Corps 2014; 160(1): 64.

14. Hettiaratchy S, Tai N, Mahoney P, Hodgetts T. UK's NHS trauma systems: lessons from military experience. The Lancet 2010; 376(9736): 149-51.

15. Postma IL, Weel H, Heetveld MJ, et al. Mass casualty triage after an airplane crash near Amsterdam. Injury 2013; 44(8): 1061-7.

16. Cone DC, Serra J, Burns K, MacMillan DS, Kurland L, Van Gelder C. Pilot test of the SALT mass casualty triage system. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2009; 13(4): 536-40.

17. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the Accuracy of Mass Casualty Triage Systems When Used in an Adult Population. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2019: 1-8.

18. Cross KP, Cicero MX. Head-to-Head Comparison of Disaster Triage Methods in Pediatric, Adult, and Geriatric Patients. Annals of emergency medicine 2013; 61(6): 668-76.e7.

19. Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. Annals of emergency medicine 2001; 38(5): 541-8.

20. Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. Emergency Medicine Journal 2006; 23(6): 475.

21. Vassallo J, Smith J, Bouamra O, Lecky F, Wallis LA. The civilian validation of the Modified Physiological Triage Tool (MPTT): an evidence-based approach to primary major incident triage. Emergency Medicine Journal 2017; 34(12): 810.

22. Turner CD, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Crit Care 2016; 20(1): 362.

23. Moran CG, Lecky F, Bouamra O, et al. Changing the System - Major Trauma Patients and Their Outcomes in the NHS (England) 2008–17. EClinicalMedicine 2018; 2: 13-21.

24. Frykberg ER, Tepas JJ, 3rd. Terrorist bombings. Lessons learned from Belfast to Beirut. Annals of surgery 1988; 208(5): 569-76.

25. American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient, 2014.

26. Stoke B. World's largest-ever explosion (almost). BBC Stoke. 2008.

27. Craigie RJ, Farrelly PJ, Santos R, Smith SR, Pollard JS, Jones DJ. Manchester Arena bombing: lessons learnt from a mass casualty incident. BMJ Military Health 2020; 166(2): 72-5.

28. Aylwin CJ, König TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. The Lancet 2006; 368(9554): 2219-25.

29. Moran CG, Webb C, Brohi K, Smith M, Willett K. Lessons in planning from mass casualty events in UK. Bmj 2017; 359: j4765.

30. Edwards RD. The Birmingham bombings 40 years on: what can we learn from IRA terror? The Telegraph. 2014.

31. King R. Detonation: Rebirth of a City; 2006.

32. Turner CDA, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Critical Care 2016; 20(1): 362.

33. BBC. Mass shootings and gun control. BBC. 2010.

34. Office of National Statistics. Police force area tables, crime in England & Wales, year ending March 2019, 2019.

35. Olding J, Olding C, Bew D, Fan K. Penetrating head & neck trauma - Epidemiology and injury characteristics in terror-related violence, interpersonal violence and deliberate self-harm at a level 1 trauma centre. Surgeon 2019; 17(3): 133-8.

36. Christensen MC, Nielsen TG, Ridley S, Lecky FE, Morris S. Outcomes and costs of penetrating trauma injury in England and Wales. Injury 2008; 39(9): 1013-25.

1. Programme JESI. JESIP Joint Doctrine: The Interoperability Framework Edition 3. 2022. https://www.jesip.org.uk/news/joint-doctrine-edition-3/.

2. Organisation WH. Mass Casualty Management Systems: Strategies and guidelines for building health sector capacity. In: Prevention HAiCIV, editor.; 2007.

3. England N. NHS Clinical Guidelines for Major Incidents and Mass Casualty Events (Version 2). 2020.

4. Trauma ACoSCo. Advanced Trauma Life Support: Student course manual (10th Edition). Chicago, IL; 2018.

5. Group ALS. Major Incident Medical Management and Support: The Practical Approach at the Scene, 3rd Edition: BMJ Books; 2011.

6. Kahn CA, Schultz CH, Miller KT, Anderson CL. Does START triage work? An outcomes assessment after a disaster. Annals of emergency medicine 2009; 54(3): 424-30, 30.e1.

7. Debacker M, Hubloue I, Dhondt E, et al. Utstein-style template for uniform data reporting of acute medical response in disasters. PLoS Curr 2012; 4: e4f6cf3e8df15a-e4f6cf3e8dfa.

8. Asaeda G. The day that the START triage system came to a STOP: observations from the World Trade Center disaster. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2002; 9(3): 255-6.

9. Perkins ZB, De'Ath HD, Aylwin C, Brohi K, Walsh M, Tai NR. Epidemiology and outcome of vascular trauma at a British Major Trauma Centre. Eur J Vasc Endovasc Surg 2012; 44(2): 203-9.

10. Pepper M, Archer F, Moloney J. Triage in Complex, Coordinated Terrorist Attacks. Prehospital and disaster medicine 2019; 34(4): 442-8.

11. Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehospital and disaster medicine 2008; 23(1): 3-8.

12. Nakao H, Ukai I, Kotani J. A review of the history of the origin of triage from a disaster medicine perspective. Acute medicine & surgery 2017; 4(4): 379-84.

13. Bricknell M. For debate: the Operational Patient Care Pathway. Journal of the Royal Army Medical Corps 2014; 160(1): 64.

14. Hettiaratchy S, Tai N, Mahoney P, Hodgetts T. UK's NHS trauma systems: lessons from military experience. The Lancet 2010; 376(9736): 149-51.

15. Postma IL, Weel H, Heetveld MJ, et al. Mass casualty triage after an airplane crash near Amsterdam. Injury 2013; 44(8): 1061-7.

16. Cone DC, Serra J, Burns K, MacMillan DS, Kurland L, Van Gelder C. Pilot test of the SALT mass casualty triage system. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2009; 13(4): 536-40.

17. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the Accuracy of Mass Casualty Triage Systems When Used in an Adult Population. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2019: 1-8.

18. Cross KP, Cicero MX. Head-to-Head Comparison of Disaster Triage Methods in Pediatric, Adult, and Geriatric Patients. Annals of emergency medicine 2013; 61(6): 668-76.e7.

19. Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. Annals of emergency medicine 2001; 38(5): 541-8.

20. Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. Emergency Medicine Journal 2006; 23(6): 475.

21. Vassallo J, Smith J, Bouamra O, Lecky F, Wallis LA. The civilian validation of the Modified Physiological Triage Tool (MPTT): an evidence-based approach to primary major incident triage. Emergency Medicine Journal 2017; 34(12): 810.

22. Turner CD, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Crit Care 2016; 20(1): 362.

23. Moran CG, Lecky F, Bouamra O, et al. Changing the System - Major Trauma Patients and Their Outcomes in the NHS (England) 2008–17. EClinicalMedicine 2018; 2: 13-21.

24. Frykberg ER, Tepas JJ, 3rd. Terrorist bombings. Lessons learned from Belfast to Beirut. Annals of surgery 1988; 208(5): 569-76.

25. Trauma ACoSCo. Resources for Optimal Care of the Injured Patient, 2014.

26. Stoke B. World's largest-ever explosion (almost). BBC Stoke. 2008.

27. Craigie RJ, Farrelly PJ, Santos R, Smith SR, Pollard JS, Jones DJ. Manchester Arena bombing: lessons learnt from a mass casualty incident. BMJ Military Health 2020; 166(2): 72-5.

28. Aylwin CJ, König TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. The Lancet 2006; 368(9554): 2219-25.

29. Moran CG, Webb C, Brohi K, Smith M, Willett K. Lessons in planning from mass casualty events in UK. Bmj 2017; 359: j4765.

30. Edwards RD. The Birmingham bombings 40 years on: what can we learn from IRA terror? The Telegraph. 2014.

31. King R. Detonation: Rebirth of a City; 2006.

32. Turner CDA, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Critical Care 2016; 20(1): 362.

33. BBC. Mass shootings and gun control. BBC. 2010.

34. Statistics OoN. Police force area tables, crime in England & Wales, year ending March 2019, 2019.

35. Olding J, Olding C, Bew D, Fan K. Penetrating head & neck trauma - Epidemiology and injury characteristics in terror-related violence, interpersonal violence and deliberate self-harm at a level 1 trauma centre. Surgeon 2019; 17(3): 133-8.

36. Christensen MC, Nielsen TG, Ridley S, Lecky FE, Morris S. Outcomes and costs of penetrating trauma injury in England and Wales. Injury 2008; 39(9): 1013-25.

37. Vale JA, Marrs TO, Maynard RC. Novichok: a murderous nerve agent attack in the UK. Clin Toxicol (Phila) 2018; 56(11): 1093-7.

38. Police TM. Grenfell Tower fire investigation. 2017.

39. (NARU) NARU. Clinical Guidance: Medical Support Minimal Requirements for a Mass Casualty Incident. www.naru.org.uk; 2014.

40. (NCEPOD) NEiPD. Trauma, who cares? NCEPOD 2007.

41. Gabbe BJ, Simpson PM, Sutherland AM, et al. Improved functional outcomes for major trauma patients in a regionalized, inclusive trauma system. Annals of surgery 2012; 255(6): 1009-15.

42. Nathens AB, Jurkovich GJ, Rivara FP, Maier RV. Effectiveness of State Trauma Systems in Reducing Injury-Related Mortality: A National Evaluation. Journal of Trauma and Acute Care Surgery 2000; 48(1): 25.

43. Manchester) TTUo. The Trauma Audit and Research Network. Continuous. www.tarn.ac.uk.

44. Malik NS, Chernbumroong S, Xu Y, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. EClinicalMedicine.

45. Vassallo J, Smith J. Major incident triage and the evaluation of the Triage Sort as a secondary triage method. Emergency medicine journal : EMJ 2019; 36(5): 281-6.

46. Eastridge BJ, Mabry RL, Blackbourne LH, Butler FK. We don't know what we don't know: prehospital data in combat casualty care. US Army Med Dep J 2011: 11-4.

47. Vassallo J, Beavis J, Smith JE, Wallis LA. Major incident triage: Derivation and comparative analysis of the Modified Physiological Triage Tool (MPTT). Injury 2017; 48(5): 992-9.

48. Allcock EC, Woolley T, Doughty H, Midwinter M, Mahoney PF, Mackenzie I. The clinical outcome of UK military personnel who received a massive transfusion in Afghanistan during 2009. J R Army Med Corps 2011; 157(4): 365-9.

49. O'Connell KM, Littleton-Kearney MT, Bridges E, Bibb SC. Evaluating the Joint Theater Trauma Registry as a data source to benchmark casualty care. Mil Med 2012; 177(5): 546-52.

50. Palm K, Apodaca A, Spencer D, et al. Evaluation of military trauma system practices related to damage-control resuscitation. The journal of trauma and acute care surgery 2012; 73(6 Suppl 5): S459-64.

51. Newport Beach CFD. START Support Services. 1983.

52. Vassallo J, Horne S, Smith JE. Triage and the Modified Physiological Triage Tool-24 (MPTT-24). BMJ Mil Health 2020; 166(1): 33-6.

53. Vassallo J, Smith J. 2 Major incident triage and the implementation of a new triage tool, the MPTT-24. Emergency Medicine Journal 2017; 34(12): A860-A1.

54. Jerome D, Savage DW, Pietrosanu M. An assessment of mass casualty triage systems using the Alberta trauma registry. Cjem 2023; 25(8): 659-66.

55. Tran MD, Garner AA, Morrison I, Sharley PH, Griggs WM, Xavier C. The Bali bombing: civilian aeromedical evacuation. Med J Aust 2003; 179(7): 353-6.

56. Wallis LA, Carley S. Validation of the Paediatric Triage Tape. Emergency Medicine Journal 2006; 23(1): 47.

57. Romig LE. Pediatric triage. A system to JumpSTART your triage of young patients at MCIs. Jems 2002; 27(7): 52-8, 60-3.

58. Statistics OoN. Population estimates 2019. 2019.

59. Navin DM, Sacco WJ, McGill G. Application of a New Resource-Constrained Triage Method to Military-Age Victims. Military Medicine 2009; 174(12): 1247-55.

60. Sacco WJ, Navin DM, Fiedler KE, Waddell RK, 2nd, Long WB, Buckman RF, Jr. Precise formulation and evidence-based application of resource-constrained triage. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2005; 12(8): 759-70.

61. Banerjee J, Baxter M, Coats T, et al. Major Trauma in Older People. 2017.

62. Vassallo JM, Smith JE, Wallis LA. Investigating the effects of under-triage by existing major incident triage tools. European journal of emergency medicine : official journal of the European Society for Emergency Medicine 2019; 26(2): 139-44.

63. Trunkey DD. Trauma centers and trauma systems. Jama 2003; 289(12): 1566-7.

64. Wachira BW, Abdalla RO, Wallis LA. Westgate Shootings: An Emergency Department Approach to a Mass-casualty Incident. Prehospital and disaster medicine 2014; 29(5): 538-41.

65. Alarhayem AQ, Myers JG, Dent D, et al. Time is the enemy: Mortality in trauma patients with hemorrhage from torso injury occurs long before the “golden hour”. The American Journal of Surgery 2016; 212(6): 1101-5.

66. Baxt WG, Upenieks V. The lack of full correlation between the Injury Severity Score and the resource needs of injured patients. Annals of emergency medicine 1990; 19(12): 1396-400.

67. Lerner EB, Cone DC, Weinstein ES, et al. Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. Disaster medicine and public health preparedness 2011; 5(2): 129-37.

68. Hodgetts TJ, Hanlan CG, Newey CG. Battlefield First Aid: a simple, systematic approach for every soldier. J R Army Med Corps 1999; 145(2): 55-9.

69. Challen K, Walter D. Major incident triage: comparative validation using data from 7th July bombings. Injury 2013; 44(5): 629-33.

70. Price CL, Brace-McDonnell SJ, Stallard N, Bleetman A, Maconochie I, Perkins GD. Performance characteristics of five triage tools for major incidents involving traumatic injuries to children. Injury 2016; 47(5): 988-92.

71. Sacco WJ, Navin DM, Waddell RK, 2nd, Fiedler KE, Long WB, Buckman RF, Jr. A new resource-constrained triage method applied to victims of penetrating injury. The Journal of trauma 2007; 63(2): 316-25.

72. Jain T, Ragazzoni L, Stryhn H, Stratton S, Della Corte F. Comparison of the Sacco Triage Method Versus START Triage Using a Virtual Reality Scenario in Advance Care Paramedic Students. CJEM 2015; -1: 1-5.

73. Paul AO, Kay MV, Huppertz T, et al. [Validation of the prehospital mSTaRT triage algorithm. A pilot study for the development of a multicenter evaluation]. Der Unfallchirurg 2009; 112(1): 23-30, 2.

74. Guideline N, E. Brooke Lerner DCC, Eric S. Weinstein, Richard B. Schwartz, Phillip L. Coule, Michael Cronin, Ian S. Wedmore, Eileen M. Bulger, Deborah Ann Mulligan, Raymond E. Swienton, Scott M. Sasser, Umair A. Shah, Leonard J. Weireter, Teri L. Sanddal, Julio Lairet, David Markenson, Lou Romig, Gregg Lord, Jeffrey Salomone, Robert O'Connor and Richard C. Hunt. Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. Disaster medicine and public health preparedness 2011; 5: 129-37.

75. Model uniform core criteria for mass casualty triage. Disaster medicine and public health preparedness 2011; 5(2): 125-8.

76. Lee CW, McLeod SL, Peddle MB. First Responder Accuracy Using SALT after Brief Initial Training. Prehospital and disaster medicine 2015; 30(5): 447-51.

77. Eastridge BJ, Mabry RL, Seguin P, et al. Death on the battlefield (2001-2011): implications for the future of combat casualty care. The journal of trauma and acute care surgery 2012; 73(6 Suppl 5): S431-7.

78. Vassallo J, Smith J, Wallis L. Major incident triage and the implementation of a new triage tool, the MPTT-24. Journal of the Royal Army Medical Corps 2018; 164: 103-6.

79. Heller AR, Salvador N, Frank M, Schiffner J, Kipke R, Kleber C. [Diagnostic quality of triage algorithms for mass casualty incidents]. Anaesthesist 2017; 66(10): 762-72.

80. Wolf P, Bigalke M, Graf BM, Birkholz T, Dittmar MS. Evaluation of a novel algorithm for primary mass casualty triage by paramedics in a physician manned EMS system: a dummy based trial. Scand J Trauma Resusc Emerg Med 2014; 22: 50.

81. Vassallo J, Smith JE, Bruijns SR, Wallis LA. Major incident triage: A consensus based definition of the essential life-saving interventions during the definitive care phase of a major incident. Injury 2016; 47(9): 1898-902.

82. Marsden MER, Mossadegh S, Marsh W, Tai N. Development of a major incident triage tool: the importance of evidence from implementation studies. J R Army Med Corps 2018.

83. Eastridge BJ, Butler F, Wade CE, et al. Field triage score (FTS) in battlefield casualties: validation of a novel triage technique in a combat environment. Am J Surg 2010; 200(6): 724-7; discussion 7.

84. Meredith W, Rutledge R, Hansen AR, et al. Field triage of trauma patients based upon the ability to follow commands: a study in 29,573 injured patients. The Journal of trauma 1995; 38(1): 129-35.

85. Bennett A. Methodologies Utilized and Lessons Learned in High Threat Environments and Mass Casualty Environments. Journal of High Threat & Austere Medicine 2019: 1-7.

86. T J Hodgetts JH, I Maconochie, C Smart. Paediatric triage tape. Pre-hospital Immediate Care 1998; 2: 155-9.

87. Pollaris G, Sabbe M. Reverse triage: more than just another method. European journal of emergency medicine : official journal of the European Society for Emergency Medicine 2016; 23(4): 240-7.

88. Heffernan RW, Lerner EB, McKee CH, et al. Comparing the Accuracy of Mass Casualty Triage Systems in a Pediatric Population. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2019; 23(3): 304-8.

89. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the Accuracy of Mass Casualty Triage Systems When Used in an Adult Population. Prehospital Emergency Care 2019: 1-8.

90. Cicero MX, Santillanes GR, Cross KP, Kaji AH, Donofrio JJ. Prehospital Disaster Triage Does Not Predict Pediatric Outcomes: Comparing the Criteria Outcomes Tool to Three Mass-Casualty Incident Triage Algorithms. Prehospital and disaster medicine 2021; 36(5): 503-10.

91. Dittmar MS, Wolf P, Bigalke M, Graf BM, Birkholz T. Primary mass casualty incident triage: evidence for the benefit of yearly brief re-training from a simulation study. Scand J Trauma Resusc Emerg Med 2018; 26(1): 35.

92. Jain TN, Ragazzoni L, Stryhn H, Stratton SJ, Della Corte F. Comparison of the Sacco Triage Method Versus START Triage Using a Virtual Reality Scenario in Advance Care Paramedic Students. CJEM 2016; 18(4): 288-92.

93. Valdez C, Sarani B, Young H, Amdur R, Dunne J, Chawla LS. Timing of death after traumatic injury--a contemporary assessment of the temporal distribution of death. The Journal of surgical research 2016; 200(2): 604-9.

94. Demetriades D, Murray J, Charalambides K, et al. Trauma fatalities: time and location of hospital deaths. J Am Coll Surg 2004; 198(1): 20-6.

95. Heuer M, Hussmann B, Kaiser GM, Lefering R, Paul A, Lendemans S. Abdominal vascular trauma in 760 severely injured patients. Eur J Trauma Emerg Surg 2013; 39(1): 47-55.

96. Demetriades D, Kimbrell B, Salim A, et al. Trauma deaths in a mature urban trauma system: is "trimodal" distribution a valid concept? J Am Coll Surg 2005; 201(3): 343-8.

97. Vassallo J, Smith JE, Wallis LA. Major incident triage and the implementation of a new triage tool, the MPTT-24. 2018; 164(2): 103-6.

98. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. The Journal of trauma 1997; 43(6): 922-5; discussion 5-6.

99. Wallis L, Carley S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. Emergency Medicine Journal 2006; 23(4): 291.

100. Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. Annals of emergency medicine 2013; 61(6): 668-76.e7.

101. Lerner EB, McKee CH, Cady CE, et al. 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 2015; 19(2): 267-71.

102. EMS FICo. National Implementation of the Model Uniform Core Criteria for Mass Casualty Incident Triage: A report of the FICEMS. 2014.

103. Vassallo J, Smith J. Major incident triage and the evaluation of the Triage Sort as a secondary triage method. Emergency Medicine Journal 2019; 36(5): 281.

104. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digital Medicine 2020; 3(1): 118.

105. European Parliament EPRS, Scientific Foresight, Unit. The impact of the General Data Protection Regulation (GDPR) on artificial

intelligence: study 2020.

106. Ramlakhan S, Saatchi R, Sabir L, et al. Understanding and interpreting artificial intelligence, machine learning and deep learning in Emergency Medicine. Emergency Medicine Journal 2022; 39(5): 380.

107. Turing AM. Computing Machinery and Intelligence. Mind 1950; 49: 433-60.

108. Price WN. Big data and black-box medical algorithms. Sci Transl Med 2018; 10(471).

109. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019; 110: 12-22.

CHAPTER TWO: PERFORMANCE OF EXISTING TRIAGE TOOLS AND THE VALIDATION OF AN ENDPOINT TO MEASURE TOOL PERFORMANCE

Introduction to Chapter 2

Major incidents are characterised by a mismatch between the immediate needs of patients and the resources available to treat them. A review of the existing literature has revealed that in the UK, man-made major incidents comprising blast and penetrating trauma have occurred with increasing frequency in recent years and continue to constitute the greatest threat to national security. Triage tools have a crucial role in ensuring that finite healthcare resources are allocated to maximise overall survival. Existing UK and international tools have largely been developed based on expert opinion and implemented without formal validation. A major challenge in furthering the science of major incident management is the lack of consensus and standardisation of end-points which best define Priority 1 (P1) status and other triage categories. A number of systems exist to define P1, including ISS>15, mortality and several intervention-based criteria, including objective and evidence-based definitions described by Lerner et al.

Chapter 2 aims to validate an objective, practically applicable and resource-based system to define major incident triage categories using a nationally representative sample of adults and children. A further aim is to compare the performance of existing international tools in adults, older adults and children across injuries of all mechanisms using a nationally representative sample of injured patients, in order to determine the best performing existing tool.

Publication number 1: The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population

Author contributions:

NM conducted a literature review prior to the study. SC and YX contributed equally to this study. **NM**, JV, AB, DB, DK, MF and GVG designed the study. **NM**, SC and YX verified the underlying data and conducted analysis. All authors contributed to data interpretation. **NM** wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

EClinicalMedicine 36 (2021) 100888

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research paper

The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population

Nabeela S. Malik[a,b,c,*], Saisakul Chernbumroong[a,d], Yuanwei Xu[d], James Vassallo[e], Justine Lee[b,g,h], Douglas M. Bowley[b,f,g], Timothy Hodgetts[i], Christopher G Moran[h,j], Janet M Lord[a,b], Antonio Belli[a,b,g], Damian Keene[f,g], Mark Foster[a,e,g], Georgios V Gkoutos[a,d,k,l]

[a] NIHR Surgical Reconstruction and Microbiological Research Centre (SRMRC), Heritage Building, Queen Elizabeth Hospital, Mindelsohn Way, Edgbaston, Birmingham B15 2TH, UK
[b] Institute of Inflammation and Ageing, University of Birmingham, Birmingham B15 2TT, UK
[c] 212 (Yorkshire) Field Hospital, Endcliffe Hall, Endcliffe Vale Road, Sheffield S10 3EU, UK
[d] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, UK
[e] Academic Department of Military Emergency Medicine, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, UK
[f] Academic Department of Military Surgery & Trauma, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, UK
[g] University Hospitals Birmingham, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, UK
[h] NHS England London, Skipton House, 80 London Road, London SE1 6LH, UK
[i] Army Health, Army Headquarters, Monxton Road, Andover SP11 8HT, UK
[j] Nottingham University Hospitals NHS Trust, Derby Road, Nottingham NG7 2UH, UK
[k] Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2TT, UK
[l] MRC Health Data Research UK (HDR UK), Midlands Site, B15 2TT UK

ARTICLE INFO

Article History:
Received 22 December 2020
Revised 19 April 2021
Accepted 20 April 2021
Available online 15 May 2021

Keywords:
Major incident
Mass casualty
Prehospital medicine
Triage
Major trauma
Disaster
Injury Severity Score
BCD Triage Sieve
Careflight
START
MSTART
MPTT-24
RAMP
JumpSTART
MIMMS
Life-saving intervention
Military medicine

ABSTRACT

Background: Natural disasters, conflict, and terrorism are major global causes of death and disability. Central to the healthcare response is triage, vital to ensure the right care is provided to the right patient at the right time. The ideal triage tool has high sensitivity for the highest priority (P1) patients with acceptably low over-triage. This study compared the performance of major incident triage tools in predicting P1 casualty status in adults in the prospective UK Trauma Audit and Research Network (TARN) registry.

Methods: TARN patients aged 16+ years (January 2008-December 2017) were included. Ten existing triage tools were applied using patients' first recorded pre-hospital physiology. Patients were subsequently assigned triage categories (P1, P2, P3, Expectant or Dead) based on pre-defined, intervention-based criteria. Tool performance was assessed by comparing tool-predicted and intervention-based priority status.

Findings: 195,709 patients were included; mortality was 7·0% (n=13,601); median Injury Severity Score (ISS) was 9 (IQR 9–17); 97·1% sustained blunt injuries. 22,144 (11·3%) patients fulfilled intervention-based criteria for P1 status, exhibiting higher mortality (12·8% *vs.* 5·0%, p<0.001), increased intensive care requirement (52·4% *vs* 5·0%, p<0.001), and more severe injuries (median ISS 21 *vs* 9, p<0.001) compared with P2 patients. In 16–64 year olds, the highest performing tool was the Battlefield Casualty Drills (BCD) Triage Sieve (Prediction of P1 status: 70·4% sensitivity, over-triage 70·9%, area under the receiver operating curve (AUC) 0·068 [95%CI 0·676–0·684]). The UK National Ambulance Resilience Unit (NARU) Triage Sieve had sensitivity of 44·9%; over-triage 56·4%; AUC 0·666 (95%CI 0·662–0·670). All tools performed poorly amongst the elderly (65+ years).

Interpretation: The BCD Triage Sieve performed best in this nationally representative population; we recommend it supersede the NARU Triage Sieve as the UK primary major incident triage tool. Validated triage category definitions are recommended for appraising future major incidents.

Funding: This study is funded by the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre. GVG also acknowledges support from the MRC Heath Data Research UK (HDRUK/CFC/01). The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the Ministry of Defence.

* Corresponding author at: NIHR Surgical Reconstruction and Microbiological Research Centre (SRMRC), Heritage Building, Queen Elizabeth Hospital, Mindelsohn Way, Edgbaston, Birmingham, B15 2TH, UK.

E-mail address: nabeelamalik@nhs.net (N.S. Malik).

https://doi.org/10.1016/j.eclinm.2021.100888
2589-5370/

Research in context

Evidence before this study

Major incidents, ranging from terrorist attacks to large-scale natural disasters, require prioritisation of limited healthcare resources in order to maximise survival amongst those injured. Selection of the optimal triage tool to prioritise patients at the scene of a major incident is an essential component of disaster preparedness and several such tools exist internationally. Those in need of urgent life-saving intervention (Priority 1, P1) are at greatest risk of adverse outcome, hence their timely and accurate identification is of greatest importance, however comparing results from existing studies is limited by lack of consensus on what endpoint best defines P1 status.

Added value of this study

Using 195,000 patients from the UK national trauma registry, this study measured the performance of ten international major incident triage tools, allowing direct comparison in their ability to predict both P1 status and mortality. We validated a system of retrospectively assigning triage categories based on commonly described pre-hospital and hospital interventions. The best performing tool was the UK military's Battlefield Casualty Drills (BCD) Triage Sieve, affording a 24–26% improvement in identifying P1 patients over the current National Ambulance Resilience Unit (NARU) Triage Sieve .

Implications of all the available evidence

Existing consensus-derived definitions of major incident triage categories, which relate directly to healthcare resource utilisation, have been validated. We recommend their use as an endpoint for future evaluations of UK and international major incidents, research and related training. A number of studies have demonstrated that the NARU Triage Sieve used currently by UK ambulance services is not the optimal tool for major incident triage; the BCD Triage Sieve may facilitate a substantial improvement in detecting patients requiring time-critical, life-saving intervention.

1. Introduction

The global incidence of natural disasters, conflict, and terrorism has risen in the last two decades [1,2]; with over two million people dead, many more wounded, and incurring far-reaching economic and societal consequences [1–4]. In the UK, recent major incidents include terrorist-related combined vehicular and stabbing attacks [2], a shrapnel-laden bomb in a concert hall [4], and a large residential fire [2]. Triage, the sorting of casualties according to priority, was conceived during the Napoleonic Wars [5]. Triage enables prioritisation for treatment and onward transfer, and selection of an appropriate destination for definitive medical care [3,6-8]. Those in need of urgent life-saving intervention (Priority 1, or P1, category, also known as Immediate, Red, and Triage category 1 or "T1" internationally) are at greatest risk of potentially preventable adverse outcome, hence their timely and accurate identification is the most important priority of major incident triage [5,8,9]. Accurate triage ensures that limited medical resources are directed towards achieving the greatest possible positive impact for the largest number of people [3,10]. Incorrect triage may fail to identify patients in need of urgent intervention (under-triage); however, its inverse (over-triage) risks overwhelming healthcare facilities with patients who do not require time-critical treatment [3,8–10].

The selection of appropriate major incident triage tools is an important component of disaster and major incident preparedness [1,3,11]. Algorithmic tools used at the scene of a major incident must be quick and simple to apply under challenging circumstances. In the UK, emergency medical services (EMS) currently utilise the National Ambulance and Resilience Unit (NARU) Triage Sieve [12], adapted from the UK military's former MIMMS Triage Sieve [5]. The Battlefield Casualty Drills (BCD) Triage Sieve, used by British soldiers faced with multiple casualties, first appeared in 1998 [13], undergoing serial updates in line with emerging evidence and changes in clinical practice. The most recent update in 2018 (**Supplementary Data Fig. 1**) incorporated assessment of mental status and a revised respiratory rate threshold (adopted from MPTT-24) [14], a revised heart rate threshold and the rolling of unresponsive patients into the three-quarter prone position. The US-based Simple Triage and Rapid Treatment (START) and modified version MSTART tools [5,15] have been evaluated following several disasters [6,8], and registry-based studies [15,16], demonstrating sensitivity of 85% to 100% in predicting P1

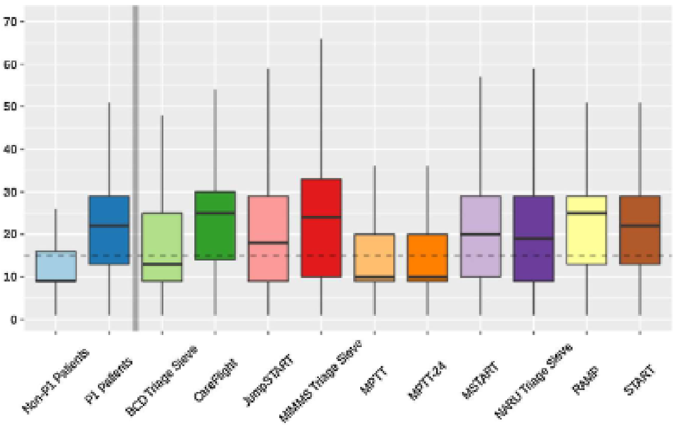


Fig. 1. Distribution of Injury Severity amongst tool-assigned P1 patients (patients aged 16–64 years)
Ledger: ISS=Injury Severity Score. Dotted horizontal line denotes ISS 15. The upper whisker extends from the hinge to the largest value no further than 1•5 * IQR from the hinge; the lower whisker extends from the hinge to the smallest value, at most 1•5 * IQR of the hinge.

status [5]. The Australian CareFlight has demonstrated appreciable sensitivity in predicting P1 status (46–82%) and mortality (AUC 0·852) in events including the 2002 Bali nightclub bombings [5], hospital [15] and trauma registry-based studies [16]. Novel tools include the Modified Physiological Triage Tool (MPTT)[14], MPTT-24 [17], and US-based Rapid Assessment of Mentation and Pulse (RAMP) [18]. JumpSTART, developed for use in children under eight years, has yet to be evaluated in adult patients [5,11]. Few studies examine the performance of triage tools in the elderly [16], who constitute a growing proportion of the UK population [19].

There is a paucity of evidence to guide policy makers in their choice of triage tool [5,8,11]. Conducting prospective studies to directly compare tool performance during major incidents is logistically challenging, given their unpredictable, infrequent nature [2,5,8,10]. Thus, existing evidence comprises post-event evaluations [6,7,10], often limited by small patient numbers with incomplete pre-hospital data [20], simulation studies [8], and studies using hospital [15] or trauma registry patients [14,16] as surrogates for those injured in major incidents. Comparing study results is further limited by lack of consensus on what endpoint best defines P1 status [5,10,15,21]. Some studies have retrospectively utilised ISS>15 as an endpoint to assign high-acuity status and justify transfer to the uppermost tier of trauma care [10,16]. Whilst ISS and mortality are commonly used, the need for life-saving intervention is recognised as the most appropriate endpoint as this relates directly to resource utilisation in the resource-constrained major incident setting [15,21]. In order to inform major incident triage practice in the UK civilian setting, this study aimed to compare the performance of major incident triage tools applicable at the scene of injury in predicting P1 casualty status amongst adults using the UK trauma registry database. A secondary aim was to assess the utility of a consensus-based system of defining triage categories for the retrospective evaluation of major incident triage [21].

2. Methods

2.1. Overview of study design

This study tests the performance of ten major incident triage tools in predicting P1 status using patients from the UK Trauma Audit and Research Network (TARN) registry. Each triage tool was applied to patients' pre-hospital physiology to determine whether the patient would have been designated P1 status. Patient records were reviewed to determine whether they required time-critical interventions from a pre-defined list, allowing assignment of "actual" triage categories (P1, P2, P3, Expectant or Dead) [21]. Tool performance was assessed by comparing tool-predicted and intervention-based P1 status.

2.2. Selection of participants

Prospectively recorded, anonymised data for TARN registry patients aged 16+ years presenting to hospitals in England and Wales between 1 January 2008 and 31 December 2017 were included. Patients with incomplete pre-hospital physiological data required to apply the triage tools (respiratory rate, heart rate, capillary refill time, Glasgow Coma Score (GCS), and GCS Motor Component) were excluded.

TARN constitutes the largest trauma registry in Europe, receiving data from all UK Major Trauma Centres and Trauma Units [22]. TARN hospital co-ordinators include injured patients fulfilling the following criteria: length of stay over 72 hours, intensive care (ICU) admission and/or in-hospital death [22]. Pre-hospital and in-hospital physiological, demographic, and outcome data are prospectively recorded into a web-based proforma [22]. TARN excludes pre-hospital deaths and elderly patients with isolated femoral neck fractures [22].

2.3. Application of major incident triage tools

It was anticipated that TARN's inclusion criterion of length of stay greater than 72 hours would result in over-representation of elderly patients within the study population. Therefore, to test tool performance, patients were categorised by age into adults (16–64 years) and the elderly (65+ years), consistent with National Health Service configuration. The BCD Triage Sieve (**Supplementary Data Fig. 1**), CareFlight [5], JumpSTART [5], MIMMS Triage Sieve [5], MPTT [14], MPTT-24 [17], MSTART [5], NARU Triage Sieve [12], RAMP [18] and START [5] tools were transcribed into computer code and applied to first recorded pre-hospital physiology to determine whether patients were P1 or non-P1. Tool characteristics including their precise components are summarised in **Table 1**. The US-based SALT [11] and German ASAV [23] tools were considered for inclusion, however these require subjective judgements, limiting reliable retrospective application.

In order to facilitate retrospective application of the triage tools (see Table 1), several assumptions were made. By virtue of fulfilling TARN inclusion criteria, all patients were assumed to be non-ambulatory. Patients who had undergone an advanced airway intervention at scene were deemed unable to breathe [24]. A respiratory rate of less than four breaths per minute was regarded undetectable by EMS personnel. The term "catastrophic haemorrhage" utilised by the BCD and NARU Triage Sieve, and MPTT-24 could not be applied retrospectively as this field is not captured by TARN. Patients with a systolic blood pressure of 90 mmHg and over were regarded as having a palpable radial pulse [25]. Patients were deemed unconscious if their GCS score was less than or equal to eight and those with a GCS of less than 12 were deemed unresponsive to voice [26]. Ability to follow commands was equated to GCS Motor Score of six by convention. In applying JumpSTART, a GCS Motor Score of three or less was regarded as equivalent to "inappropriate response to painful stimulus (*e.g.* posturing) or unresponsive to noxious stimulus [26]."

2.4. Outcome measures

The primary outcome measure was the ability of triage tools to predict P1 status, defined as the need for time-critical lifesaving intervention(s) [21]. Each patient was assigned a triage category (Dead, Expectant, P1, P2 or P3) based on a pre-defined system utilising EMS and hospital-based interventions described by Lerner *et al*, using equivalent TARN terminology **(see Supplementary Data Table 1)** [21]. As TARN does not include patients with chemical, biological, radiological, and nuclear injuries, criteria relevant to this injury mechanism were not included [21,22]. TARN records the timing of hospital arrival and each intervention, allowing incorporation of this interval into the time-critical definitions constituting P1 status. To assess the validity of Lerner's system of classification [21], patients within each category were compared with regard to mortality, ICU admission, hospital length of stay (LOS) and ISS.

Secondary outcome measures included prediction of mortality and ISS>15, and distribution of ISS amongst tool-assigned P1 patients, which may provide further discriminative value and appreciation of tool characteristics.

2.5. Data processing and analyses

TARN data were received in SPSS Version 24·0 (Armonk NY: IBM Corp 2015) and processed using R software (Version 3·6, R Core Team, New Zealand, 2000). Non-parametric data are presented as median and interquartile range; categorical data as frequency and percent. To indicate whether the differences between P1 and P2 patients as designated by Lerner's criteria [21] were statistically significant, the Chi-squared test (comparing mortality and ICU admission) and Mood's median test [27] (comparing ISS) were utilised. Performance characteristics included sensitivity, specificity, positive

Table 1
Summary of triage tool characteristics.

Tool	Description and geographical use	Tool components 1st step	2nd step	3rd step	4th step	5th step	6th step	7th step	Interventions permitted
Battlefield Casualty Drills (BCD) Triage Sieve	Current UK military tool for use in adults (introduced in 1998, revised 2018).	Catastrophic haemorrhage?	Walking?	Breathing?	Responds to voice?	Breathing rate between 12–23	Heart Rate more than 100	–	Apply tourniquet, open airway, place casualty in the $^{3}/_{4}$ prone recovery position
CareFlight	Australian tool used in adults and children (introduced in 2001).	Walks?	Obeys command?	Palpable radial pulse? OR Breathes with open airway?	–	–	–	–	Open airway
Jump Simple Triage and Rapid Treatment (JumpSTART)	United States, used in several states in children (introduced in 2001).	Able to walk?	Spontaneous breathing (check radial pulse if apnoeic)	Respiratory rate <15 or >45	Palpable pulse?	Neurological Assessment (AVPU)	–	–	Airway positioning, 5 rescue breaths if apnoeic
Major Incident Medical Management and Support (MIMMS) Triage Sieve	Former UK military adult triage tool (introduced in 1995).	Walking	Breathing	Respiratory rate <10 or ≥30	Capillary refill >2 seconds	–	–	–	Open airway
Modified Physiological Triage Tool (MPTT)	UK-based tool* modelled in a military cohort (described in 2017).	Walking?	Breathing?	Respiratory rate <12 or ≥22	Heart rate ≥100	GCS <14	–	–	–
Modified Physiological Triage Tool 24 (MPTT-24)	UK-based tool*, modification of MPTT (described in 2017).	Catastrophic Haemorrhage?	Walking?	Breathing?	Responds to voice	Respiratory rate <12 or ≥24	Heart rate ≥100	–	Apply tourniquet or haemostatic dressing
Modified Simple Triage and Rapid Treatment (MSTART)	United States, modification of START (described in 2006).	Able to walk?	Spontaneous breathing	Respiratory rate >30	Radial pulse absent	Obey commands	–	–	Position airway
National Ambulance and Resilience Unit (NARU) Triage Sieve	Current UK civilian adult tool, adapted from the MIMMS Triage Sieve (this version was introduced in 2013)	Catastrophic haemorrhage	Are they injured	Walking	Breathing	Unconscious	Respiratory rate <10 or ≥30	Pulse >120 or capillary refill >2 sec	Apply tourniquet/ haemostatic dressing, open airway, place in recovery position
Rapid Assessment of Mentation and Pulse (RAMP)	United States, used by the Rocky Mountain Fire Department, Colorado (introduced in 2018).	Casualty without signs of obvious death	Casualty follows commands	Radial pulse present?	–	–	–	–	Control massive haemorrhage, open airway, chest decompression
Simple Triage and Rapid Treatment (START)	United States (introduced in 1983).	Able to walk?	Spontaneous breathing	Respiratory rate >30	Capillary refill >2 sec	Obey commands	–	–	Position airway
Sort, Assess, Life-saving interventions, Treatment/ Transport (SALT)	United States (introduced in 2008 by the Centre of Disease Control).	Sort**	Breathing	Obeys commands or makes purposeful movements?	Has peripheral pulse?	Not in respiratory distress?	Major haemorrhage is controlled?	Likely to survive given current resources?	Control major haemorrhage, open airway (if child, consider 2 rescue breaths), chest

(continued on next page)

Table 1 (*Continued*)

Tool	Description and geographical use	Tool components 1st step	2nd step	3rd step	4th step	5th step	6th step	7th step	Interventions permitted
									decompression, auto injector antidotes
Amberg-Schwandorf Algorithm (ASAV)***	German adaptation of the MSTART (described in 2013).	Ambulating?	Deadly injured?	Breathing difficulties?	Spurting haemorrhage?	Radial pulse absent?	Unable to follow simple commands?		Keep airway open

Ledger: Respiratory rate and heart rate are measured as breaths and beats per minute, respectively. AVPU refers to the Alert, Voice, Pain, Unresponsive scale; GCS=Glasgow Coma Score. All tools described are applicable at the scene of a major incident (primary triage tools). *Has yet to undergo practical use or implementation studies. SALT and ASAV were not evaluated in this study as there were major limitations in applying these retrospectively. **SALT involves sorting according to the following: walk, wave/purposeful movement, still/obvious life threat; as well as the subjective judgements: "Minor injuries only?" and "Likely to survive given current resources?" ***ASAV includes the subjective judgement "Deadly injured?" and assessment of breathing status as follows: "airway obstructed, bradypnoea, apnoea, dyspnoea, tachypnoea (not obviously psychogenic) and cyanosis."

Table 2
Patient and injury characteristics (n=195,709).

Characteristic	n (%)
Gender	
Male	104,019 (53·1%)
Female	91,690 (46·9%)
Missing data	0 (0·0%)
Injury Severity Score (ISS)	
Median (IQR)	9 (9–17)
Missing data	0 (0·0%)
Age	
Median (IQR), years	66·2 (47·3–83·0)
Patients aged 16–64 years	95,306 (48·7%)
Patients aged 65+ years	100,403 (51·3%)
Missing data	0 (0·0%)
Discharge status	
Alive	182,107 (93·0%)
Dead	13,601 (7·0%)
Missing data	1 (0·0%)
Mode of injury	
Blunt	190,048 (97·1%)
Penetrating	5660 (2·9%)
Missing data	1 (0·0%)
Mechanism of injury	
Fall less than 2m	113,319 (57·9%)
Vehicle Incident/Collision	41,590 (21·3%)
Fall more than 2m	25,194 (12·9%)
Blow(s)	6827 (3·5%)
Stabbing	4105 (2·1%)
Other	2609 (1·3%)
Crush	1355 (0·7%)
Shooting	440 (0·2%)
Blast	142 (0·1%)
Burn	128 (0·1%)
Missing data	0 (0·0%)

predictive value, negative predictive value, under-triage (1-sensitivity), over-triage (1-positive predictive value), and Area Under the Receiver Operating Curve (AUC). 95% confidence intervals were calculated using the Wilson Score with continuity correction for binomial proportions, and DeLongs Algorithm for comparing AUC curves [28]. To estimate bias, patients included in the study were compared to those excluded with respect to clinical and demographic characteristics (**Supplementary Data** Table 2). A value of $p<0{\cdot}05$ was considered statistically significant.

Ethical approval: The UK Health Research Authority Patient Information Advisory Group (Section 20) have granted ethical approval and waived the requirement for individual participant consent for research using anonymised TARN data.

Role of Funding: The funding source had no role to play in the in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

3. Results

3.1. Characteristics of the study population

TARN captured 215,632 patients. 19,923 (9·2%) patients were excluded due to incomplete data; hence, 195,709 patients were included. Patient and injury characteristics are summarised in Table 2. There was a slight male preponderance (53·1% male *vs.* 46·9% female). Elderly patients constituted approximately half of the study population (n=100,403, 51·3%). Mortality was 7·0% (n=13,601), median ISS was 9 (IQR 9–17). Blunt-injury patients accounted for 97·1% of patients, with low-level falls (n=113,319, 57·9%) and vehicle collision (n=41,590, 21·3%) recorded as the most prevalent injury mechanisms. Penetrating trauma constituted only 2·9% (n=5660), largely comprising stabbings (n=4105, 2·1%). Comparison between included and excluded patients is shown in **Supplementary Data** Table 2.

Table 3
Comparison of outcome characteristics between patients in each triage category.

Triage category	Total, n (%)	Mortality, n (%)	Intensive care admission, n (%)	Length of stay (days), median (IQR)	ISS, median (IQR)
Dead	282 (0·1)	282 (100·0)	145 (51·4)	1 [1, 3]	27 [25, 41]
Expectant	1879 (1·0)	1862 (99·1)	1021 (54·3)	1 [1, 4]	26 [25, 38]
Priority 1 (Immediate)	22,144 (11·3)	2839 (12·8)	11,593 (52·4)	11 [4, 26]	21 [10, 29]
Priority 2 (Urgent)	171,404 (87·6)	8618 (5·0)	8661 (5·0)	10 [5, 18]	9 [9, 16]

Ledger: IQR=interquartile range, ISS=Injury Severity Score.

Table 4
Breakdown of time-critical life-saving interventions constituting Priority 1 status.

Subcomponents of the Priority 1 triage category	n (% of P1 patients)
An advanced airway intervention (*e.g.* intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	18,890 (85·3%)
Chest tube placed within 2 hours of arrival at hospital	4301 (19·4%)
Neurological, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	3427 (15·5%)
Arrived in the ED with uncontrolled haemorrhage	1979 (8·9%)
Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	606 (2·7%)
IV vasopressors administered within 2 hours of arrival at hospital	361 (1·6%)
Patient who required EMS initiation of CPR (*i.e.* had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	182 (0·8%)
Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	3 (0·0%)
Total number of P1 patients	22,144 (100·0%)

Ledger: There is overlap between life-saving interventions (LSI): 78·0% (n=17,272) of P1 patients required one LSI, 15·9% (n=3520) required two LSI, and 6·0% (n=1328) required 3 or more LSI.

3.2. *Designation of intervention-based triage categories (using Lerner's criteria)*

A minority of patients met criteria for the "Dead" category (n=282, 0·1%) whilst 1·0% (n=1879) of the study population were classed as Expectant due to non-survivable burns (n=36) and catastrophic head injury (n=1843) (**Table 3**). 22,144 (11·3%) of patients satisfied criteria for P1 status; most commonly requiring airway intervention (n=18,890, 85·3%), emergent chest tube placement (n=4301, 19·4%), and emergency surgery (n=3427, 15·5%) (**Table 4**). No TARN patients met criteria for the minimally injured P3 group (**Supplementary Data Table 1**). The remaining patients (n=171,402, 87·0%) were assigned P2, forming the largest triage category.

Comparison of patient characteristics across the four triage categories revealed important differences (**Table 3**). Mortality was universal in those classed "Dead" and 99·1% in those labelled Expectant, both groups exhibited very severe injuries (median ISS 27 and 26, respectively). P1 patients demonstrated more than double the mortality (12·8% *vs.* 5·0%; p<0.001), ten times the rate of ICU admission (52·4% *vs.* 5·0%; p<0.001), and more severe injuries (median ISS 21 *vs.* median ISS 9; p<0.001, respectively) compared to patients designated P2.

Table 5
Tool Performance in Predicting Intervention-based Priority 1 Status in Adults aged 16–64 years.

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	70·4 [69·7, 71·1]	65·6 [65·3, 66·0]	29·1 [28·6, 29·6]	91·7 [91·5, 91·9]	29·6 [28·9, 30·3]	70·9 [70·4, 71·4]	0·680 [0·676, 0·684]
CareFlight	43·3 [42·6, 44·1]	92·8 [92·7, 93·0]	54·8 [53·9, 55·7]	89·1 [88·9, 89·3]	56·7 [55·9, 57·4]	45·2 [44·3, 46·1]	0·681 [0·677, 0·685]
JumpSTART	46·8 [46·1, 47·6]	89·3 [89·0, 89·5]	46·6 [45·8, 47·4]	89·3 [89·1, 89·6]	53·2 [52·4, 53·9]	53·4 [52·6, 54·2]	0·681 [0·676, 0·685]
MIMMS Triage Sieve	41·8 [41·0, 42·5]	93·4 [93·3, 93·6]	56·0 [55·1, 56·9]	88·9 [88·7, 89·1]	58·2 [57·5, 59·0]	44·0 [43·1, 44·9]	0·676 [0·672, 0·680]
MPTT	49·9 [49·1, 50·7]	59·1 [58·7, 59·4]	19·6 [19·2, 20·0]	85·5 [85·2, 85·8]	50·1 [49·3, 50·9]	80·4 [80·0, 80·8]	0·545 [0·541, 0·549]
MPTT-24	47·9 [47·1, 48·7]	62·9 [62·6, 63·2]	20·6 [20·1, 21·0]	85·8 [85·5, 86·1]	52·1 [51·3, 52·9]	79·4 [79·0, 79·9]	0·554 [0·550, 0·558]
MSTART	57·2 [56·5, 58·0]	89·0 [88·8, 89·3]	51·1 [50·4, 51·9]	91·2 [91·0, 91·4]	42·8 [42·0, 43·5]	48·9 [48·1, 49·6]	0·731 [0·727, 0·735]
NARU Triage Sieve	44·9 [44·1, 45·7]	88·4 [88·2, 88·6]	43·6 [42·9, 44·4]	88·9 [88·7, 89·1]	55·1 [54·3, 55·9]	56·4 [55·6, 57·1]	0·666 [0·662, 0·670]
RAMP	39·4 [38·6, 40·1]	93·3 [93·1, 93·5]	54·1 [53·2, 55·0]	88·5 [88·3, 88·7]	60·6 [59·9, 61·4]	45·9 [45·0, 46·8]	0·663 [0·660, 0·667]
START	53·7 [52·9, 54·5]	90·9 [90·7, 91·1]	54·2 [53·4, 55·0]	90·7 [90·5, 90·9]	46·3 [45·5, 47·1]	45·8 [45·0, 46·6]	0·723 [0·719, 0·727]

Ledger: Results are accompanied by 95% confidence intervals. PPV=positive predictive value, NPV=negative predictive value. AUC=Area Under the Receiver Operating Curve.

Table 6
Tool Performance in Predicting Intervention-based Priority 1 Status amongst the Elderly (aged 65± years).

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	56·7 [55·5, 57·9]	72·7 [72·4, 73]	12·1 [11·7, 12·5]	96·2 [96·1, 96·3]	43·3 [42·1, 44·5]	87·9 [87·5, 88·3]	0·647 [0·641, 0·653]
CareFlight	33·5 [32·3, 34·7]	93·4 [93·3, 93·6]	25·3 [24·4, 26·3]	95·5 [95·4, 95·6]	66·5 [65·3, 67·7]	74·7 [73·7, 75·6]	0·635 [0·629, 0·641]
JumpSTART	36·1 [34·9, 37·3]	90·7 [90·5, 90·9]	20·5 [19·7, 21·2]	95·5 [95·4, 95·7]	63·9 [62·7, 65·1]	79·5 [78·8, 80·3]	0·634 [0·628, 0·640]
MIMMS Triage Sieve	34·7 [33·5, 35·9]	92·8 [92·7, 93·0]	24·3 [23·4, 25·2]	95·5 [95·4, 95·7]	65·3 [64·1, 66·5]	75·7 [74·8, 76·6]	0·638 [0·632, 0·644]
MPTT	45·4 [44·1, 46·6]	66·4 [66·1, 66·7]	8·2 [7·9, 8·5]	94·8 [94·7, 95·0]	54·6 [53·4, 55·9]	91·8 [91·5, 92·1]	0·559 [0·553, 0·565]
MPTT-24	43·1 [41·9, 44·3]	69·9 [69·6, 70·2]	8·7 [8·4, 9·0]	94·9 [94·7, 95·0]	56·9 [55·7, 58·1]	91·3 [91·0, 91·6]	0·565 [0·559, 0·571]
MSTART	48·6 [47·4, 49·9]	88·5 [88·3, 88·7]	21·8 [21·2, 22·5]	96·3 [96·2, 96·4]	51·4 [50·1, 52·6]	78·2 [77·5, 78·8]	0·686 [0·679, 0·692]
NARU Triage Sieve	33·2 [32·1, 34·4]	89·6 [89·4, 89·8]	17·5 [16·9, 18·2]	95·3 [95·2, 95·4]	66·8 [65·6, 67·9]	82·5 [81·8, 83·1]	0·614 [0·609, 0·620]
RAMP	31·3 [30·1, 32·4]	93·7 [93·5, 93·9]	24·7 [23·8, 25·7]	95·4 [95·2, 95·5]	68·7 [67·6, 69·9]	75·3 [74·3, 76·2]	0·625 [0·619, 0·631]
START	45·9 [44·7, 47·2]	89·9 [89·7, 90·1]	23·2 [22·5, 24·0]	96·2 [96·0, 96·3]	54·1 [52·8, 55·3]	76·8 [76·0, 77·5]	0·679 [0·673, 0·686]

Ledger: Results are accompanied by 95% confidence intervals. PPV=positive predictive value, NPV=negative predictive value. AUC=Area Under the Receiver Operating Curve.

Table 7
Tool performance in predicting mortality in adults aged 16–64 years.

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	85·2 [83·8, 86·6]	60·9 [60·5, 61·2]	5·6 [5·4, 5·8]	99·3 [99·3, 99·4]	14·8 [13·4, 16·2]	94·4 [94·2, 94·6]	0·730 [0·723, 0·738]
CareFlight	69·6 [67·8, 71·4]	88·3 [88·1, 88·6]	14·0 [13·4, 14·6]	99·1 [99·0, 99·1]	30·4 [28·6, 32·2]	86·0 [85·4, 86·6]	0·790 [0·781, 0·799]
JumpSTART	70·0 [68·2, 71·8]	84·7 [84·5, 84·9]	11·1 [10·6, 11·6]	99·0 [99·0, 99·1]	30·0 [28·2, 31·8]	88·9 [88·4, 89·4]	0·774 [0·765, 0·783]
MIMMS Triage Sieve	63·3 [61·4, 65·2]	88·9 [88·7, 89·1]	13·5 [12·9, 14·2]	98·9 [98·8, 99·0]	36·7 [34·8, 38·6]	86·5 [85·8, 87·1]	0·761 [0·752, 0·771]
MPTT	34·2 [32·3, 36·1]	57·4 [57·0, 57·7]	2·1 [2·0, 2·3]	97·0 [96·8, 97·1]	65·8 [63·9, 67·7]	97·9 [97·7, 98·0]	0·458 [0·448, 0·467]
MPTT-24	33·4 [31·6, 35·3]	61·0 [60·6, 61·3]	2·3 [2·1, 2·4]	97·1 [97·0, 97·2]	66·6 [64·7, 68·4]	97·7 [97·6, 97·9]	0·472 [0·463, 0·481]
MSTART	77·3 [75·6, 78·9]	82·9 [82·7, 83·2]	11·0 [10·5, 11·5]	99·3 [99·2, 99·3]	22·7 [21·1, 24·4]	89·0 [88·5, 89·5]	0·801 [0·793, 0·809]
NARU Triage Sieve	72·7 [70·9, 74·4]	84·3 [84·1, 84·6]	11·2 [10·8, 11·7]	99·1 [99·1, 99·2]	27·3 [25·6, 29·1]	88·8 [88·3, 89·2]	0·785 [0·776, 0·794]
RAMP	50·6 [48·6, 52·6]	88·9 [88·7, 89·1]	11·1 [10·5, 11·7]	98·5 [98·4, 98·6]	49·4 [47·4, 51·4]	88·9 [88·3, 89·5]	0·698 [0·688, 0·707]
START	75·3 [73·6, 77·0]	85·1 [84·8, 85·3]	12·1 [11·6, 12·6]	99·2 [99·1, 99·3]	24·7 [23·0, 26·4]	87·9 [87·4, 88·4]	0·802 [0·794, 0·810]

Ledger: Results are accompanied by 95% confidence intervals. PPV=positive predictive value, NPV=negative predictive value. AUC=Area Under the Receiver Operating Curve.

Table 8
Tool Performance in Predicting Mortality amongst the Elderly (aged 65± years).

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	49·7 [48·8, 50·7]	73·4 [73·1, 73·7]	18·8 [18·4, 19·3]	92·2 [92·0, 92·4]	50·3 [49·3, 51·2]	81·2 [80·7, 81·6]	0·616 [0·611, 0·621]
CareFlight	26·9 [26·1, 27·8]	94·1 [93·9, 94·2]	36·1 [35·0, 37·1]	91·2 [91·0, 91·4]	73·1 [72·2, 73·9]	63·9 [62·9, 65·0]	0·605 [0·601, 0·609]
JumpSTART	23·5 [22·7, 24·3]	90·6 [90·4, 90·8]	23·6 [22·9, 24·5]	90·5 [90·3, 90·7]	76·5 [75·7, 77·3]	76·4 [75·5, 77·1]	0·571 [0·566, 0·575]
MIMMS Triage Sieve	21·5 [20·8, 22·3]	92·7 [92·5, 92·9]	26·8 [25·8, 27·7]	90·5 [90·3, 90·7]	78·5 [77·7, 79·2]	73·2 [72·3, 74·2]	0·571 [0·567, 0·575]
MPTT	48·6 [47·7, 49·5]	67·5 [67·2, 67·8]	15·6 [15·2, 16·0]	91·4 [91·2, 91·6]	51·4 [50·5, 52·3]	84·4 [84·0, 84·8]	0·580 [0·575, 0·585]
MPTT-24	45·9 [44·9, 46·8]	71·0 [70·7, 71·3]	16·4 [16·0, 16·8]	91·4 [91·2, 91·6]	54·1 [53·2, 55·1]	83·6 [83·2, 84·0]	0·584 [0·579, 0·589]
MSTART	35·7 [34·8, 36·6]	88·9 [88·7, 89·1]	28·4 [27·7, 29·2]	91·8 [91·6, 92·0]	64·3 [63·4, 65·2]	71·6 [70·8, 72·3]	0·623 [0·618, 0·627]
NARU Triage Sieve	29·4 [28·6, 30·3]	90·4 [90·2, 90·6]	27·5 [26·7, 28·4]	91·2 [91·0, 91·4]	70·6 [69·7, 71·4]	72·5 [71·6, 73·3]	0·599 [0·595, 0·604]
RAMP	25·0 [24·2, 25·9]	94·3 [94·1, 94·4]	35·1 [34·1, 36·2]	91·0 [90·8, 91·2]	75·0 [74·1, 75·8]	64·9 [63·8, 65·9]	0·597 [0·592, 0·601]
START	33·5 [32·7, 34·4]	90·3 [90·1, 90·5]	30·1 [29·3, 30·9]	91·6 [91·5, 91·8]	66·5 [65·6, 67·3]	69·9 [69·1, 70·7]	0·619 [0·615, 0·624]

Ledger: Results are accompanied by 95% confidence intervals. PPV=positive predictive value, NPV=negative predictive value. AUC=Area Under the Receiver Operating Curve.

3.3. *Triage tool performance*

3.3.1. *Prediction of P1 status*

Tool performance in predicting P1 status in adults aged 16–64 years is shown in Table 5. The NARU Triage Sieve exhibited sensitivity of 44·9% in predicting P1 status with associated over-triage of 56·4% and an AUC of 0·666 (95% CI 0·662–0·670). The BCD Triage Sieve demonstrated the highest sensitivity (70·4%) in predicting P1 status with associated over-triage of 70·9% and an AUC of 0·680 (95% CI 0·676–0·684). MSTART demonstrated the next highest sensitivity in predicting P1 status (57·2%) with the most favourable AUC of 0·731 (95% CI 0·727–0·735); followed closely by parent tool START. The novel RAMP demonstrates lower overall performance to CareFlight. MPTT and MPTT-24 demonstrated moderate sensitivity (49.9% and 47.9%), the highest over-triage rates (80·4% and 79·4%) and lowest specificity (59·1% and 62·9%). Amongst adults aged 65+ years, tools demonstrated high over-triage and all tools except for the BCD Triage Sieve achieved less than 50% sensitivity in predicting P1 status (Table 6).

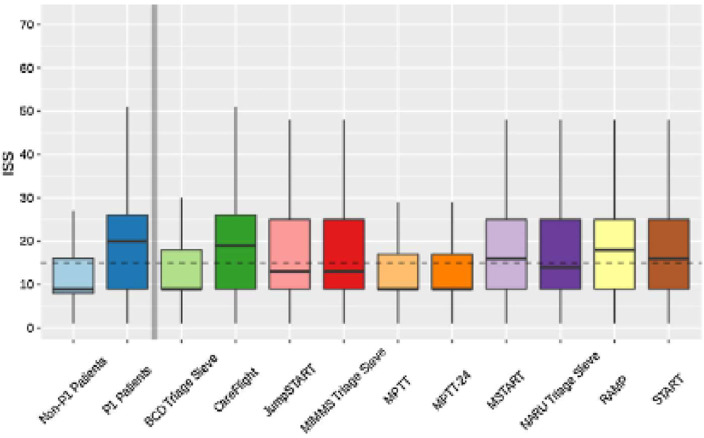


Fig. 2. Distribution of Injury Severity amongst tool-assigned P1 patients (patients aged 65± years)
Ledger: ISS=Injury Severity Score. Dotted horizontal line denotes ISS 15. The upper whisker extends from the hinge to the largest value no further than 1·5 * IQR from the hinge; the lower whisker extends from the hinge to the smallest value, at most 1·5 * IQR of the hinge.

3.3.2. Prediction of mortality

Amongst 16–64 year olds, there was greater variation in tool performance in predicting mortality (Table 7). The BCD Triage Sieve exhibited the highest sensitivity (85·2%) in predicting mortality, followed by MSTART (77·3%), START (75·3%), and NARU Triage Sieve (72.7%); all four tools attained comparable AUC (0·730–0·802). MPTT and MPTT-24 predict mortality with sensitivity of 34·2% and 33·4%, respectively, with AUC below 0·5. All tools performed poorly (sensitivity 21·5–49·7%; AUC 0·571–0·623) in predicting mortality in the elderly (Table 8).

3.3.3. Injury Severity Score

Patients with ISS>15 totalled 62,402; of these, only 24·1% (n=15,058) met criteria for intervention-based P1 status whilst 75·9% (n=47,344) were non-P1. One third (n=7086) of intervention-based P1 patients had an ISS≤15. There was great variation in the distribution of ISS amongst tool-assigned P1 patients Figs. 1 and 2). Tool performance in predicting ISS>15 is included in **Supplementary data Tables 3 and 4**.

4. Discussion

Globally, natural disasters, conflict, and terrorism pose significant and often unexpected threats, incurring substantial societal and economic impact [1,2,29]. Review of the 2017 Manchester Arena attack highlighted the importance of well-co-ordinated, multiagency collaboration in processing casualties and the negative impacts of an inadequate EMS response [4]. Meticulous disaster planning, including selection of an effective triage tool, is crucial to maximising survival [2,3,8,10,29]. This study measured the performance of ten major incident triage tools in predicting P1 status and mortality using a nationally representative UK adult patient population. A system for defining major incident triage categories relating directly to healthcare resource utilisation has been validated, yielding patient groups with distinct clinical characteristics. ISS>15 correlated poorly with the need for life-saving intervention, making this a suboptimal endpoint for evaluating major incidents. The NARU Triage Sieve, currently used in UK practice, is outperformed by several other triage tools. The military-derived BCD Triage Sieve demonstrates the greatest sensitivity in predicting both P1 status (sensitivity 70·4%, AUC 0·680, over-triage 70·9%) and mortality (sensitivity 85·2%, AUC 0·730). All tools performed poorly amongst the elderly.

Whilst the American College of Surgeons has established standards for the performance of pre-hospital triage tools for individual patients (acceptable rates of under-triage of up to 5% and 25–50% over-triage, respectively) [7], no national or international standards exist to govern tool performance in major incidents [3,11,12]. From a clinical perspective, under-triage of critically unwell (P1) patients leads to absolute harm arising from delayed care or transfer to an inappropriate medical facility. Conveying critically injured patients to the highest tier of trauma care within trauma networks is associated with decreased mortality: in the UK, bypass of local hospitals to designated Major Trauma Centres has been associated with a 19% increase in the adjusted odds of survival following severe injury [30]. However, minimising over-triage in the resource-constrained major incident setting is also crucially important as overwhelming medical facilities with patients suffering non-critical injury can impair care for those requiring time-critical interventions[8–10]: a study encompassing 3357 casualties from 220 bombing incidents demonstrated a direct linear relationship between over-triage and critical mortality [9]. Policymakers are therefore likely to favour the clinically relevant measures of under-triage and over-triage over AUC, which offers an aggregate measure of performance, selecting tools that align with local casualty distribution plans and available resources [3,4]. Whilst the need for life-saving intervention is of prime importance in guiding resource allocation in major incidents, predicting all-cause in-hospital mortality in patients with diverse, multi-system injuries is an additional useful measure of tool performance, given that the priority is to maximise overall survival [3,5,12,16]. The NARU Triage Sieve currently used in UK practice performs suboptimally in predicting both P1 status and mortality. Based on the markedly superior sensitivity in predicting P1 status and mortality in patients in this study, we recommend the BCD Triage Sieve as the primary triage tool for adult patients in the UK major incident setting. This change may afford a 24–26% improvement in detecting patients at the scene of a major incident who require time-critical life-saving intervention (and an additional 13–20% of patients who may suffer in-hospital mortality), facilitating their immediate transfer to the highest tier of trauma care and thereby maximising their chances of survival. The BCD Triage Sieve is similar in format to the NARU Triage Sieve, a potential advantage in the retraining of EMS personnel. Furthermore, omission of the need to measure capillary refill may reduce the time taken to perform triage.

The MSTART, START, and CareFlight tools demonstrate considerable sensitivity in predicting mortality, consistent with the findings of a large US registry-based study [16]. CareFlight involves four simple qualitative assessments, apparently achievable within 15 seconds [5]. CareFlight's previously demonstrated superior performance in children confers the potential advantage that a single tool used across all ages would simplify EMS training and practice [16]. Such strengths render it a viable choice for non-clinical UK emergency services personnel, where its relative ease of applicability and low over-triage rate may prove advantageous. In adults, JumpSTART is substantially outperformed by several other tools, limiting its potential utility as a single tool for use across all ages. MPTT and MPTT-24 have comparable over-triage rates to the BCD Sieve, however the BCD Sieve offers a 20·5–22·5% sensitivity advantage in predicting the need for life-saving intervention. Amongst the elderly, all tools performed poorly, consistent with the findings of a large US registry-based study [16]; this may be attributable to age-related physiological changes, chronic illness and polypharmacy. The elderly represent 18·3% of the total UK population [19], yet constitute 51·3% of the study population (51·3%), likely due to the TARN inclusion criterion of admission exceeding 72 hours[22]: this over-representation has been mitigated by analysing this cohort as a separate subgroup. Further research is needed to determine how triage can be improved in the elderly.

A key strength of this study is the use of a nationally representative sample of patients, with detailed physiological data, injured by a range of mechanisms. Our findings are therefore applicable to UK adults involved in all-hazard major incidents. ISS has again been shown to correlate poorly with resource use and need for intervention [31]. Our study has externally validated a consensus-derived definition of triage categories [21]. Application of Lerner's criteria defined P1 patients with characteristics (mortality 12·8%, 52·4% ICU admission, median ISS of 21) appropriate for the highest tier of trauma care [4,7,12]; whilst those designated P2 (mortality 5·0%, 5·0% ICU admission rate, median ISS of 9) may appropriately be treated in second-tier centres [21]. Incorporating timing of intervention to distinguish P1 from P2 patients is also clinically meaningful in the major incident setting. Furthermore, the proportion of P1 casualties (11·3%) yielded is comparable to the 10·8–17·7% critical injury rate reported in recent UK [4,10] and international major incidents [8,29]. We recommend that Lerner's criteria [21] be employed as the gold standard for future use in post-event evaluations, research, and training purposes in the UK major incident setting. This will enable standardised assessment of major incidents and triage systems, aiding resource planning and policy refinement.

The Expectant category (also known as P1 hold, P4 or T4) has previously been defined as "*casualties whose condition is so severe that they cannot survive despite the best available care and whose treatment would divert medical resources from salvageable patients who may then* *be compromised* [12]." However, triage category definitions proposed by Lerner did not take into account resource limitations[21]: the definition employed in this study was associated with near-universal mortality (99·1%) and assigned to patients whose injuries would be non-survivable under any circumstances (*e.g.* 90% burns). The Expectant category has never been assigned in UK civilian practice[4]: rationing healthcare to severely injured, living patients is ethically challenging [11,12]. We recommend that this non-resource dependant definition of the Expectant category is appropriate for academic use and that in practice, Expectant status should only be assigned by a senior clinician.

Weaknesses of this study include use of singly injured TARN patients as surrogates for those injured in major incidents, in whom outcomes (*e.g.* mortality) may be considerably worse. Blunt trauma predominates in this registry population: therefore, triage tool performance may not be completely generalisable to penetrating or blast mechanism incidents. Some assumptions made to facilitate retrospective triage tool application (*e.g.* patients meeting TARN inclusion criteria are non-ambulatory) may not hold true in real-life. Study conclusions were unlikely to be biased by patients (9·2%) excluded due to missing data. TARN inclusion criteria are biased towards capturing the severely injured, however this can be viewed as a strength since these patients are at greatest risk of adverse outcome: care received is unlikely to greatly influence outcome in the large number of "walking wounded" patients who predominate in major incidents [1,6,10]. Additionally, TARN excludes pre-hospital deaths. Given that most trauma deaths occur pre-hospital [8,10,29], our study fails to capture the main indicators of early mortality by only analysing tool performance in patients who reach hospital alive; however, those who do not make it to hospital alive under normal circumstances are even less likely to do so during a major incident, where transfer will likely be delayed due to pre-hospital resource constraints. Furthermore, our study focussed on tool performance in predicting P1 status: further research is required to evaluate tool performance in predicting non-P1 categories. Notably, triage tools commonly assign P3 status to those able to walk, however a Dutch study of an aeroplane crash involving 135 casualties revealed serious underlying injuries in 17% of ambulatory patients [20]. Finally, computed retrospective application of triage tools is used as a surrogate for EMS personnel conducting real-time triage under challenging circumstances. This does not account for human error, or the variation in tool complexity which affects their ease and accuracy of application [5,11]. However, the current study design allows the physiological discriminative capability of tools to be assessed independently of human error, as well as overcoming the challenges of conducting prospective studies during major incidents. Consequently, registry-based studies such as this are likely to form the highest level of evidence to guide major incident triage practice and policy.

In practice, policy makers must consider several factors before adopting a triage tool for widespread use [3,5,8,11]. This includes ability of the tool to differentiate categories accurately across a variety of relevant injury mechanisms [5,16]; time taken and inter-rater reliability when applied [5,8,23]; the degree of over-triage that health systems can accommodate [1,10]; and interoperability between the multiple agencies involved in major incidents [3,8,10,11]. Additionally, regular major incident training exercises have been credited for improved performance during the 2017 Manchester Arena [4] and 2015 Paris terrorist attacks [29]. Whilst simple triage algorithms have been employed since the 1980s [5,11], further work is required to determine how technology (*e.g.* portable device applications, electronic patient management systems, wearable devices) may be effectively incorporated into clinical use to augment triage.

In conclusion, based on current available evidence and the findings of this study, the NARU Triage Sieve used currently in UK civilian practice is not the optimal tool for major incident triage. We recommend its replacement by the BCD Triage Sieve, which may afford a 24–26% improvement in detecting patients in need of time-critical, life-saving intervention at the scene of a major incident, thus facilitating their immediate transfer to the highest tier of trauma care and maximising chances of survival. ISS>15 correlates poorly with the need for life-saving intervention, making this a suboptimal endpoint in evaluating major incidents. We have validated a system of retrospectively assigning triage categories based on commonly described EMS and hospital interventions, resulting in clinically distinct groups of patients: we recommend its use in future evaluations of UK major incidents.

5. Author contributions

NM conducted a literature review prior to the study. SC and YX contributed equally to this study. NM, JV, AB, DB, DK, MF and GVG designed the study. NM, SC and YX accessed the database, verified the underlying data and conducted analysis. All authors contributed to data interpretation. NM wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

Declaration of Competing Interest

The authors confirm that they have no conflicts of interest to declare.

Data sharing statement

De-identified patient data utilised for this study are proprietary to the Trauma Audit and Research Network, University of Manchester and may be requested directly from TARN[22].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2021.100888.

References

[1] Centre for Research on the Epidemiology of Disasters and UN Office for Disaster Risk Reduction. Human cost of disasters: an overview of the last 20 years 2000-2019. Emergency Events Database (EM-DAT) Centre for Research on the Epidemiology of Disasters; 2020. p. 1–17.

[2] Moran CG, Webb C, Brohi K, Smith M, Willett K. Lessons in planning from mass casualty events in UK. BMJ 2017;359:j4765.

[3] World Health Organisation. Mass casualty management systems: strategies and guidelines for building health sector capacity. Prevention HAiCIV; 2007 editor.

[4] Deeming H. The Kerslake Report: an independent review into the preparedness for, and emergency response to, the Manchester Arena attack on 22nd May 2017. 2018.

[5] Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehospital Disaster Med 2008;23(1):3–8.

[6] Kahn CA, Schultz CH, Miller KT, Anderson CL. Does START triage work? An outcomes assessment after a disaster. Ann Emerg Med 2009;54(3):424–30 30.e1.

[7] American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient, 2014.

[8] Turner CD, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Crit Care 2016;20(1):362.

[9] Frykberg ER, Tepas 3rd JJ. Terrorist bombings. Lessons learned from Belfast to Beirut. Ann Surg 1988;208(5):569–76.

[10] Aylwin CJ, König TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. Lancet North Am Ed 2006;368(9554):2219–25.

[11] Lerner EB, Schwartz RB, Coule PL, et al. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. Disaster Med Public Health Preparedness 2008;2(1):S25–34 **Suppl**.

[12] National Ambulance Resilience Unit (NARU). Clinical guidance: medical support minimal requirements for a mass casualty incident. www.naru.org.uk; 2014. Accessed June 2020

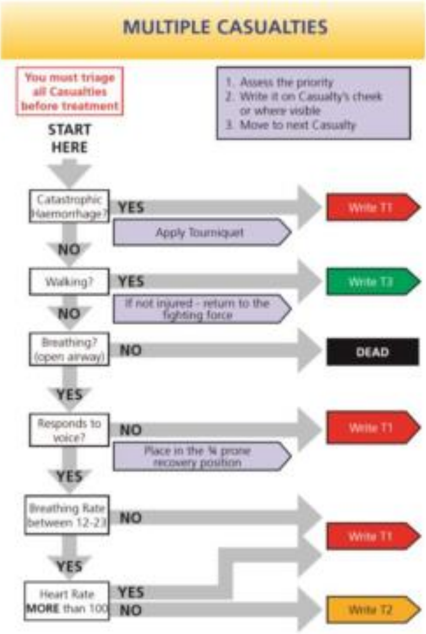
[13] Hodgetts TJ, Hanlan CG, Newey CG. Battlefield first aid: a simple, systematic approach for every soldier. J R Army Med Corps 1999;145(2):55–9.

[14] Vassallo J, Smith J, Wallis L. Major incident triage and the implementation of a new triage tool, the MPTT-24. J R Army Med Corps 2018;164:103–6.

[15] Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. Ann Emerg Med 2001;38(5):541–8.
[16] Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. Ann Emerg Med 2013;61(6):668–76 e7.
[17] Vassallo J, Smith JE, Wallis LA. Major incident triage and the implementation of a new triage tool, the MPTT-24 2018;164(2):103–6.
[18] Bennett A. Methodologies utilized and lessons learned in high threat environments and mass casualty environments. J High Threat Austere Med 2019:1–7.
[19] Office of National Statistics. Population estimates 2019. 2019. Accessed June 2020
[20] Postma IL, Weel H, Heetveld MJ, et al. Mass casualty triage after an airplane crash near Amsterdam. Injury 2013;44(8):1061–7.
[21] Lerner EB, McKee CH, Cady CE, et al. A consensus-based gold standard for the evaluation of mass casualty triage systems. Prehospital Emergency Care 2015;19 (2):267–71.
[22] The University of Manchester. The trauma audit and research network. Continuous. www.tarn.ac.uk. Accessed June 2020
[23] Wolf P, Bigalke M, Graf BM, Birkholz T, Dittmar MS. Evaluation of a novel algorithm for primary mass casualty triage by paramedics in a physician manned EMS system: a dummy based trial. Scand J Trauma Resusc Emerg Med 2014;22:50.
[24] Cobas MA, De la Peña MA, Manning R, Candiotti K, Varon AJ. Prehospital intubations and mortality: a level 1 trauma center perspective. Anesth Analg 2009;109 (2):489–93.
[25] Deakin CD, Low JL. Accuracy of the advanced trauma life support guidelines for predicting systolic blood pressure using carotid, femoral, and radial pulses: observational study. BMJ (Clinical research ed) 2000;321(7262):673–4.
[26] Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale. Ann Emerg Med 2004;44(2):108–13.
[27] Hollander M, Wolfe DA, Chicken E. Nonparametric statistical methods. Wiley; 2013.
[28] Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Process Lett 2014;21(11):1389–93.
[29] Hirsch M, Carli P, Nizard R, et al. The medical response to multisite terrorist attacks in Paris. Lancet 2015;386:2535–8 10012.
[30] Moran CG, Lecky F, Bouamra O, et al. Changing the system - major trauma patients and their outcomes in the NHS (England) 2008–17. EClin Med 2018;2:13–21.
[31] Baxt WG, Upenieks V. The lack of full correlation between the injury severity score and the resource needs of injured patients. Ann Emerg Med 1990;19 (12):1396–400.

SUPPLEMENTARY DATA:

Supplementary Data Figure 1: Battlefield Casualty Drills Triage Sieve (UK Military Primary Triage Tool)



Supplementary Data Table 1: Designation of Triage Categories based on EMS and Hospital Interventions received:

	DEAD	TARN equivalent
1.	A lack of palpable pulse and/or respiratory effort (i.e. cardiac or respiratory arrest) at initial EMS assessment that is not responsive to airway positioning or needle decompression	Any cardiorespiratory resuscitation at scene AND Inpatient death
2.	Lack of pulse or respiratory effort within 15 minutes of EMS arrival at scene	Any cardiorespiratory resuscitation at scene AND Inpatient death
		↓All no
	EXPECTANT	**TARN equivalent**
1.	In patients aged 0 to 49 years old, third degree (full thickness) burns to >90% of the body	90% Total Body Surface Burns
2.	In patients over 50 years old, third degree (full thickness) burns to >80% of the body	Age≥50 years AND 40-89% Total Body Surface Burns
3.	Penetrating or blunt trauma to the head which crosses the midline with agonal respirations and/or no motor response decorticate posturing or decerebrate posturing (i.e. a motor GCS of 3 or less)	Any head injury AND Total GCS=3 or Motor GCS≤3 at scene AND Inpatient death
4.	Uncontrolled haemorrhage that resulted in cardiac arrest prior to EMS transport	*No TARN equivalent*
		↓All no
	PRIORITY 1 (IMMEDIATE)	**TARN equivalent**
1.	Neurologic, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	At scene: Pericardial decompression, pericardiocentesis, thoracotomy, heart surgery In hospital (including ED): Evacuation of EDH/SDH, Evacuation ICH, Elevation depressed cranium, Repair Cranium fracture, Open Craniotomy, Burrhole of Cranium, Lobectomy of brain, Repair of Dura, Craniectomy Laparotomy, Splenectomy, Nephrectomy, Resection Liver, Repair Spleen, Repair Kidney laceration, Abdominal Packing, Caesarian Delivery for trauma, Colostomy, Hemicolectomy/Colectomy, Ileectomy, Repair Liver laceration, Repair Colon laceration, Repair Rupture to Bladder, Repair mesentery of small bowel, Repair mesentery of colon, Excision of Pancreas, Repair of Duodenum, Repair of Jejunum, Repair of Ileum, Repair of Stomach, Rectal operation, Bowel operations (specified), Surgery involving the Iliac artery, Surgery involving the Subclavian artery, Aortic Repair, Pericardiocentesis, Thoracotomy, Aortic Repair, Pneumonectomy, Heart Surgery, Repair of lung, Repair Oesophagus, Diaphragm repair External Fixation of Pelvis, Fixation of Pelvic Ring, Fixation of Acetabulum
2.	Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	Fasciotomy, Surgery to the brachial or femoral artery, Amputation of upper/lower limb. Any injury to the brachial or femoral artery (not time dependent)
3.	Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	Escharotomy
4.	Chest tube placed within 2 hours of arrival at hospital	Insertion of chest tube at scene or in-hospital
5.	An advanced airway intervention (e.g. intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	Airway obstruction at scene or in-hospital, Airway support required at scene or in-hospital, Intubation and mechanical ventilation at scene or in-hospital, CPAP administration in ED or Critical Care, Cricothyroidotomy or tracheostomy required at scene or in-hospital
6.	IV vasopressors administered within 2 hours of arrival at hospital	Administration of vasopressors/inotropes
7.	Arrived in the ED with uncontrolled haemorrhage	Administration of 4 or more units of blood products within 24 hours of admission, Any use of Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA), Interventional radiology/embolisation within 4 hours of hospital arrival
9.	Patient who required EMS initiation of CPR (i.e. had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	Any cardiopulmonary resuscitation in hospital (in the absence of cardiopulmonary resuscitation at scene)
		↓All no
	PRIORITY 3 (MINIMAL)	**TARN equivalent**
1.	Discharged from the ED with no X-rays or an extremity X-ray that was negative or showed an uncomplicated fracture (i.e. a closed extremity fracture without significant displacement or neurovascular compromise); no laboratory testing; received only simple wound repair (single layer suturing only); and received no medications intravenously (does not include fluids), or inhaled (does not include oxygen) from EMS or in the hospital	*Not included in TARN database (these patients would not meet TARN inclusion criteria)*
		↓Any no
	PRIORITY 2 (DELAYED)	**TARN equivalent**
1.	All remaining patients	All remaining TARN patients who do not fulfil any criteria for previous triage categories

Ledger: Cardiac arrest is defined as a lack of palpable pulse and EMS initiation of CPR. CPR=cardiopulmonary resuscitation, EMS=Emergency Medical Services, GCS=Glasgow Coma Scale, EDH=extradural haematoma, SDH=subdural haematoma, ICH=intracranial haemorrhage, LMA=laryngeal mask airway, CPAP=continuous positive airway pressure. Chemical, biological, radiological and neurological (CBRN) entities were excluded from triage category definitions as no CBRN patients exist within the TARN database. There is international variation in the names of triage categories with approximate equivalence as follows: Priority 1 (Immediate, Red, T1); Priority 2 (Delayed, T2, Urgent); Priority 3 (Minimal, Minor, T3); Expectant (Unsalvageable, T1 hold, P4); Dead (Black).

Supplementary Data Table 2: Comparison of Characteristics between Included and Excluded Patients

Patient and injury characteristics	Included, n (%)	Excluded, n (%)
	n=195,709	n=19,923
Gender		
Male	104,019 (53·1%)	11,823 (59·3%)
Female	91,690 (46·9%)	8,100 (40·7%)
Injury Severity Score (ISS)		
Median (IQR)	9 (9-17)	10 (9-22)
Age		
Median (IQR), years	66·2 (47·3-83·0)	58·7 (38·1-79·3)
Discharge status		
Alive	182,107 (93·0%)	17,501 (87·8%)
Dead	13,601 (7·0%)	2,422 (20·5%)
Mode of injury		
Blunt	190,048 (97·1%)	18,855 (94·6%)
Penetrating	5,660 (2·9%)	1,067 (9·0%)
Mechanism of injury		
Fall less than 2m	113,319 (57·9%)	9,016 (45·3%)
Vehicle Incident/Collision	41,590 (21·3%)	5,613 (28·2%)
Fall more than 2m	25,194 (12·9%)	2,659 (13·3%)
Blow(s)*	6,827 (3·5%)	784 (3·9%)
Stabbing	4,105 (2·1%)	806 (4·0%)
Other	2,609 (1·3%)	744 (3·7%)
Crush	1,355 (0·7%)	163 (0·8%)
Shooting	440 (0·2%)	69 (0·3%)
Blast	142 (0·1%)	25 (0·1%)
Burn	128 (0·1%)	44 (0·2%)

Supplementary Data Table 3: Tool Performance in Predicting ISS>15 in adults aged 16-64 years

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	56·7 [56·1, 57·2]	64·8 [64·5, 65·2]	45·5 [45·0, 46·0]	74·3 [73·9, 74·6]	43·3 [42·8, 43·9]	54·5 [54·0, 55·0]	0·608 [0·604, 0·611]
CareFlight	28·4 [28·0, 28·9]	94·7 [94·5, 94·9]	73·6 [72·8, 74·4]	71·9 [71·5, 72·2]	71·6 [71·1, 72·0]	26·4 [25·6, 27·2]	0·616 [0·613, 0·618]
JumpSTART	29·4 [28·9, 29·9]	89·8 [89·6, 90·0]	59·9 [59·2, 60·7]	71·1 [70·7, 71·4]	70·6 [70·1, 71·1]	40·1 [39·3, 40·8]	0·596 [0·593, 0·599]
MIMMS Triage Sieve	22·2 [21·8, 22·7]	94·6 [94·5, 94·8]	68·2 [67·3, 69·1]	70·1 [69·8, 70·4]	77·8 [77·3, 78·2]	31·8 [30·9, 32·7]	0·584 [0·582, 0·587]
MPTT	49·1 [48·5, 49·6]	61·0 [60·7, 61·4]	39·5 [39·0, 40·0]	69·8 [69·4, 70·2]	50·9 [50·4, 51·5]	60·5 [60·0, 61·0]	0·551 [0·547, 0·554]
MPTT-24	44·2 [43·6, 44·7]	65·3 [64·9, 65·6]	39·7 [39·2, 40·2]	69·3 [68·9, 69·6]	55·8 [55·3, 56·4]	60·3 [59·8, 60·8]	0·547 [0·544, 0·550]
MSTART	36·4 [35·9, 36·9]	90·5 [90·3, 90·7]	66·5 [65·8, 67·2]	73·3 [73·0, 73·6]	63·6 [63·1, 64·1]	33·5 [32·8, 34·2]	0·634 [0·632, 0·637]
NARU Triage Sieve	29·0 [28·5, 29·5]	90·7 [90·5, 90·9]	61·8 [61·0, 62·6]	71·1 [70·8, 71·4]	71·0 [70·5, 71·5]	38·2 [37·4, 39·0]	0·599 [0·596, 0·601]
RAMP	25·7 [25·3, 26·2]	94·9 [94·7, 95·1]	72·4 [71·6, 73·2]	71·1 [70·8, 71·4]	74·3 [73·8, 74·7]	27·6 [26·8, 28·4]	0·603 [0·601, 0·606]
START	33·8 [33·3, 34·3]	92·4 [92·2, 92·6]	69·9 [69·2, 70·6]	72·9 [72·6, 73·2]	66·2 [65·7, 66·7]	30·1 [29·4, 30·8]	0·631 [0·629, 0·634]

Supplementary Data Table 4: Tool Performance in Predicting ISS>15 in older adults (aged 65+ years)

Tool	Sensitivity	Specificity	PPV	NPV	Undertriage (1-sensitivity)	Overtriage (1-PPV)	AUC
BCD Triage Sieve	39·7 [39·1, 40·2]	73·1 [72·8, 73·5]	38·5 [37·9, 39·0]	74·1 [73·8, 74·4]	60·3 [59·8, 60·9]	61·5 [61·0, 62·1]	0·564 [0·561, 0·567]
CareFlight	18·3 [17·9, 18·8]	96·0 [95·9, 96·2]	66·2 [65·2, 67·2]	73·5 [73·2, 73·8]	81·7 [81·2, 82·1]	33·8 [32·8, 34·8]	0·572 [0·570, 0·574]
JumpSTART	17·2 [16·8, 17·6]	91·7 [91·5, 91·9]	46·7 [45·8, 47·6]	72·3 [72·1, 72·6]	82·8 [82·4, 83·2]	53·3 [52·4, 54·2]	0·545 [0·542, 0·547]
MIMMS Triage Sieve	12·5 [12·2, 12·9]	94·0 [93·8, 94·1]	46·8 [45·7, 47·9]	71·7 [71·4, 72·0]	87·5 [87·1, 87·8]	53·2 [52·1, 54·3]	0·533 [0·530, 0·535]
MPTT	41·8 [41·3, 42·4]	68·9 [68·5, 69·2]	36·3 [35·7, 36·8]	73·7 [73·3, 74·0]	58·2 [57·6, 58·7]	63·7 [63·2, 64·3]	0·554 [0·550, 0·557]
MPTT-24	36·7 [36·2, 37·3]	72·9 [72·6, 73·2]	36·5 [35·9, 37·0]	73·1 [72·8, 73·5]	63·3 [62·7, 63·8]	63·5 [63·0, 64·1]	0·548 [0·545, 0·551]
MSTART	24·7 [24·2, 25·2]	90·7 [90·5, 91·0]	53·0 [52·2, 53·8]	74·0 [73·7, 74·3]	75·3 [74·8, 75·8]	47·0 [46·2, 47·8]	0·577 [0·574, 0·580]
NARU Triage Sieve	17·9 [17·5, 18·4]	92·0 [91·8, 92·2]	48·6 [47·7, 49·6]	72·6 [72·3, 72·9]	82·1 [81·6, 82·5]	51·4 [50·4, 52·3]	0·550 [0·547, 0·552]
RAMP	17·5 [17·0, 17·9]	96·2 [96·1, 96·3]	66·1 [65·0, 67·1]	73·4 [73·1, 73·6]	82·5 [82·1, 83·0]	33·9 [32·9, 35·0]	0·568 [0·566, 0·571]
START	23·1 [22·6, 23·6]	92·3 [92·1, 92·5]	55·9 [55·0, 56·8]	73·9 [73·6, 74·2]	76·9 [76·4, 77·4]	44·1 [43·2, 45·0]	0·577 [0·574, 0·580]

Publication number 2: Paediatric major incident triage: UK military tool offers best performance in predicting the need for time-critical major surgical and resuscitative intervention

Author contributions:

NM conducted a literature review prior to the study. **NM**, DB, DK and GVG designed the study. **NM**, SC and YX accessed the database, verified the underlying data and conducted analysis. All authors contributed to data interpretation. **NM** wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

EClinicalMedicine 40 (2021) 101100

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research paper

Paediatric major incident triage: UK military tool offers best performance in predicting the need for time-critical major surgical and resuscitative intervention

Nabeela S. Malik[a,b,c,*], Saisakul Chernbumroong[a,d], Yuanwei Xu[d], James Vassallo[e], Justine Lee[b,f,g], Christopher G. Moran[g,h], Tina Newton[i], G. Suren Arul[c,i], Janet M. Lord[a,b], Antonio Belli[a,b,f], Damian Keene[f,j], Mark Foster[a,f,j], Timothy Hodgetts[k], Douglas M. Bowley[b,f,j], Georgios V. Gkoutos[a,d,l,m]

[a] NIHR Surgical Reconstruction and Microbiological Research Centre (SRMRC), Heritage Building, Queen Elizabeth Hospital, Mindelsohn Way, Edgbaston, Birmingham B15 2TH, United Kingdom
[b] Institute of Inflammation and Ageing, University of Birmingham, Birmingham B15 2TT, United Kingdom
[c] 212 (Yorkshire) Field Hospital, Endcliffe Hall, Endcliffe Vale Road, Sheffield S10 3EU, United Kingdom
[d] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom
[e] Academic Department of Military Emergency Medicine, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, United Kingdom
[f] University Hospitals Birmingham, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, United Kingdom
[g] NHS England London, Skipton House, 80 London Road, London SE1 6LH, United Kingdom
[h] Nottingham University Hospitals NHS Trust, Derby Road, Nottingham NG7 2UH, United Kingdom
[i] Birmingham Children's Hospital, Steelhouse Lane, Birmingham B4 6NH, United Kingdom
[j] Academic Department of Military Surgery & Trauma, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, United Kingdom
[k] Headquarters Defence Medical Services, Coltman House, Lichfield WS14 9PY, United Kingdom
[l] Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2TT, United Kingdom
[m] MRC Health Data Research UK (HDR UK), Midlands Site, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

Article History:
Received 2 June 2021
Revised 6 August 2021
Accepted 6 August 2021
Available online 23 August 2021

Keywords:
Major incident
Mass casualty
Prehospital medicine
Triage
Paediatric trauma
Major trauma
Disaster
Injury Severity Score
BCD Triage Sieve
CareFlight
START
MSTART
MPTT-24
RAMP
JumpSTART
MIMMS
Paediatric Triage Tape

ABSTRACT

Background: Children are frequently injured during major incidents (MI), including terrorist attacks, conflict and natural disasters. Triage facilitates healthcare resource allocation in order to maximise overall survival. A critical function of MI triage tools is to identify patients needing time-critical major resuscitative and surgical intervention (Priority 1 (P1) status). This study compares the performance of 11 MI triage tools in predicting P1 status in children from the UK Trauma Audit and Research Network (TARN) registry.

Methods: Patients aged <16 years within TARN (January 2008-December 2017) were included. 11 triage tools were applied to patients' first recorded pre-hospital physiology. Patients were retrospectively assigned triage categories (P1, P2, P3, Expectant or Dead) using predefined intervention-based criteria. Tools' performance in <16s were evaluated within four-yearly age subgroups, comparing tool-predicted and intervention-based priority status.

Findings: Amongst 4962 patients, mortality was 1.1% (n = 53); median Injury Severity Score (ISS) was 9 (IQR 9–16). Blunt injuries predominated (94.4%). 1343 (27.1%) met intervention-based criteria for P1, exhibiting greater intensive care requirement (60.2% vs. 8.5%, $p < 0.01$) and ISS (median 17 vs 9, $p < 0.01$) compared with P2 patients. The Battlefield Casualty Drills (BCD) Triage Sieve had greatest sensitivity (75.7%) in predicting P1 status in children <16 years, demonstrating a 38.4–49.8% improvement across all subgroups of children <12 years compared with the UK's current Paediatric Triage Tape (PTT). JumpSTART demonstrated low sensitivity in predicting P1 status in 4 to 8 year olds (35.5%) and 0 to 4 year olds (28.5%), and was outperformed by its adult counterpart START (60.6% and 59.6%).

Interpretation: The BCD Triage Sieve had greatest sensitivity in predicting P1 status in this paediatric trauma registry population: we recommend it replaces the PTT in UK practice. Users of JumpSTART may consider alternative tools. We recommend Lerner's triage category definitions when conducting MI evaluations.

Funding: This study is funded by the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre. GVG also acknowledges support from the MRC Heath Data Research UK

* Corresponding author at: NIHR Surgical Reconstruction and Microbiological Research Centre (SRMRC), Heritage Building, Queen Elizabeth Hospital, Mindelsohn Way, Edgbaston, Birmingham B15 2TH, United Kingdom.
E-mail address: nabeelamalik@nhs.net (N.S. Malik).

https://doi.org/10.1016/j.eclinm.2021.101100

JumpSTART
Life-saving inte

(HDRUK/CFC/01). The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the Ministry of Defence.

Research in context

In both ordinary times and during crises, strong primary health-care systems play a key role in saving lives and ensuring the continuity of health-care services. Primary health-care providers, such as general practitioners, counsellors, or pharmacists, represent the first point of contact between the health-care system and individuals, families, and communities. The role of primary care providers in providing people with the most comprehensive care possible, from prevention to treatment, is pivotal in ensuring basic health-care services and guaranteeing a health system that meets the needs of the population it serves.

The COVID-19 pandemic exacerbated pre-existing weaknesses in health-care access and delivery: primary care systems worldwide were not able to deflect the pressure from hospitals and, due to lock-down restrictions, were unable to fully support communities with maintaining essential health services. The pandemic has unequally affected many vulnerable groups, bringing the discussion around the social determinants of health and inequity to the forefront of public health. In a nutshell, access to health-care services is still not for all.

However, this is not a new issue. Heads of state across the globe met for the first time in 1978 in Alma-Ata, Kazakhstan, to commit to the ideal of health for all. The meeting recognised the existence of inequality in the access to health care between advantaged and disadvantaged individuals, both socioeconomically and geographically; acknowledged the situation as unacceptable; and endorsed primary health care as a core pillar to the attainment of health for all. The Alma-Ata declaration reinterpreted primary health care, envisioning it beyond hospitals and health services, but as a reflection of the social determinants of health. The declaration's main pillars were to build a primary health-care system close to all communities, regardless of their geographical location, to allow universal accessibility to individuals through their full participation and empowerment. In October, 2018, the commitment to health for all was reaffirmed by heads of state during the Global Conference on Primary Healthcare in Kazakhstan in the Declaration of Astana. This political re-affirmation, together with a renewed commitment globally towards universal health coverage and the development of Sustainable Development Goals, are welcome steps in the right direction. However, progress has been slow with the goal of Alma-Ata of 'health for all' unlikely to be met for many decades.

Joint data on access to health care released before the COVID-19 pandemic from the World Bank and WHO on Dec 13, 2017, worryingly highlighted that the trajectory to attain health for all was very far from being reached. At least half of the world's population lack access to essential health services and 800 million people spend at least 10% of their budget on health-care expenses. The same report also warned that a shortfall of 18 million health-care workers will occur by 2030, which will hamper universal health-care delivery and is likely to widen the gaps in health equity. Data from the Universal Health Coverage Global monitoring report published by WHO in 2019 reiterated these issues and warned that over 5 billion people will be unable to access health care by 2030. This report called countries for increased spending on primary care by at least 1% of the gross domestic product to close health coverage gaps and highlighted that more people were paying for services out of their own pocket, with 925 million people spending more than 10% of their household income on health-care expenses and 200 million people spending more than 25%. Although there is variation across world regions, it is evident that there is room for substantial improvement globally.

We must act now to reframe primary health care based on the Alma-Ata principles, tailoring strategies around the actual burden of diseases and the additional challenges brought about by the COVID-19 pandemic. It is time to carve policies that reflect patients' needs to build strong primary health-care systems globally that serve the needs of all populations. It is essential that the delivery of primary health-care services is reorganised, establishing a strong link between basic practices and community services, and expanding home-based programmes and strategies to reach isolated communities. In this context, the role of community health workers should be further utilised to provide timely information and direct access to care. In addition, the leverage of remote services and use of big data that were successfully exploited during the pandemic should be further implemented as tools to help maintain continuity of care and to develop tailored interventions. Lastly, and importantly, governments should invest in modernising health-care systems and supporting the delivery of essential services, an aim that cannot be reached without a plan oriented at increasing the number of health-care workers.

Many countries are still facing major challenges posed by the COVID-19 pandemic, which has shone an uncomfortable light on the health inequities that already existed. Unfortunately, these gaps in health-care access and delivery have also widened over the past 2 years. 'Health for all' is certainly one of the major public health challenges of the 21st century and currently looks unlikely to be met in the next few decades. It will require input and commitment from governments at a global level, but also immediate action to reframe primary care and bring it closer to people. 43 years have passed and it is finally time to transform the Alma-Ata vision into action.

1. Introduction

Children are often injured during major incidents (MI) including natural disasters, conflict and terrorist attacks, where their immediate needs exceed the resources available to treat them [1–4]. For example, following the 2017 Manchester Arena Bombing, children constituted 44/153 (29%) casualties attending various Emergency Departments (ED) [5] and 7/22 (32%) of those killed [2]. During MI, resources are best directed towards maximising overall survival amongst those affected. Selection of a triage tool for use at scene is an important aspect of disaster planning, enabling patients to be prioritised for treatment and onward transfer, particularly those in need of immediate life-saving intervention[1,3,4]. Children display age-dependent normal vital signs and, thus adult tools may assign an incorrect triage category when applied in children [6]. There is a natural tendency for first responders to assign a higher triage category to children [6]; uninjured ambulatory children are often conveyed to hospital from MIs, as reported following the Fairchild and Columbine School massacres [4]. Incorrect triage of children may fail to identify those needing urgent intervention (under-triage); however, assigning P1 status to children who do not require time-critical treatment (over-triage) risks overwhelming dedicated paediatric resources at scene and in hospital, and potentially directing resources away from others who require intervention more urgently [7]. As such, objective assessment of injured children using an appropriate triage tool is crucial to maximising overall survival following a MI.

The ideal MI triage tool is quick and simple to apply, with high sensitivity in identifying those for whom timely intervention is likely to alter overall outcomes (P1 patients) and an acceptably low rate of overtriage [8,9]. Two dedicated paediatric primary MI tools are in current use internationally. The Paediatric Triage Tape (PTT) (adapted from the adult MIMMS Triage Sieve in 1998) is applied to children <12 years in UK MIs [6]. PTT utilises physiological parameters relating to the height (or weight) of the child to determine the child's triage category. The US-based JumpSTART is a paediatric adaptation of the adult Simple Triage and Rapid Treatment (START) for use in children <8 years [10]. The Australian CareFlight triage tool has been applied in both adults and children, achieving good performance (AUC 0.852) in predicting mortality to discharge [11] and the need for intervention [12,13]. Although several other adult triage tools have been developed and are in use, their performance in children remains largely unvalidated [1]. These include the Major Incident Medical Management and Support (MIMMS) Triage Sieve [14], the Modified Physiological Triage Tool (MPTT) [15], its derivative MPTT-24 [15], the modified START (MSTART) [14], National Ambulance Resilience Unit (NARU) Triage Sieve [16] and the US-based Rapid Assessment of Mentation and Pulse (RAMP) [17]. The Battlefield Casualty Drills (BCD) Triage Sieve, used by British soldiers faced with multiple casualties, first appeared in 1998 [18], and was updated in 2018 in line with emerging evidence and changes in clinical practice (Fig. 1). The BCD Triage Sieve was recently identified as the most sensitive of multiple MI triage tools in an adult population [18].

Few studies exist to inform the choice of paediatric MI triage tools [1,11,19]. One challenge in interpreting existing studies lies with variation in age used to define a child [11,19,20]. Whilst the UK usually employs age <16 years as the cut-off for paediatric healthcare services, MI casualty distribution planning in some regions stipulates that children aged 12 to 16 years should be conveyed to adult facilities in order to preserve specialist paediatric services for the youngest patients. However no studies have examined tool performance in the 12 to 16 year old subgroup; in whom the NARU Triage Sieve would be applied (UK practice) [16]. Studies measuring triage tool performance have often focussed on predicting an Injury Severity Score (ISS) >15 as the end-point [13,20], despite a lack of correlation between ISS and requirement for medical resources [21]. There is growing consensus that the ability to predict requirement for urgent life-saving resuscitative and/or surgical intervention is the most meaningful measure of performance in MI triage tools [9,13,22]. Our primary aim was to determine which tool performs best in children (<16 years), in order to inform UK policy. A secondary aim was to analyse the performance of tools in subgroups of children by age, in order to determine the appropriateness of the age cut-offs applied by the paired adult and paediatric tools, namely the UK's Paediatric Triage Tape (<12 years) with the NARU Triage Sieve [6,16], and the US JumpSTART (below eight years) with START [10].

2. Methods

2.1. Overview of study design

This study utilises the physiology and outcomes of injured children within the UK national Trauma Audit and Research Network (TARN) as a surrogate for those injured in a MI. Two paediatric and nine adult triage tools have been applied to each patient's first recorded pre-hospital physiology. Patient records have been assessed to determine which triage category they would fulfil (P1, P2, Expectant or Dead) on the basis of required interventions, using pre-defined (Lerner's) criteria [9]; Priority 1 status was defined as patients requiring time-critical, major resuscitative and/or surgical intervention(s). Tool performance is reported against intervention-based Priority 1 status.

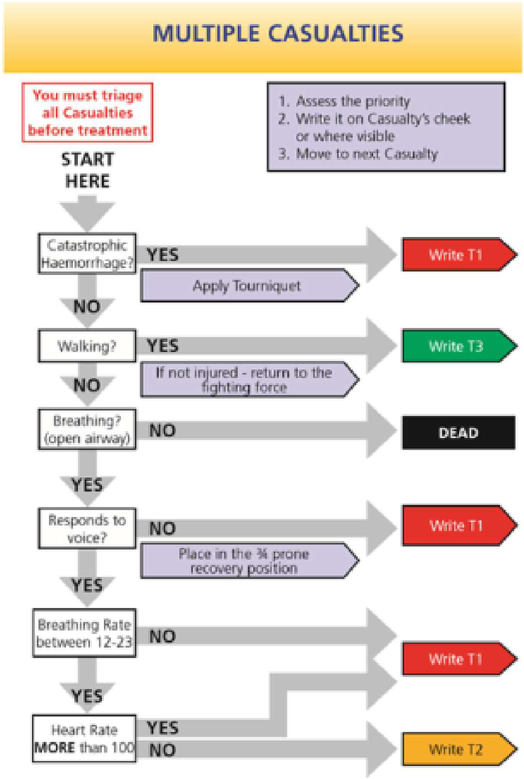


Fig. 1. Battlefield Casualty Drills Triage Sieve (UK Military Primary Triage Tool).

2.2. Study population

TARN co-ordinators capture pre-hospital, clinical management and outcome data from 169 trauma receiving hospitals in England and Wales, including all paediatric major trauma centres, constituting the largest trauma registry in Europe [23]. TARN includes injured patients attending hospital with length of stay over 48 h, intensive care admission and/or in-hospital death [23]. Pre-hospital deaths are excluded.

All patients aged <16 years submitted to TARN by hospitals in England and Wales between 1 January 2008 and 31 December 2017 were included. Those patients missing pre-hospital physiological data required for tool application (respiratory rate, heart rate, systolic blood pressure, Glasgow Coma Score (GCS), and GCS Motor Component) were excluded.

2.3. Application of triage tools

The BCD Triage Sieve (Fig. 1), CareFlight [14], JumpSTART [10], MIMMS Triage Sieve [14], MPTT [15], MPTT-24 [15], MSTART [14], NARU Triage Sieve[16], RAMP[17], START[14] and PTT[6] tools (see Table 1) were converted into computer code: these were verified by clinician co-authors and by application in an adult population [18].

Table 1
Summary of triage tool characteristics.

Tool	Description and geographical use	Tool components							
		1st step	2nd step	3rd step	4th step	5th step	6th step	7th step	Interventions permitted
Battlefield Casualty Drills (BCD) Triage Sieve	Current UK military tool for use in adults (introduced in 1998, revised 2018).	Catastrophic haemorrhage?	Walking?	Breathing?	Responds to voice?	Breathing rate between 12 and 23	Heart Rate more than 100	–	Apply tourniquet, open airway, place casualty in the ¾ prone recovery position
CareFlight	Australian tool used in adults and children (introduced in 2001).	Walks?	Obeys command?	Palpable radial pulse? OR Breathes with open airway?	–	–	–	–	Open airway
Jump Simple Triage and Rapid Treatment (JumpSTART)	United States, used in several states in children (introduced in 2001).	Able to walk?	Spontaneous breathing (check radial pulse if apnoeic)	Respiratory rate <15 or >45	Palpable pulse?	Neurological Assessment (AVPU)	–	–	Airway positioning, 5 rescue breaths if apnoeic
Major Incident Medical Management and Support (MIMMS) Triage Sieve	Former UK military adult triage tool (introduced in 1995).	Walking	Breathing	Respiratory rate <10 or ≥30	Capillary refill >2 s	–	–	–	Open airway
Modified Physiological Triage Tool (MPTT)	UK-based tool* modelled in a military cohort (described in 2017).	Walking?	Breathing?	Respiratory rate <12 or ≥22	Heart rate ≥100	GCS <14	–	–	–
Modified Physiological Triage Tool 24 (MPTT-24)	UK-based tool*, modification of MPTT (described in 2017).	Catastrophic Haemorrhage?	Walking?	Breathing?	Responds to voice	Respiratory rate <12 or ≥24	Heart rate ≥100	–	Apply tourniquet or haemostatic dressing
Modified Simple Triage and Rapid Treatment (MSTART)	United States, modification of START (described in 2006).	Able to walk?	Spontaneous breathing	Respiratory rate >30	Radial pulse absent	Obey commands	–	–	Position airway
National Ambulance and Resilience Unit (NARU) Triage Sieve	Current UK civilian adult tool, adapted from the MIMMS Triage Sieve (this version was introduced in 2013)	Catastrophic haemorrhage	Are they injured	Walking	Breathing	Unconscious	Respiratory rate <10 or ≥30	Pulse >120 or capillary refill >2 s	Apply tourniquet/haemostatic dressing, open airway, place in recovery position
Paediatric Triage Tape (PTT)	Current UK paediatric tool (<12 years) (adapted from MIMMS Triage Sieve in 1998).	Alert and moving all limbs (children <100 cm height) or Walking	Use tape to gauge child's length in order to determine which set of physiological values to compare the child against	Breathing?	Respiratory rate (height-specific threshold)	Capillary refill <2 s (use child's forehead)	Pulse rate (heigh-specific threshold)	–	Position airway
Rapid Assessment of Mentation and Pulse (RAMP)	United States, used by the Rocky Mountain Fire Department, Colorado (introduced in 2018).	Casualty without signs of obvious death	Casualty follows commands	Radial pulse present?	–	–	–	–	Control massive haemorrhage, open airway, chest decompression
Simple Triage and Rapid Treatment (START)	United States (introduced in 1983).	Able to walk?	Spontaneous breathing	Respiratory rate >30	Capillary refill >2 s	Obey commands	–	–	Position airway

Ledger: AVPU refers to the Alert, Voice, Pain, Unresponsive scale; GCS=Glasgow Coma Score. All tools described are applicable at the scene of a major incident (primary triage tools). *Has yet to undergo practical use or implementation studies. SALT and ASAV were not evaluated in this study as there were major limitations in applying these retrospectively. **SALT involves sorting according to the following: walk, wave/purposeful movement, still/obvious life threat; as well as the subjective judgements: "Minor injuries only?" and "Likely to survive given current resources?" ***ASAV includes the subjective judgement "Deadly injured?" and assessment of breathing status as follows: "airway obstructed, bradypnoea, apnoea, dyspnoea, tachypnoea (not obviously psychogenic) and cyanosis."

Tool codes were applied to patients' first recorded pre-hospital physiology (assuming that these preceded any intervention), to determine whether patients would be assigned P1 or non-P1 status, as per a recent adult study methodology [18].

Where tools employed parameters not recorded in TARN, approximations were made based on available information. TARN patients were assumed to be non-ambulatory. Those who underwent advanced airway interventions at scene were considered unable to breathe [24]. A respiratory rate below four breaths per minute was deemed undetectable by EMS personnel. No approximation for the term "catastrophic haemorrhage" (utilised by MPTT-24, BCD and NARU Triage Sieve) could be identified, hence this term was not applied. Children with a systolic blood pressure of ≥60 mmHg (<12 years) or ≥90 mmHg (≥12 years), were regarded as having a palpable radial pulse [25]. Patients with a GCS of ≤8 were deemed unconscious, those with a GCS <12 were deemed unresponsive to voice [26]. GCS Motor Score of six indicated ability to follow commands. For JumpSTART, a GCS Motor Score ≤3 was equated to "inappropriate response to painful stimulus (e.g. posturing) or unresponsive to noxious stimulus" [26].

Tool performance was measured in children <16 years and in subgroups based on age: 0 to <4 years (pre-school), 4 to <8 years, 8 to <12 years and age 12 to <16 years. These subgroups were selected in line with thresholds employed by the dedicated paediatric tools (PTT <12 years, JumpSTART <8 years).

2.4. Outcome measures

The primary outcome was the ability of triage tools to predict P1 status, defined as the need for any one or more of eight time-critical major resuscitative or surgical interventions (Table 2)[9]. Each patient was assigned a triage category (Dead, Expectant, Priority 1 [P1], Priority 2 [P2] or Priority 3[P3]) based on a pre-defined system utilising EMS and hospital-based interventions described by Lerner et al., using TARN terminology which best matched each criterion (see Supplementary Data Table 1). Since TARN does not include patients with chemical, biological, radiological, and nuclear injuries, criteria relevant to these mechanisms were not considered [9]. Two further paediatric-specific measures for "presented to ED with uncontrollable haemorrhage" were included: administration of a fluid bolus of 20 ml/kg within an hour of arrival in ED [22] and/or the requirement for blood products within an hour of ED arrival. In order to calculate the fluid bolus volume, weight was estimated using age as recorded by TARN and World Health Organisation male and female charts for infants up to 12 months [27], or the formula (age+2)x4 for children aged >12 months [28].

TARN records the timing of hospital arrival and each intervention, allowing incorporation of this into the time-critical definitions constituting P1 status. To assess the validity of Lerner's classification, patients within each category were compared by mortality, ICU admission, hospital LOS and ISS.

Secondary outcome measures included prediction of mortality and ISS>15 (see Supplementary data Table 2 and 3), and distribution of ISS amongst tool-assigned P1 patients (Fig. 2), which may provide further discriminative value and appreciation of tool characteristics.

Table 2
Breakdown of time-critical major operative and resuscitative interventions constituting Priority 1 status.

Subcomponents of the Priority 1 triage category	n (%)
An advanced airway intervention (e.g. intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 h of arrival at hospital	808 (75·2%)
Arrived in the ED with uncontrolled haemorrhage	310 (28·9%)
Neurological, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 h of arrival to hospital	169 (15·7%)
Chest tube placed within 2 h of arrival at hospital	70 (6·5%)
Limb-conserving surgery performed within 4 h of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	23 (2·1%)
IV vasopressors administered within 2 h of arrival at hospital	6 (0·6%)
Patient who required EMS initiation of CPR (i.e. had a cardiac arrest) during transport, in the ED, or within 4 h of arrival at a hospital	1 (0·1%)
Escharotomy performed on a patient with burns within 2 h of arrival at a hospital	1 (0·1%)
Total number of P1 patients	1343 (100·0%)

Ledger: There is overlap between life-saving interventions (LSI): 73·2% (*n* = 1006) of P1 patients required one LSI, 21·8% (*n* = 299) required two LSI, and 5·0% (*n* = 69) required 3 or more LSI.

Table 3
Summary of patient and injury characteristics.

Variable		
Gender, *n* (%)	Male	3447 (69·5%)
	Female	1515 (30·5%)
	Missing data	0 (0·0%)
Injury Severity Score, median (IQR)		9 (9, 17)
	Missing data	0
Age (years)	Median (IQR)	11·9 (8·0, 14·2)
	<16 years, n (%)	4962 (100·0%)
	0 to <4	467 (9·4%)
	4 to <8	768 (15·5%)
	8 to <12	1281 (25·80%)
	12 to <16	2446 (49·3%)
	Missing data	0 (0·0%)
Outcome at discharge, *n* (%)	Alive	4909 (98·9%)
	Dead	53 (1·1%)
	Missing data	0 (0·0%)
Injury type, *n* (%)	Blunt	4733 (95·4%)
	Penetrating	229 (4·6%)
	Missing data	0 (0·0%)
Injury mechanism, *n* (%)	Vehicle incident/collision	2459 (49·6%)
	Fall less than 2m	1187 (23·9%)
	Fall more than 2m	645 (13·0%)
	Blow(s)	327 (6·6%)
	Stabbing	130 (2·6%)
	Crush	42 (0·9%)
	Shooting	16 (0·3%)
	Blast	5 (0·1%)
	Burn	4 (0·1%)
	Other	147 (3·0%)
	Missing data	0 (0·0%)

2.5. Data processing and analyses

TARN data were received in SPSS Version 24·0 (Armonk NY: IBM Corp 2015) and processed using Python (Version 3.7.4) and R software (Version 3·6, R Core Team, New Zealand, 2000). Non-parametric data are presented as median and interquartile range; categorical data as frequency and percent. D'Agostino and Pearson's test was used to confirm the non-parametric nature of data distribution [29]. Differences between P1 and P2 patients as designated by Lerner's criteria [9] were compared using the Chi-squared test (mortality and ICU admission) and Mood's median test (ISS). Performance characteristics included sensitivity, specificity, under-triage (1-sensitivity) and over-triage (1-positive predictive value). Area Under the receiver operating Curve (AUC) was calculated using the trapezoidal rule [30]. 95% confidence intervals were calculated using the Wilson Score with continuity correction for binomial proportions, and DeLongs Algorithm for comparing AUC curves [31]. Included patients were compared to those excluded with respect to clinical and demographic characteristics. Fisher's exact test was used to compare categorical variables (gender, mortality and mode of injury (blunt and penetrating)). Continuous variables were compared using a Two-sample

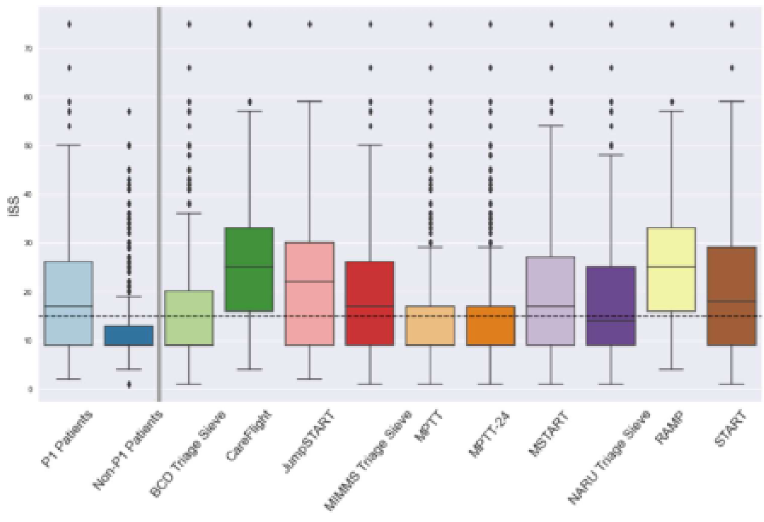


Fig. 2. Distribution of ISS amongst tool P1 patients (patients aged ≤16 years) Ledger: ISS=Injury Severity Score. Dotted horizontal line denotes ISS 15. The upper whisker extends from the hinge to the largest value no further than 1•5 * IQR from the hinge; the lower whisker extends from the hinge to the smallest value, at most 1•5 * IQR of the hinge.

Table 4
Comparison of the characteristics of included versus excluded patients.

Variable	Included patients	Excluded patients
Male gender, *n* (%)	3447 (69·5%)*	6847 (67·3%)
ISS, median (IQR)	9 (9, 17)*	9 (9, 16)
Age (years), median (IQR)	11·9 (8·0, 14·2)*	3·9 (1·6, 10·1)
Mortality, *n* (%)	53 (1·1%)	316 (3·1%)*
Injury type		
Blunt	4733 (95·4%)	9935 (97·7%)*
Penetrating	229 (4·6%)*	236 (2·3%)
Injury mechanism		
Vehicle incident/collision	2459* (49·6%)	2262 (22·2%)
Fall less than 2m	1187* (23·9%)	4827 (47·5%)
Blow(s)	327 (6·6%)	943 (9·3%)
Crush	42 (0·9%)	118 (1·1%)
Fall more than 2m	645* (13·0%)	869 (8·5%)
Other	147 (3·0%)	971* (9·6%)
Burn	4 (0·1%)	40* (0·4%)
Stabbing	130* (2·6%)	99 (1·0%)
Shooting	16 (0·3%)	26 (0·3%)
Blast	5 (0·1%)	16 (0·2%)

Ledger: Patients were excluded on the basis of insufficient pre-hospital physiological data required to apply the tools (see Methods). OR=Odds ratio. Percentages represent the proportion of patients within the included (or excluded) group with the characteristic described (e.g. 69·5% of all included patients were of male gender). * indicates the group that has higher number of incidents than expected, and is statistically significant ($p < 0·05$).

Kolmogorov-Smirnov Test. Differences in injury mechanism were estimated using the Chi-square test, where results were significant, post-hoc tests were performed to generate a p value. P values were adjusted using the Bonferroni correction. A value of $p < 0·05$ was considered statistically significant.

Ethical approval: The UK Health Research Authority Patient Information Advisory Group (Section 20) has granted ethical approval and waived the requirement for individual patient consent for research using anonymised TARN data.

Role of Funding: The funding source played no role in study design; in data collection, analysis or interpretation; in the writing of the report; or decision to submit the paper for publication.

3. Results

3.1. Characteristics of the study population

Of the 15,133 TARN patients identified, 10,171 (67·2%) patients were excluded due to incomplete pre-hospital physiological data (see Supplementary Data: Analysis of missing data), therefore, 4962 patients were included.

Patient and injury characteristics are presented in Table 3. Two thirds (69·5%) of patients were male, half (49·3%) were aged 12 to 16 years, whilst less than 10% (n = 467) were aged under four years. Mortality was 1·1% (53/4962), median ISS was 9 (IQR 9–17). 94·4% (n = 4733) of patients suffered blunt injuries, mainly comprising vehicle collisions (n = 2459, 49·6%) and falls under two metres (n = 1187,

Table 5
Comparison of outcome characteristics between patients in each triage category.

Triage category	Total, n (%)	Mortality, n (%)	ICU admission, n (%)	LOS (days), median (IQR)	ISS, median (IQR)
Dead	2 (0·04%)	2 (100·0%)	1 (50·0%)	11 [6, 15]	22 [21, 24]
Expectant	29 (0·58%)	29 (100·0%)	27 (93·1%)	1 [1, 3]	41 [29, 50]
Priority 1 (Immediate)	1343 (27·1%)	19 (1·4%)	809 (60·2%)	7 [3, 15]	17 [9, 26]
Priority 2 (Urgent)	3588 (72·3%)	3 (0·08%)	304 (8·5%)	5 [3, 9]	9 [9, 12]

Ledger: ISS=Injury Severity Score, IQR=interquartile range, LOS=Length of stay, ICU=Intensive Care Unit.

Table 6a
Triage tool performance in predicting Priority 1 status in children ≤16 years (the need for time-critical major operative or resuscitative measures).

Age group	Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
All (<**16 years**)	BCD Triage Sieve	75·7 (73·3, 78·0)	42·0 (40·3, 43·5)	24·3	67·4	0·588 (0·571, 0·606)
	CareFlight	40·4 (37·8, 43·1)	94·8 (94·0, 95·5)	59·6	25·7	0·676 (0·661, 0·692)
	JumpSTART	35·5 (33·0, 38·2)	93·5 (92·6, 94·3)	64·5	33·1	0·645 (0·629, 0·661)
	MIMMS Triage Sieve	37·5 (34·9, 40·2)	87·8 (86·7, 88·9)	62·5	46·7	0·627 (0·610, 0·643)
	MPTT	59·2 (56·5, 61·8)	34·8 (33·3, 36·4)	40·8	74·8	0·470 (0·452, 0·488)
	MPTT-24	56·7 (54·0, 59·3)	39·2 (37·6, 40·8)	43·3	74·3	0·479 (0·461, 0·497)
	MSTART	50·9 (48·2, 53·6)	88·0 (86·9, 89·0)	49·1	38·9	0·695 (0·679, 0·709)
	NARU Triage Sieve	48·3 (45·6, 51·0)	78·8 (77·4, 80·1)	51·7	54·2	0·636 (0·619, 0·652)
	PTT*	44·8 (40·9, 48·7)	87·2 (85·6, 88·7)	55·2	45·5	0·660 (0·637, 0·683)
	RAMP	39·7 (37·1, 42·4)	94·9 (94·2, 95·6)	60·3	25·6	0·673 (0·657, 0·689)
	START	49·2 (46·5, 51·9)	88·6 (87·6, 89·7)	50·8	38·3	0·689 (0·674, 0·705)

Ledger: BCD Triage Sieve =Battlefield Casualty Drills Triage Sieve, MIMMS Triage Sieve=Major Incident Medical Management System Triage Sieve, MPTT=Modified Physiological Triage Tool, MSTART=Modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve, RAMP=Rapid Assessment of Mentation and Pulse, START=Simple Triage and Rapid Treatment, PTT=Paediatric Triage Tape; *The PTT is only applicable to those under 12 years (n = 2516, 50·7%)·.

Table 6b
Triage tool performance in predicting Priority 1 status in children ≤12 years (the need for time-critical major operative or resuscitative measures).

Age group	Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
All (<**12 years**)	BCD Triage Sieve	85·6 (82·6, 88·2)	28·3 (26·2, 30·4)	14·4	71	0·570 (0·545, 0·595)
	CareFlight	43·7 (39·8, 47·6)	95·5 (94·4, 96·3)	56·3	23·3	0·696 (0·674, 0·710)
	JumpSTART	35·6 (31·9, 39·4)	95·6 (94·6, 96·5)	64·4	26·5	0·656 (0·633, 0·679)
	MIMMS Triage Sieve	45·7 (41·8, 49·7)	82·4 (80·6, 84·1)	54·3	53	0·641 (0·617, 0·664)
	MPTT	66·5 (62·6, 70·1)	21·5 (19·7, 23·5)	33·5	77·5	0·440 (0·413, 0·466)
	MPTT-24	64·1 (60·3, 67·8)	26·2 (24·2, 28·3)	35·9	77·1	0·452 (0·425, 0·478)
	MSTART	59·8 (55·8, 63·6)	85·1 (83·4, 86·7)	40·2	42·1	0·724 (0·704, 0·745)
	NARU Triage Sieve	60·2 (56·3, 64·0)	70·4 (68·3, 72·4)	39·8	59	0·653 (0·630, 0·676)
	PTT	44·8 (40·9, 48·7)	87·2 (85·6, 88·7)	55·2	45·5	0·660 (0·637, 0·683)
	RAMP	43·7 (39·8, 47·6)	95·5 (94·4, 96·3)	56·3	23·3	0·696 (0·674, 0·718)
	START	59·6 (55·7, 63·4)	85·1 (83·4, 86·7)	40·4	42·2	0·724 (0·703, 0·744)

Ledger: BCD Triage Sieve =Battlefield Casualty Drills Triage Sieve, MIMMS Triage Sieve=Major Incident Medical Management System Triage Sieve, MPTT=Modified Physiological Triage Tool, MSTART=Modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve, RAMP=Rapid Assessment of Mentation and Pulse, START=Simple Triage and Rapid Treatment, PTT=Paediatric Triage Tape;.

23·9%). Penetrating trauma constituted only 4·6% (n = 229), mainly stabbing (2·6%, n = 130).

A comparison between included and excluded patients is shown in Table 4. Excluded patients had a comparable injury ISS (median 9 [IQR9–16] vs. median ISS 9 [IQR9–17], $p < 0·01$, respectively) and a higher mortality (3·1% vs. 1·1%, $p < 0·01$) relative to included patients. Excluded patients were more likely to have suffered burns and falls below two metres.

3.2. Intervention-based designation of triage categories (using Lerner's criteria)

Fewer than 1% of patients (n = 31) met criteria for the "Dead" and "Expectant" category, with universal mortality across both groups (Table 5). 1343 patients fulfilled the criteria for P1 status: three quarters (n = 808) required advanced airway intervention, 28·9% (n = 310) arrived in ED with uncontrolled haemorrhage and 15·7% (n = 169) required time-critical major surgical intervention (Table 2). The remaining patients (n = 3588, 72·3%) were designated P2, representing the largest triage category. By virtue of TARN's inclusion criteria, no patients met criteria for the P3 (minor) category (Supplementary data Table 1).

Patients assigned P1 based on Lerner's criteria suffered higher in-hospital mortality (1·41% vs. 0·08%, $p < 0·01$), had longer LOS (median 7 vs. 5 days, $p < 0·01$), suffered more severe injuries (median ISS 17 [IQR 9–26] vs. 9[IQR9–12], $p < 0·01$), and were seven times more likely to require ICU admission (60·2% vs. 8·5%, $p < 0·01$) than patients designated P2.

3.3. Triage tool performance

Tool prediction of P1 status in all children <16 is shown in Table 6a. Overall, the BCD Triage Sieve demonstrated the highest sensitivity (75·8%), with an over-triage rate of 67·4%. The PTT had a much lower sensitivity at 44·8%, with over-triage of 45·5%. CareFlight and RAMP had very similar performance characteristics in <16 s, achieving the highest specificity (over 94%) and lowest over-triage rates (25·7% and 25·6%, respectively) of all the tools. MPTT and MPTT-24 exhibited the second highest sensitivity (59·2% and 56·7%, respectively), however these tools also exhibited the highest over-triage rates (74·8% and 74·3%). For comparison, tool performance in children <12 is shown in Table 6b.

Tool prediction of P1 status in children within each four yearly age subgroups is demonstrated in Table 7. The BCD Triage Sieve exhibited the highest sensitivity (66·7–90·2%) in all subgroups of children, demonstrating a 38·4–49·8% higher sensitivity in detecting P1 status in all subgroups of children <12 years as compared with the PTT. The BCD Triage Sieve exceeded the sensitivity of the NARU Triage Sieve in 12–16 year olds by 29·2% (sensitivity of 66·7% vs 37·5%). Amongst 12–16 years olds, CareFlight demonstrated identical sensitivity to the NARU Triage Sieve (37·5%) but with markedly lower over-triage (28·2% vs. 44·8%, respectively).

Table 7
Tool performance by age subgroup.

Age group	Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
12–16 years*	BCD Triage Sieve	66·7 (63·0, 70·1)	56·7 (54·3, 59·0)	33·3	61·8	0·616 (0·593, 0·640)
	CareFlight	37·5 (33·9, 41·2)	94·1 (92·9, 95·1)	62·5	28·1	0·658 (0·635, 0·680)
	JumpSTART	35·5 (32·0, 39·2)	91·2 (89·7, 92·4)	64·5	38·2	0·633 (0·610, 0·656)
	MIMMS Triage Sieve	30·1 (26·7, 33·6)	93·6 (92·4, 94·7)	69·9	34·5	0·619 (0·595, 0·642)
	MPTT	52·6 (48·8, 56·3)	49·1 (46·7, 51·5)	47·4	70·6	0·508 (0·483, 0·534)
	MPTT-24	49·9 (46·1, 53·6)	53·2 (50·8, 55·5)	50·1	70·0	0·515 (0·490, 0·540)
	MSTART	42·9 (39·2, 46·6)	91·1 (89·6, 92·3)	57·1	34·1	0·670 (0·647, 0·692)
	NARU Triage Sieve	37·5 (33·9, 41·2)	87·8 (86·1, 89·3)	62·5	44·8	0·626 (0·603, 0·650)
	RAMP	36·0 (32·5, 39·7)	94·4 (93·2, 95·4)	64·0	27·9	0·652 (0·629, 0·675)
	START	39·7 (36·1, 43·5)	92·4 (91·1, 93·6)	60·3	32·1	0·661 (0·638, 0·683)
8–12 years	BCD Triage Sieve	82·6 (77·6, 86·7)	38·8 (35·8, 42·0)	17·4	72·0	0·607 (0·572, 0·642)
	CareFlight	44·6 (38·8, 50·6)	96·4 (95·0, 97·4)	55·4	22·0	0·705 (0·674, 0·736)
	JumpSTART	39·4 (33·7, 45·3)	95·5 (93·9, 96·6)	60·6	28·5	0·674 (0·642, 0·707)
	MIMMS Triage Sieve	44·2 (38·5, 50·2)	89·94(87·9, 91·7)	55·8	44·1	0·671 (0·638, 0·704)
	MPTT	61·0 (55·1, 66·6)	29·2 (26·4, 32·1)	39·0	80·1	0·451 (0·412, 0·489)
	MPTT-24	57·8 (51·9, 63·6)	35·9 (32·9, 39·0)	42·2	79·3	0·468 (0·431, 0·507)
	MSTART	58·9 (52·9, 64·6)	90·1 (88·1, 91·9)	41·1	36·7	0·745 (0·716, 0·774)
	NARU Triage Sieve	51·9 (45·9, 57·8)	81·5 (78·9, 83·8)	48·1	55·3	0·667 (0·634, 0·700)
	PTT	46·7 (40·8, 52·6)	90·4 (88·4, 92·2)	53·3	41·5	0·686 (0·654, 0·718)
	RAMP	44·6 (38·8, 50·6)	96·4 (95·0, 97·4)	55·4	22·0	0·705 (0·674, 0·736)
	START	58·9 (52·9, 64·6)	90·1 (88·1, 91·9)	41·1	36·7	0·745 (0·716, 0·774)
4 –8 years	BCD Triage Sieve	90·2 (85·0, 93·7)	14·8 (12·1, 18·1)	9·9	72·4	0·525 (0·479, 0·571)
	CareFlight	44·3 (37·4, 51·5)	97·2 (95·3, 98·3)	55·7	15·1	0·708 (0·669, 0·746)
	JumpSTART	35·5 (28·9, 42·5)	97·9 (96·2, 98·9)	64·5	14·3	0·667 (0·626, 0·707)
	MIMMS Triage Sieve	46·3 (39·3, 53·4)	81·6 (78·1, 84·6)	53·7	52·5	0·640 (0·597, 0·682)
	MPTT	72·9 (66·2, 78·8)	10·4 (8·1, 13·3)	27·1	77·4	0·417 (0·370, 0·464)
	MPTT-24	70·9 (64·1, 77·0)	12·9 (10·3, 16·0)	29·1	77·4	0·419 (0·373, 0·466)
	MSTART	61·1 (54·0, 67·8)	87·1 (84·0, 89·7)	38·9	37·1	0·741 (0·705, 0·777)
	NARU Triage Sieve	63·6 (56·6, 70·1)	65·5 (61·4, 69·4)	36·5	60·2	0·645 (0·603, 0·687)
	PTT	40·4 (33·7, 47·5)	91·0 (88·2, 93·2)	59·6	38·4	0·657 (0·616, 0·698)
	RAMP	44·3 (37·4, 51·5)	97·2 (95·3, 98·3)	55·7	15·1	0·708 (0·669, 0·746)
	START	60·6 (53·5, 67·3)	87·1 (84·0, 89·7)	39·4	37·2	0·738 (0·702, 0·775)
0–4 years	BCD Triage Sieve	85·4 (78·6, 90·5)	19·0 (14·9, 23·8)	14·6	66·5	0·522 (0·466, 0·578)
	CareFlight	41·1 (33·2, 49·4)	89·6 (85·5, 92·6)	58·9	34·7	0·653 (0·602, 0·704)
	JumpSTART	28·5 (21·6, 36·5)	92·1 (88·4, 94·7)	71·5	36·8	0·603 (0·550, 0·656)
	MIMMS Triage Sieve	47·7 (39·6, 55·9)	60·1 (54·5, 65·5)	52·3	63·6	0·539 (0·484, 0·594)
	MPTT	68·2 (60·1, 75·4)	17·4 (13·5, 22·1)	31·8	71·7	0·428 (0·372, 0·484)
	MPTT-24	66·9 (58·7, 74·2)	19·3 (15·2, 24·2)	33·1	71·6	0·431 (0·375, 0·487)
	MSTART	59·6 (51·3, 67·4)	65·8 (60·3, 71·0)	40·4	54·6	0·627 (0·575, 0·679)
	NARU Triage Sieve	71·5 (63·5, 78·4)	44·3 (38·8, 50·0)	28·5	62·0	0·579 (0·525, 0·633)
	PTT	47·0 (38·9, 55·3)	70·3 (64·8, 75·2)	53·0	57·0	0·586 (0·533, 0·640)
	RAMP	41·1 (33·2, 49·4)	89·6 (85·5, 92·6)	58·9	34·7	0·653 (0·602, 0·704)
	START	59·6 (51·3, 67·4)	65·8 (60·3, 71·0)	40·4	54·6	0·627 (0·575, 0·679)

Ledger: BCD Triage Sieve =Battlefield Casualty Drills Triage Sieve, MIMMS Triage Sieve=Major Incident Medical Management System Triage Sieve, MPTT=Modified Physiological Triage Tool, MSTART=Modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve, RAMP=Rapid Assessment of Mentation and Pulse, START=Simple Triage and Rapid Treatment, PTT=Paediatric Triage Tape. *The PTT is only applicable to those under 12 years (n = 2516, 50·7%).

CareFlight and RAMP exhibited very similar performance characteristics to each other as well as the most consistent performance across all age subgroups, with sensitivities in the region of 36·0 to 44·6% and specificity consistently ≥90%.

JumpSTART demonstrated low sensitivity in predicting P1 status in 4 to 8 year olds (35·5%) and 0 to 4 year olds (28·5%), accompanied by high specificity (over 90%) and therefore low over-triage rates (14·3% and 36·8%, respectively). However, START demonstrated nearly double the sensitivity (60.6% and 59.6%) in both these subgroups.

1617 of all patients <16 years had ISS>15; of these, only 52·5% (n = 849) met criteria for intervention-based P1 status. One third (n = 494) of intervention-based P1 patients had an ISS≤15. Amongst the tools, the BCD Triage Sieve exhibited the highest sensitivity in predicting mortality (94·3%) and ISS>15 (73·6%) (Supplementary data Tables 2 and 3).

4. Discussion

Care of children during MIs is challenging and emotive, and specialist paediatric trauma resources are less available than adult services. As such, the objective and accurate triage of children in MIs is vital to ensure that healthcare resources are appropriately allocated. This study has assessed the performance of 11 primary MI triage tools using data from 4962 injured children from the UK national TARN registry. The PTT, currently employed by UK ambulance services for use in children <12 years, correctly identified only 45% of children requiring time-critical major resuscitative and surgical interventions (P1 patients); whilst the highest sensitivity (75·7%) was demonstrated by the UK military adult tool, the BCD Triage Sieve. The US-based JumpSTART demonstrated low sensitivity in predicting P1 status in children <8 years (35·5% in 4 to 8 year olds and 28·5% in 0 to 4 year olds) and was outperformed by its adult counterpart START (60% sensitivity in children <8 years). Lerner's criteria with paediatric-specific fluid resuscitation measures have been used to define triage categories in a paediatric population, yielding clinically meaningful differences between patient groups.

Despite utilising age-specific paediatric respiratory and heart rate thresholds, the PTT is outperformed by the adult BCD Triage Sieve. This may be attributable to the BCD Triage Sieve's early application of the mental status assessment "Responds to voice?" (approximately equivalent to a GCS of 12) [26], as mental status correlates strongly with outcomes following trauma [11,12]. A US registry-based study examining triage tool performance, in which 36,618 out of 530,695 patients were aged <16 years, demonstrated that GCS was a strong predictor of mortality at hospital discharge in patients of all ages (AUC 0·825), particularly in children aged 0 to 8 years (AUC 0·964), where GCS outperformed CareFlight and START [11]. In our study, the BCD Triage Sieve has demonstrated twice the sensitivity (75·7% vs. 35·5%) in predicting the need for time-critical major resuscitative and/or surgical intervention (Priority 1 status) in injured children when compared with the currently utilised PTT, as well as enhanced sensitivity in predicting mortality (94·3%) and ISS>15 (73·6%). We recommend that the BCD Triage Sieve replace the PTT (and the NARU Triage Sieve used in children ≥12 years) as the primary MI triage tool for patients aged <16 years in the UK. The BCD Triage Sieve has a higher rate of over-triage compared to PTT (67·4% vs. 45·5%, respectively). This is comparable to the BCD Triage Sieve's 70·9% over-triage rate demonstrated in an adult study [18], which demonstrated the BCD Triage Sieve offers optimal performance in predicting P1 status in adults [18]. Having one primary MI triage tool for use across all ages would simplify EMS training and improve the consistency of triage. Human factors affect performance during MIs [8], and in incidents involving an ongoing threat, first responders trained in triage methods that employ arithmetic often resort to more simplistic means, as noted following the San Bernadino shootings [3]. Avoiding tools that involve the application of more complex, age-specific physiological parameters (e.g. PTT) in the triage of children is likely to be associated with more reliable triage in practice. Care providers may also choose to employ the BCD Triage Sieve in casualty clearing stations and at hospital reception in the absence of an effective secondary paediatric MI triage tool.

Our study demonstrated that CareFlight and PTT had similar sensitivity in predicting P1 status (40·4% and 44·7%, respectively), consistent with a prospective South African study of 3461 children (<13 years) presenting to ED, which demonstrated that CareFlight and PTT had comparable sensitivity (46% and 41·5%, respectively) in predicting the need for urgent non-orthopaedic surgery or other resuscitative intervention [19]. This study, similar to ours, also highlighted that JumpSTART had the lowest sensitivity (0·8%) of all tools tested: JumpSTART is intended to replace START in children <8 years; however, in our study it is outperformed by START in all age subgroups. Based on this and prior evidence, regions employing JumpSTART may wish to consider alternative methods to triage children in MIs.

A key strength of this study is use of Lerner's criteria (expanded to include paediatric-specific fluid resuscitation measures) to define triage categories, which has several advantages over using ISS>15 or intervention-based criteria described previously [12,22]. ISS>15 is widely used in quality assurance as the threshold to justify the highest tier of trauma care in the UK and US: however ISS correlates poorly with the need for medical intervention, as our study confirms [23]. In 2001, Garner described criteria for defining Priority 1 status, including non-orthopaedic surgery within 6 h and other resuscitative measures [12]. Garner included in their definition of P1 patients who received over 1000 ml of fluid to maintain a blood pressure above 89 mmHg (which is less common in current practice, with preferential use of blood and blood products), and those undergoing invasive intracranial pressure monitoring, which has since been shown to lack correlation with neurological outcome. In 2006, Wallis described triage categories (P1, P2 and P3 equivalents) derived using a Delphi consensus of experts, however, these were only applicable to children and outlined aggressive time cut-offs more akin to combat casualty care (e.g. P1 casualties are those requiring laparotomy or thoracotomy within one hour) [22]. By comparison, our study has utilised Lerner's criteria (derived by expert consensus and literature review), rather than author-defined criteria alone [9]. Lerner's system has multiple advantages: it defines all possible triage categories, it considers a broad range of injury mechanisms including burns, its use has been validated in adults [18] and it is applicable to patients of all ages (allowing children and adults to be considered simultaneously), which may facilitate more equitable resource allocation [9]. Furthermore, we have demonstrated clinically meaningful differences in mortality, ISS and ICU requirement in patient groups constituting each triage category. In a previous Utstein-style consensus on the reporting of the acute medical response to disasters, experts highlighted the need to define a universally accepted measure of triage accuracy, particularly to establish whether criteria used to sort injured survivors into categories are clinically meaningful and are adequately predictive of survivability [32]. We recommend use of Lerner's definitions of triage categories (with paediatric-specific fluid resuscitation measures, where applicable) as an objective, evidence-based means by which to model novel tools and to define triage categories when conducting post-event evaluations of UK and international MI triage. Uniformity in reporting of MI triage will allow meaningful comparison between studies and thereby facilitate refinement in MI policy [4].

Other study strengths include use of trauma registry data, allowing tool performance to be assessed on a large, nationally representative sample of injured children, incorporating multiple injury mechanisms. This overcomes the practical limitations of conducting studies during actual MIs. Computed application of triage tools has allowed the inherent discriminatory capability of triage tools to be assessed independently of human error.

Study limitations include under-representation of burns and blast injury mechanisms. The low proportion of patients with gunshot wounds is representative of the UK, where mass shootings are rare following the introduction of strict gun laws after the 1996 Dunblane Massacre [4,6]. Our study findings may be less generalisable to other nations and during conflict [4]. The term "catastrophic haemorrhage" utilised by four tools could not be applied using registry information; however, there is abundant evidence that haemorrhage is the leading preventable cause of death following trauma and that control of bleeding improves survival [33]. This study focusses on the ability of triage tools to predict P1 status only. Future studies should evaluate the ability of tools to predict other triage categories (e.g. over-triage of P3 patients as P2 may impact hospital resources) and further reduce over-triage rates. Over-triage has the advantage of rapidly removing children from the scene; however, there is a direct correlation between over-triage and mortality [7]. Further work should focus on developing tools that do not involve arithmetic calculation: CareFlight and RAMP employ qualitative assessments alone, however, both have demonstrated sensitivity <50% in predicting P1 status in this and other studies. Our study findings may be biased by patients excluded due to missing pre-hospital physiological data: excluded patients had a higher mortality and younger age when compared with included patients. In particular, our study's estimation of tools' ability to predict mortality as an outcome measure is likely further biased by TARN's exclusion of pre-hospital deaths. It is unclear why such a large proportion of children (67·2%) within the trauma registry are missing pre-hospital data as compared with 9·2% of adults in a similar study [18]. Possible explanations include challenges in collecting prehospital observations in young children, expedited transfer of paediatric casualties to hospital or shortfalls in submitting data to TARN. We strongly recommend that care providers explore and address why the quality of paediatric prehospital data is remarkably different from that of adults within the same trauma registry. Several post-event evaluations have cited the availability of pre-hospital data as a barrier to determining MI triage tool performance [3,4,32]. We considered data imputation and use of first recorded hospital physiology (which may be influenced by treatments administered prior to hospital arrival), however these may further bias results. Although not without limitation, use of national trauma registry data may represent the largest UK population of injured children in whom triage tool performance can be assessed.

In conclusion, based on performance assessed using this trauma registry population, we recommend that the BCD Triage Sieve should be applied to both children and adults injured in UK MIs, which would simplify both training and application of the triage process while improving in parallel the accuracy in identifying patients in need of time-critical major resuscitative and surgical intervention. The methodology used in this study (Lerner's criteria, incorporating paediatric-specific fluid resuscitation measures) uses outcome data to identify appropriateness of original triage category. This method provides an objective standard for developing novel triage tools as well as conducting post-event evaluations of future UK MIs.

5. Author contributions

NM conducted a literature review prior to the study. NM, DB, DK and GVG designed the study. NM, SC and YX accessed the database, verified the underlying data and conducted analysis. All authors contributed to data interpretation. NM wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

Declaration of Competing Interest

The authors confirm that they have no conflicts of interest to declare.

Funding

This study is funded by the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre. GVG also acknowledges support from the MRC Heath Data Research UK (HDRUK/CFC/01). The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the Ministry of Defence.

Data sharing statement

De-identified patient data utilised for this study are proprietary to the Trauma Audit and Research Network, University of Manchester and may be requested directly from TARN.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2021.101100.

References

[1] Kilner TM, Brace SJ, Cooke MW, Stallard N, Bleetman A, Perkins GD. In 'big bang' major incidents do triage tools accurately predict clinical priority?: a systematic review of the literature. Injury 2011;42(5):460–8.
[2] Panel T.K.R. The Kerslake report: an independent review into the preparedness for, and emergency response to, the Manchester arena attack on 22nd may 2017, 2018.
[3] Pepper M, Archer F, Moloney J. Triage in complex, coordinated terrorist attacks. Prehosp Disaster Med 2019;34(4):442–8.
[4] Turner CD, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Crit Care 2016;20(1):362.
[5] Dark P, Smith M, Ziman H, Carley S, Lecky F. Healthcare system impacts of the 2017 Manchester Arena bombing: evidence from a national trauma registry patient case series and hospital performance data. Emerg Med J 2021:0:1–10 emermed-2019. doi: 10.1136/emermed-2019-208575.
[6] Hodgetts TJ, Hall J, Maconochie I, Smart C. Paediatric triage tape. Pre Hosp Immed Care 1998;2:155–9.
[7] Frykberg ER, Tepas JJ. Terrorist bombings. Lessons learned from Belfast to Beirut. Ann Surg 1988;208(5):569–76.
[8] American Academy of Pediatrics. American College of Emergency Physicians. American College of Surgeons - Committee on Trauma. American Trauma Society. Children's National Medical Center. Child Health Advocacy Institute. Emergency Medical Services for Children National Resource Center. International Association of Emergency Medical Services Chiefs. National Association of County and City Health Officials. National Association of Emergency Medical Technicians. National Association of EMS Physicians. National Association of State EMS Officials. National Disaster Life Support Education Consortium. National EMS Management Association. Society for the Advancement of Violence and Injury Research. Health Resources and Services Administration/Maternal and Child Health Bureau Emergency Medical Services for Children Program. Model uniform core criteria for mass casualty triage. Disaster Med Public Health Prep 2011;5(2):125–8.
[9] Lerner EB, McKee CH, Cady CE, et al. A consensus-based gold standard for the evaluation of mass casualty triage systems. Prehosp Emerg Care 2015;19(2):267–71 official journal of the National Association of EMS Physicians and the National Association of State EMS Directors.
[10] Romig LE. Pediatric triage. A system to JumpSTART your triage of young patients at MCIs. JEMS 2002;27(7):60–3 52-8.
[11] Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. Ann Emerg Med 2013;61(6):668–76 e7.
[12] Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. Ann Emerg Med 2001;38(5):541–8.
[13] Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. Emerg Med J 2006;23(6):475.
[14] Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehosp Disaster Med 2008;23(1):3–8.
[15] Vassallo J., Smith J.E., Wallis L.A. Major incident triage and the implementation of a new triage tool, the MPTT-24. 2018; 164(2): 103–6.
[16] National Ambulance Resilience Unit (NARU). Clinical guidance: medical support minimal requirements for a mass casualty incident. National Ambulance Resilience Unit (NARU); 2014 www.naru.org.uk.
[17] Bennett A. Methodologies utilized and lessons learned in high threat environments and mass casualty environments. J High Threat Austere Med 2019:1–7.
[18] Malik NS, Chernbumroong S, Xu Y, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: comparative analysis using the UK national trauma registry population. EClinicalMedicine 2021;36:100888.
[19] Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. Emerg Med J EMJ 2006;23(6):475–8.
[20] Ardolino A, Cheung CR, Lawrence T, et al. The accuracy of existing prehospital triage tools for injured children in England: an analysis using emergency department data. Emerg Med J 2015;32(5):397.
[21] Baxt WG, Upenieks V. The lack of full correlation between the injury severity score and the resource needs of injured patients. Ann Emerg Med 1990;19(12):1396–400.
[22] Wallis L, Carley S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. Emerg Med J EMJ 2006;23(4):291–5.
[23] The University of Manchester. The trauma audit and research network. Continuous. 2021.www.tarn.ac.uk.
[24] Cobas MA, De la Peña MA, Manning R, Candiotti K, Varon AJ. Prehospital intubations and mortality: a level 1 trauma center perspective. Anesth Analg 2009;109(2):489–93.
[25] Deakin CD, Low JL. Accuracy of the advanced trauma life support guidelines for predicting systolic blood pressure using carotid, femoral, and radial pulses: observational study. BMJ Clin Res Ed 2000;321(7262):673–4.
[26] Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow coma scale. Ann Emerg Med 2004;44(2):108–13.
[27] Royal College of Paediatrics and Child Health. UK-WHO growth charts - 0-4 years. 2013. https://www.rcpch.ac.uk/resources/uk-who-growth-charts-0-4-years (accessed 17 May 2021.
[28] Resuscitation Council UK. Paediatric advanced life support guidelines. 2021. https://www.resus.org.uk/library/2021-resuscitation-guidelines/paediatric-advanced-life-support-guidelines.
[29] D'Agostino R, Pearson ES. Tests for departure from normality. Empirical results for the distributions of $b_2$and $\sqrt{b_1}$. Biometrika 1973;60(3):613–22.
[30] Learn S. Compute area under the curve (AUC) using the trapezoidal rule. Continuous. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html (accessed Accessed July 2021.
[31] Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Process Lett 2014;21(11):1389–93.
[32] Debacker MHI, Dhondt E, Rockenschaub G, Rüter A, Codreanu T, Koenig KL, et al. Utstein-style template for uniform data reporting of acute medical response in disasters. PLoS Curr. 2012;4.
[33] Arul G, Reynolds J, DiRusso S, et al. Paediatric admissions to the British military hospital at Camp Bastion, Afghanistan. Ann R Coll Surg Engl 2012;94(1):52–7.

<u>Supplementary Data Table 1</u>: <u>Designation of Triage Categories based on EMS and Hospital Interventions received</u>:

	DEAD	**TARN equivalent**
1·	A lack of palpable pulse and/or respiratory effort (i·e· cardiac or respiratory arrest) at initial EMS assessment that is not responsive to airway positioning or needle decompression	Any cardiorespiratory resuscitation at scene AND Inpatient death
2·	Lack of pulse or respiratory effort within 15 minutes of EMS arrival at scene	Any cardiorespiratory resuscitation at scene AND Inpatient death
		↓ **All no**
	EXPECTANT	**TARN equivalent**
1·	In patients aged 0 to 49 years old, third degree (full thickness) burns to >90% of the body	90% Total Body Surface Burns
2·	In patients over 50 years old, third degree (full thickness) burns to >80% of the body	Age≥50 years AND 40-89% Total Body Surface Burns
3·	Penetrating or blunt trauma to the head which crosses the midline with agonal respirations and/or no motor response decorticate posturing or decerebrate posturing (i·e· a motor GCS of 3 or less)	Any head injury AND Total GCS=3 or Motor GCS≤3 at scene AND Inpatient death
4·	Uncontrolled haemorrhage that resulted in cardiac arrest prior to EMS transport	*No TARN equivalent*
		↓ **All no**
	PRIORITY 1 (IMMEDIATE)	**TARN equivalent**
1·	Neurologic, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	<u>At scene:</u> Pericardial decompression, pericardiocentesis, thoracotomy, heart surgery <u>In hospital (including ED):</u> Evacuation of EDH/SDH, Evacuation ICH, Elevation depressed cranium, Repair Cranium fracture, Open Craniotomy, Burrhole of Cranium, Lobectomy of brain, Repair of Dura, Craniectomy Laparotomy, Splenectomy, Nephrectomy, Resection Liver, Repair Spleen, Repair Kidney laceration, Abdominal Packing, Caesarian Delivery for trauma, Colostomy, Hemicolectomy/Colectomy, Ileectomy, Repair Liver laceration, Repair Colon laceration, Repair Rupture to Bladder, Repair mesentery of small bowel, Repair mesentery of colon, Excision of Pancreas, Repair of Duodenum, Repair of Jejunum, Repair of Ileum, Repair of Stomach, Rectal operation, Bowel operations (specified), Surgery involving the Iliac artery, Surgery involving the Subclavian artery, Aortic Repair, Pericardiocentesis, Thoracotomy, Aortic Repair, Pneumonectomy, Heart Surgery, Repair of lung, Repair Oesophagus, Diaphragm repair External Fixation of Pelvis, Fixation of Pelvic Ring, Fixation of Acetabulum
2·	Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	Fasciotomy, Surgery to the brachial or femoral artery, Amputation of upper/lower limb· Any injury to the brachial or femoral artery (not time dependent)
3·	Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	Escharotomy
4·	Chest tube placed within 2 hours of arrival at hospital	Insertion of chest tube at scene or in-hospital
5·	An advanced airway intervention (e·g· intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	Airway obstruction at scene or in-hospital, Airway support required at scene or in-hospital, Intubation and mechanical ventilation at scene or in-hospital, CPAP administration in ED or Critical Care, Cricothyroidotomy or tracheostomy required at scene or in-hospital
6·	IV vasopressors administered within 2 hours of arrival at hospital	Administration of vasopressors/inotropes
7·	Arrived in the ED with uncontrolled haemorrhage	Administration of 4 or more units of blood products within 24 hours of admission, Any use of Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA), Interventional radiology/embolisation within 4 hours of hospital arrival ***Administration of blood and/or blood products within 1 hour of hospital arrival** ***Administration of >20ml/kg of intravenous fluids within 1 hour of hospital arrival**
9·	Patient who required EMS initiation of CPR (i·e· had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	Any cardiopulmonary resuscitation in hospital (in the absence of cardiopulmonary resuscitation at scene)
		↓ **All no**
	PRIORITY 3 (MINIMAL)	**TARN equivalent**
1·	Discharged from the ED with no X-rays or an extremity X-ray that was negative or showed an uncomplicated fracture (i·e· a closed extremity fracture without significant displacement or neurovascular compromise); no laboratory testing; received only simple wound repair (single layer suturing only); and received no medications intravenously (does not include fluids), or inhaled (does not include oxygen) from EMS or in the hospital	*Not included in TARN database (these patients would not meet TARN inclusion criteria)*
		↓ **Any no**
	PRIORITY 2 (DELAYED)	**TARN equivalent**
1·	All remaining patients	All remaining TARN patients who do not fulfil any criteria for previous triage categories

Ledger: *These parameters are specific to children (aged <16years). Cardiac arrest is defined as a lack of palpable pulse and EMS initiation of CPR. CPR= cardiopulmonary resuscitation. Chemical, biological, radiological and neurological (CBRN) entities were excluded from triage category definitions as no CBRN patients exist within the TARN database. There is international variation in the names of triage categories with approximate equivalence as follows: Priority 1 (Immediate, Red, T1); Priority 2 (Delayed, T2, Urgent); Priority 3 (Minimal, Minor, T3); Expectant (Unsalvageable, T1 hold, P4); Dead (Black).

Supplementary Table 2: Tool prediction of in-hospital mortality (patients under 16 years)

Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
BCD Triage Sieve	94·3 (83·4, 98·5)	37·5 (36·1, 38·9)	5·7	98·4	0·659 (0·594, 0·724)
CareFlight	83·0 (69·7, 91·5)	86·0 (85·0, 87·0)	17·0	94·0	0·845 (0·809, 0·881)
JumpSTART	86·8 (74·0, 94·1)	86·4 (85·4, 87·4)	13·2	93·5	0·866 (0·834, 0·898)
MIMMS Triage Sieve	88·7 (76·3, 95·3)	81·7 (80·6, 82·8)	11·3	95·0	0·852 (0·817, 0·887)
MPTT	18·9 (9·9, 32·4)	36·0 (34·6, 37·3)	81·1	99·7	0·274 (0·196, 0·352)
MPTT-24	17·0 (8·5, 30·3)	39·8 (38·5, 41·2)	83·0	99·7	0·284 (0·205, 0·363)
MSTART	92·5 (80·9, 97·6)	78·2 (77·0, 79·3)	7·5	95·6	0·853 (0·819, 0·888)
NARU Triage Sieve	84·9 (71·9, 92·8)	72·1 (70·8, 73·3)	15·1	96·8	0·785 (0·738, 0·832)
PTT*	69·8 (55·5, 81·3)	86·2 (85·2, 87·1)	30·2	94·8	0·780 (0·732, 0·828)
RAMP	90·6 (78·6, 96·5)	79·1 (78·0, 80·3)	9·4	95·5	0·849 (0·813, 0·884)
START	91·7 (71·5, 98·5)	79·7 (78·1, 81·3)	8·3	95·8	0·801 (0·807, 0·907)

Ledger: BCD Triage Sieve =Battlefield Casualty Drills Triage Sieve, MIMMS Triage Sieve=Major Incident Medical Management System Triage Sieve, MPTT=Modified Physiological Triage Tool, MSTART=Modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve, RAMP=Rapid Assessment of Mentation and Pulse, START=Simple Triage and Rapid Treatment, PTT=Paediatric Triage Tape. *The PTT is only applicable to those under 12 years (n=2516, 50·7%).

Supplementary Table 3: Tool prediction of ISS>15

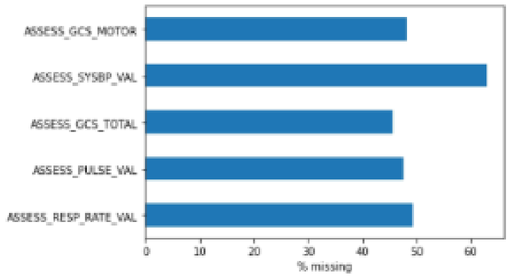
Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
BCD Triage Sieve	73·6 (71·4, 75·7)	42·4 (40·7, 44·1)	26·4	61·8	0·580 (0·563, 0·596)
CareFlight	35·3 (33·0, 37·7)	95·2 (94·4, 95·9)	64·7	21·9	0·653 (0·637, 0·668)
JumpSTART	29·9 (27·6, 32·2)	93·1 (92·2, 94·0)	70·1	32·3	0·615 (0·599, 0·631)
MIMMS Triage Sieve	32·8 (30·5, 35·1)	87·6 (86·4, 88·7)	67·2	43·9	0·602 (0·586, 0·618)
MPTT	61·7 (59·3, 64·1)	35·6 (33·9, 37·2)	38·3	68·4	0·486 (0·469, 0·504)
MPTT-24	59·4 (57·0, 61·8)	40·2 (38·5, 41·8)	40·6	67·6	0·498 (0·481, 0·515)
MSTART	43·5 (41·1, 46·0)	87·6 (86·4, 88·7)	56·5	37·1	0·656 (0·640, 0·671)
NARU Triage Sieve	43·1 (40·7, 45·6)	78·5 (77·0, 79·9)	56·9	50·8	0·608 (0·592, 0·624)
PTT*	34·5 (32·2, 36·9)	95·3 (94·5, 96·0)	65·5	22·1	0·649 (0·633, 0·665)
RAMP	42·4 (40·0, 44·8)	88·4 (87·3, 89·5)	57·6	36·1	0·654 (0·638, 0·670)
START	36·4 (33·0, 40·0)	85·6 (83·8, 87·2)	63·6	48·2	0·610 (0·587, 0·633)

Ledger: BCD Triage Sieve =Battlefield Casualty Drills Triage Sieve, MIMMS Triage Sieve=Major Incident Medical Management System Triage Sieve, MPTT=Modified Physiological Triage Tool, MSTART=Modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve, RAMP=Rapid Assessment of Mentation and Pulse, START=Simple Triage and Rapid Treatment, PTT=Paediatric Triage Tape. *The PTT is only applicable to those under 12 years (n=2516, 50·7%).

Supplementary data: Analysis of missing data (relationship with patient characteristics/outcome and age)

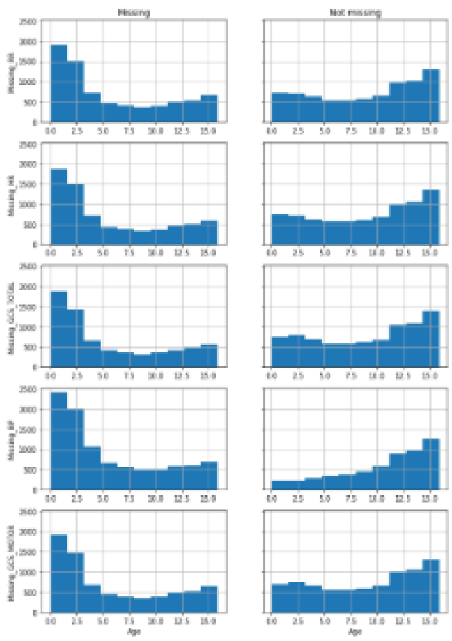
Number of patients with missing data:

Respiratory rate (ASSESS_RESP_RATE_VAL)	7477
Heart rate (ASSESS_PULSE_VAL)	7194
Total GCS Score (ASSESS_GCS_TOTAL)	6887
Systolic Blood Pressure (ASSESS_SYSBP_VAL)	9534
GCS Motor Score (ASSESS_GCS_MOTOR)	7275



We examined whether missing physiology has any association with patients' characteristics i.e. gender, injury type, injury mechanism, outcome, and age.

- There is an association between missing systolic blood pressure and patient gender (p = 0.0129). Females were more likely to have missing BP.
- There are associations between missing physiology and injury type (P < 0.0001). Patients with blunt injury were more likely to have missing HR, GCS TOTAL, GCS MOTOR, and BP. Patients with penetrating injury were more likely to have missing RR.
- There is an association between missing physiological data and outcome (p < 0.01). Patients with missing RR, HR, GCS TOTAL, GCS MOTOR were less likely to die. Patients with missing BP were more likely to die.
- There is an association between missing physiological data and INJURY MECHANISM (p < 0.01). Patients with missing RR, HR, GCS TOTAL and MOTOR were more likely to have the following injury: blow, fall less than 2M, and other. Patients with missing BP were also more likely to have burn injury.
- Patients with missing physiological data tend to be younger (median age 3.6 *vs.* 9.9 years old). The following graph demonstrates the relationship between age and missing physiological data.



Ledger: GCS=Glasgow Coma Score, BP=systolic blood pressure, HR=heart rate, RR=respiratory rate

Chapter 2 Conclusions

Evidence-based triage category definitions have been validated amongst adults and children from the UK national trauma registry. This system developed by Lerner et al defined P1 patients as those requiring any one of eight time-critical life-saving surgical and major resuscitative interventions, directly reflecting patient resource requirements, and is shown to correlate well with casualty distribution within UK regionalised systems of trauma care. Hence, it is recommended that Lerner's criteria should become the gold standard in reporting triage tool performance in future UK major incidents.

In an effort to examine the performance of existing tools amongst adults and children, these intervention-based triage category definitions were compared with the triage category assignments of ten existing international triage tools. A key limitation of this approach is the use of singly-injured TARN patients as surrogates for those injured in MIs, in whom injury characteristics and severity as well as outcome may differ. However, use of this large and highly complete dataset of a nationally representative patient population has yielded valuable evidence which may inform national policy development. This comparative analysis demonstrated that the current UK national triage tool, the NARU Triage Sieve, is poorly sensitive in identifying P1 patients who require time-critical, life-saving interventions, and was surpassed in sensitivity by several tools, the best of which was the UK military's BCD Triage Sieve (Chapter 2, publication number 1). Similarly, in children from the national trauma registry, the current UK tools, the Paediatric Triage Sieve and JumpSTART, demonstrated suboptimal performance and were substantially outperformed by several tools, with the BCD Triage Sieve again demonstrating the greatest sensitivity (Chapter 2, publication number 2).

CHAPTER 3: MACHINE LEARNING DEVELOPMENT AND VALIDATION OF NOVEL PRIMARY AND SECONDARY TRIAGE TOOLS

Introduction to Chapter 3

In Chapter 2, an objective and evidence-based system to define triage categories has been validated amongst adults and children from the UK national trauma registry (TARN). This system developed by Lerner et al defined P1 patients as those requiring any one of eight time-critical life-saving surgical and major resuscitative interventions, directly reflecting patient resource requirements in the MI setting. Existing international adult and paediatric MI triage tools have been evaluating using this proposed gold standard. This comparative analysis identified the BCD Triage Sieve as the tool with the greatest sensitivity in predicting P1 status amongst both adults and children, surpassing the current UK adult (NARU Triage Sieve) and Paediatric (Paediatric Triage Tape) MI triage tools. The BCD Triage Sieve can thus act as a useful comparator for the modelling of novel tools, however it is noted that this expert-derived tool has a considerably high over-triage rate of 68-72%, suggesting that there is room for improvement, where the sensitivity of BCD Triage Sieve is preserved or increased but with a more acceptable over-triage rate. Additionally, many of the existing tools outlined in Chapter 2, including the BCD Triage Sieve, utilise arithmetic calculations as part of patient assessment. This is in spite of several post-event evaluations of real-life major incidents reporting that counting is difficult to perform accurately in practice, and leads to the abandonment of the use of tools in situations where there is an ongoing threat to care providers.

Machine learning (ML) explores the construction and study of algorithms that can “learn” or make predictions on data, without following strictly static computer programming instructions. ML has several advantages over traditional statistics, for example by using algorithms to identify previously unrecognised input variables that predict outcome, unlike the requirement for human-determined variables in traditional statistics. Additionally, non-linear relationships between variables and outcomes can be harnessed, making ML particularly useful in identifying complex or unclear data inter-relationships. In healthcare, ML has the potential to impact and progress patient triage, diagnosis of disease, prediction of patient outcomes and estimation of prognosis. Tree-based machine learning models have demonstrated utility in clinical risk stratification and the development of decision support tools, with the ability to capture non-linear interactions between input variables. Although there is a paucity of databases encompassing data from major incident victims, large trauma registry databases exist, encompassing a broad range of patient demographics and injury mechanisms. There is potential for machine learning applied to these large databases to result in models which outperform existing tools in predicting the need for life-saving interventions in injured patients, particularly as previous triage tools are largely based on expert opinion. Hence, there is scope to develop machine-learning models that can be adapted into primary and secondary MI triage tools. Crucially, any such models must be externally validated using an independent population of injured patients prior to recommending these for practical use in MIs. The aim of Chapter 3 is to use machine learning to develop novel primary and secondary triage tools for the UK civilian setting and to validate these externally using a geographically distinct population of injured patients.

Publication number 3: Triage in major incidents: development and external validation of novel machine learning-derived primary and secondary triage tools

Author contributions:

NM and YX contributed equally to this study. **NM** and GVG designed the study. YX, **NM** and SC accessed the TARN database, verified the underlying data and conducted analysis. JV and YX accessed JTTR data, JV verified the underlying data and YX conducted analysis. All authors contributed to data interpretation. **NM** (clinician) and YX (machine learning expertise) wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

OPEN ACCESS

Triage in major incidents: development and external validation of novel machine learning-derived primary and secondary triage tools

Yuanwei Xu,[1,2] Nabeela Malik ,[3,4,5,6] Saisakul Chernbumroong,[1,3] James Vassallo,[7,8] Damian Keene,[4,6] Mark Foster,[3,4,6] Janet Lord,[3,5] Antonio Belli,[3,4] Timothy Hodgetts,[9] Douglas Bowley,[4,6] George Gkoutos[1,2,3,10,11]

Handling editor Shammi L Ramlakhan

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2022-212440).

For numbered affiliations see end of article.

Correspondence to
Miss Nabeela Malik, NIHR Surgical Reconstruction Microbiology Research Centre, Birmingham B15 2TH, UK; nabeelamalik@nhs.net

YX and NM are joint first authors.

Received 7 March 2022
Accepted 12 August 2023

ABSTRACT

Background Major incidents (MIs) are an important cause of death and disability. Triage tools are crucial to identifying priority 1 (P1) patients—those needing time-critical, life-saving interventions. Existing expert opinion-derived tools have limited evidence supporting their use. This study employs machine learning (ML) to develop and validate models for novel primary and secondary triage tools.

Methods Adults (16+ years) from the UK Trauma Audit and Research Network (TARN) registry (January 2008–December 2017) served as surrogates for MI victims, with P1 patients identified using predefined criteria. The TARN database was split chronologically into model training and testing (70:30) datasets. Input variables included physiological parameters, age, mechanism and anatomical location of injury. Random forest, extreme gradient boosted tree, logistic regression and decision tree models were trained to predict P1 status, and compared with existing tools (Battlefield Casualty Drills (BCD) Triage Sieve, CareFlight, Modified Physiological Triage Tool, MPTT-24, MSTART, National Ambulance Resilience Unit Triage Sieve and RAMP). Primary and secondary candidate models were selected; the latter was externally validated on patients from the UK military's Joint Theatre Trauma Registry (JTTR).

Results Models were internally tested in 57 979 TARN patients. The best existing tool was the BCD Triage Sieve (sensitivity 68.2%, area under the receiver operating curve (AUC) 0.688). Inability to breathe spontaneously, presence of chest injury and mental status were most predictive of P1 status. A decision tree model including these three variables exhibited the best test characteristics (sensitivity 73.0%, AUC 0.782), forming the candidate primary tool. The proposed secondary tool (sensitivity 77.9%, AUC 0.817), applicable via a portable device, includes a fourth variable (injury mechanism). This performed favourably on external validation (sensitivity of 97.6%, AUC 0.778) in 5956 JTTR patients.

Conclusion Novel triage tools developed using ML outperform existing tools in a nationally representative trauma population. The proposed primary tool requires external validation prior to consideration for practical use. The secondary tool demonstrates good external validity and may be used to support decision-making by healthcare workers responding to MIs.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ During major incidents (MIs) (eg, terrorist attacks), triage tools have a crucial role in maximising overall survival by identifying priority 1 (P1) patients.
⇒ Existing tools, derived using expert opinion, have limited evidence to support their use.

WHAT THIS STUDY ADDS

⇒ In this study, novel machine learning-based primary and secondary triage tools surpassed the current UK National Ambulance Resilience Unit Triage Sieve and other existing tools in identifying P1 patients within a nationally representative trauma population.
⇒ The secondary tool demonstrated favourable external validity. However, the primary tool could not be externally validated due to missing GCS component data.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The proposed secondary tool, applicable using a portable device, may be used to support decision-making among healthcare workers responding to MIs.

INTRODUCTION

In the immediate aftermath of a major incident (MI), patient needs exceed the resources available to treat them[1–5]: triage tools seek to categorise patients, to guide the order of treatment, transport from the scene and the choice of medical facility for definitive care.[5 6] A vital function of triage tools is to identify patients requiring time-critical, life-saving interventions (priority 1 or P1 patients). Failure to identify these patients (undertriage) is associated with absolute harm arising from delays in care or selection of an inappropriate medical facility.[6 7] However, overtriage may risk overwhelming healthcare facilities with patients not requiring time-critical treatment.[2]

Primary triage, conducted at the scene of an MI, uses paper-based flow diagrams that are quick and simple to apply under challenging conditions.[8] Existing primary triage tools have largely been developed using expert opinion, often with limited evidence to support their use.[6] These include the National Ambulance Resilience Unit (NARU) Triage Sieve (current UK tool for adults), the Australian CareFlight and the US Simple Triage And Rapid Treatment (START).[6 9 10] These tools

To cite: Xu Y, Malik N, Chernbumroong S, *et al*. *Emerg Med J* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/emermed-2022-212440

use ambulatory status to designate priority 3 (minor) category, followed by physiological assessments to distinguish P1 from P2 (less critical) patients. A recent study demonstrated that the UK military's Battlefield Casualty Drills (BCD) Triage Sieve attained greatest sensitivity among 10 international primary triage tools in detecting P1 status among adults; however, this was associated with an overtriage rate of 72%.[10]

Primary triage is often, but not always, followed by a further targeted prehospital clinical assessment of patients known as secondary triage. This is usually undertaken in a place of relative safety (eg, Casualty Clearing Station or hospital reception area)[1 8]; thus, the additional use of medical equipment and/or portable devices is more plausible. Two existing secondary MI triage tools are the UK's Major Incident Medical Management and Support Triage Sort which has suboptimal sensitivity (15.7%) in predicting the need for life-saving intervention,[11] and the US points-based Sacco Triage Method (developed to predict mortality) which is time-consuming and complex to apply.[9]

Anatomical assessment of injuries has yet to feature in any MI triage tool, yet this is commonly used in the field triage of singly injured patients.[12] Advanced age is associated with worse outcomes following injury; however, existing tools do not incorporate this in patient assessment.[13] There is scope to develop evidence-based primary and secondary MI triage tools which offer greater sensitivity while decreasing overtriage compared with the BCD Triage Sieve, yet preserve applicability. Tree-based machine learning models have demonstrated utility in clinical risk stratification, with the ability to capture non-linear interactions between input variables.[14 15] This study aimed to develop machine learning models that can be adapted into primary and secondary MI triage tools and to externally validate these models using an independent population of injured patients.

METHODS

Database for model training and internal testing

Model development and validation were conducted according to Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis guidelines.[16] Adult (16+ years) patients from the Trauma Audit and Research Network (TARN) registry presenting between 1 January 2008 and 31 December 2017 were included.[17] The TARN registry prospectively captures prehospital and hospital data from 169 hospitals in England and Wales for patients who meet the following inclusion criteria: length of stay >72 hours or admission to intensive care and/or death in hospital.[17] TARN does not include prehospital deaths. Patients for whom any input variables required for modelling were missing were excluded. Using hospital arrival dates recorded by TARN, the database was split temporally (70:30) to generate model training and internal testing datasets, respectively.

Primary outcome of interest

The primary outcome of interest was P1 status, defined as the need for time-critical life or limb-saving surgery and/or advanced resuscitative measures.[18] Each patient was retrospectively designated a triage category (priority 1, priority 2, priority 4/expectant or dead) (see flow diagram in online supplemental figure 1) using validated, consensus-derived definitions (table 1).[10 18] Prior to the modelling phase, patients were designated either P1 or non-P1. The small numbers of P4 and dead patients (who share physiological similarities to P1 patients) were excluded from the modelling as these may impede model performance.

Input variables selected for modelling

Input variables differ in their complexity and time taken for measurement. Variables that can be readily assessed by first responders in the MI setting were included in the modelling process (summarised in online supplemental table 1). This included all physiological parameters used by existing MI triage tools (first-recorded prehospital HR, RR and systolic BP) with the exception of capillary refill time, which has been found to be a poor reflection of circulatory status and is difficult to measure reliably in challenging settings and in non-white patients.[6 10 19] In addition to the ability to follow commands (GCS Motor) used by the CareFlight triage tool, all subcomponents of the GCS were included.[6] However, total GCS score, although known to be an important predictor of outcomes in injured patients,

Table 1 Triage category definitions

Dead	▶ Cardiac and/or respiratory arrest at initial prehospital evaluation that is not responsive to needle decompression or airway positioning (or the delivery of two rescue breaths in children less than 12 years old) ▶ Lack of palpable pulse and need for CPR (ie, cardiac arrest) within the first 15 min of EMS arrival on scene
Priority 4 (expectant)	▶ In patients aged 0–49 years: third-degree (full thickness) burns to >90% of the body ▶ In patients aged 50 years and over: third-degree (full thickness) burns to >80% of the body ▶ Penetrating trauma to the head that crosses the midline with agonal respirations and/or no motor response, decorticate posturing or decerebrate posturing (ie, GCS Motor ≤3) ▶ Blunt trauma to the head with agonal respirations and/or no motor response, decorticate posturing or decerebrate posturing (ie, GCS Motor ≤3) ▶ Uncontrolled haemorrhage that resulted in cardiac arrest (defined as a lack of palpable pulse and EMS initiation of CPR) prior to EMS transport
Priority 1	▶ Neurological, vascular or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital ▶ Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery ▶ Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital ▶ Chest tube placed within 2 hours of arrival at hospital ▶ An advanced airway intervention (eg, intubation, LMA, surgical airway) performed in the prehospital setting or within 4 hours of arrival at hospital ▶ IV vasopressors administered within 2 hours of arrival at hospital ▶ Arrived in the ED with uncontrolled haemorrhage ▶ Patient who required EMS initiation of CPR (ie, had a cardiac arrest) during transport, in the ED or within 4 hours of arrival at a hospital
Priority 2	▶ All patients who do not meet the criteria for the other categories are considered priority 2
Priority 3	▶ Discharged from ED with no X-rays or an extremity X-ray that was negative or showed an uncomplicated fracture (ie, a closed extremity fracture without significant displacement or neurovascular compromise); no laboratory testing; received only simple wound repair (single-layer suturing only); and received no medications intravenously (does not include fluids), or inhaled (does not include oxygen) from EMS or in the hospital

These definitions were derived by expert consensus and have been validated in a UK trauma population. Priority 4 (expectant) denotes injuries which are incompatible with life.
CPR, cardiopulmonary resuscitation; EMS, emergency medical services; IV, intravenous; LMA, laryngeal mask airway.

was not included.[7,12] Total GCS is time-consuming to calculate, with evidence suggesting that scores by paramedics frequently differ from those assigned by emergency physicians; hence, measurement under MI conditions may lack accuracy.[5,19,20] The ability to breathe spontaneously is an important determinant of outcome and is assessed early within several existing triage tools.[6,10] TARN does not explicitly record whether patients are spontaneously breathing at the scene of injury, nor does it record the indication for airway interventions.[17] We assumed that all patients who received an advanced airway intervention at the scene (defined as intubation and ventilation and/or surgical airway and/or the need for airway support) were unable to breathe spontaneously.[10,21]

The presence of injury in anatomical regions including the head, face, chest and limb(s) was included as input variables for modelling using retrospectively calculated Abbreviated Injury Severity (AIS) scores (TARN records AIS based on hospital rather than prehospital data). A binary input (AIS=0, AIS >0) was used rather than a graded assessment of severity. Due to the known difficulties in identifying intra-abdominal injuries based on clinical assessment alone, and the requirement to undertake detailed clinical assessment in order to reliably identify spinal injuries, the presence of spinal and abdominal injuries was not included as input variables.[22,23] Patient age was dichotomised into age ≥65 years (yes or no), which may be reliably identified by first responders.[12] Broad injury mechanism (blunt or penetrating) was included.

Input variables described thus far were deemed appropriate for inclusion in both primary and secondary triage tools. Although not conducive to primary triage due to the need for calculation, shock index (HR/systolic BP), which may correlate better with outcome than HR or systolic BP alone, was included in the modelling process as a potential component of a secondary triage tool.[24]

Model training and internal testing

Four machine learning methods were applied to the model training dataset to distinguish P1 from non-P1 patients. Decision tree (RPART) methodology was included because models can be visualised as bifurcating trees, closely resembling the format of existing primary triage tools. Two other tree-based models with demonstrated value in clinical risk stratification, random forest (RF) and eXtreme Gradient Boosting (XGB), were trained.[25,26] Further methodological details are presented as online supplemental material. Finally, we included an L1-regularised logistic regression model. We anticipated that non-P1 patients would substantially outnumber P1 patients; hence, we adopted an undersampling strategy to balance the data by leaving out random samples of non-P1 patients.[14] For each of these models, fivefold cross-validation was applied.[26]

To generate models that were no more complex to apply than existing primary triage tools, modelling included all possible combinations of 3–7 of the available 13 input variables. Model building and selection strategy are summarised in online supplemental figure 2. Models trained using all 13 input variables, although too complex for practical application as triage tools, were also considered as comparators (online supplemental table 2). Additionally, we compared the triage assignments (namely, P1 status) of 10 existing international primary triage tools to the testing dataset (online supplemental table 3).[10]

Previous studies demonstrate that elders (aged 65+ years) are over-represented in the TARN population while constituting 18.3% of the UK population[10]; hence, during testing, we split the TARN testing set by age (ages 16–64 years and 65+ years) to further evaluate model performance.

Determining feature importance

We assessed the relative importance of individual features (input variables) in model predictions using the TreeSHAP method, a model-agnostic, individualised feature attribution method for explaining predictions.[27] The resulting Shapley value for a particular feature measures the expected change in model prediction when that feature is present relative to the average model prediction. Additionally, feature importance was estimated by the contribution of each feature to the overall XGB model-predictive performance.[27]

Selection of models as candidates for primary and secondary triage tools

We sought to identify models that achieved the best possible performance (maximal sensitivity in identifying P1 patients, but also favourable overtriage rate and area under the receiver operating curve (AUC)) across all ages as well as age subgroups, using the minimal number of input variables, to maintain practical applicability. We predetermined that selected models must outperform the best performing existing triage tool, as identified by our study.

In keeping with existing practice, the primary tool candidate was intended to be a paper-based, simple algorithm. The model selected as a secondary tool was adapted into a web-based prototype using the R shiny application.

External validation of models using the Joint Theatre Trauma Registry database

The UK military's Joint Theatre Trauma Registry (JTTR) (February 2002–December 2016) was used to externally validate the selected models. JTTR includes consecutive patients who triggered trauma team activation at a deployed medical treatment facility, largely comprising combat casualties during military operations in Iraq and Afghanistan.

Children (<16 years), patients with erroneous data (eg, age over 110 years) and those with injuries recorded as both blunt and penetrating were excluded from the validation (see online supplemental figure 1). As we expected a paucity of prehospital data in this population,[28] patients' first recorded hospital physiology was used. Patients with missing data for the input variables were not excluded. Subcomponents of GCS are not routinely recorded within JTTR; these were derived for patients with GCS 15 and unavailable for those with GCS <15. Furthermore, we evaluated candidate models on a subset of JTTR patients with sufficient data to apply the best performing existing tool (subsequently found to be the BCD Triage Sieve), thereby facilitating direct comparison. Triage category definitions were applied as described earlier (table 1): since JTTR does not record the time of interventions, those performed at deployed medical treatment facilities were presumed to have occurred within 4 hours.[28]

Statistical analyses

Patient characteristics across the model training, internal testing and external validation datasets were compared using the X^2 test (Injury Severity Score (ISS) and age compared using Mann-Whitney U test); $p<0.05$ was considered statistically significant. Model performance is reported as sensitivity, specificity, undertriage (1-sensitivity) and overtriage (1-positive predictive value). The 95% CIs for the AUC were calculated using deLong's method (pROC R package, V.1.17.0.1).[29] The 95% CIs for

Table 2 Patient and injury characteristics for the model training, testing and external validation cohorts

	Model training dataset (70% TARN: 1 Jan 2008–14 Jul 2016)	Model testing dataset (30% TARN: 15 Jul 2016–31 Dec 2017)	External validation dataset (JTTR: 1 Feb 2002–31 Dec 2016)
Gender			
Male	72 817 (53.8%)	29 532 (50.9%)	5830 (97.9%)
Female	62 465 (46.2%)	28 447 (49.1%)	106 (1.8%)
Missing data	0 (0.0%)	0 (0.0%)	20 (0.3%)
Injury Severity Score			
Median (IQR)	9 (9–16)	9 (9–17)	8 (2–17)
Missing data	0 (0.0%)	0 (0.0%)	13 (0.2%)
Age			
Median (IQR)	64.3 (45.6–82.3)	70.9 (51.6–84.5)	24 (21–28)
16–64 years	69 237 (51.2%)	24 769 (42.7%)	5256 (88.2%)
65+ years	66 045 (48.8%)	33 210 (57.3%)	25 (0.4%)
Missing data	0 (0.0%)	0 (0.0%)	675 (11.3%)
Discharge status			
Alive	127 624 (94.3%)	54 383 (93.8%)	5681 (95.4%)
Dead	7657 (5.7%)	3596 (6.2%)	275 (4.6%)
Missing data	1 (0.0%)	0 (0.0%)	0 (0.0%)
Injury mode			
Blunt	131 208 (97.0%)	56 473 (97.4%)	1092 (18.3%)
Penetrating	4074 (3.0%)	1506 (2.6%)	4864 (81.7%)
Missing data	0 (0.0%)	0 (0.0%)	0 (0.0%)
Injury mechanism			
Fall less than 2 m	76 169 (56.3%)	36 380 (62.7%)	78 (1.3%)
Vehicle incident	30 195 (22.3%)	10 744 (18.5%)	389 (6.5%)
Fall more than 2 m	17 838 (13.2%)	6725 (11.6%)	37 (0.6%)
Blow(s)	4871 (3.6%)	1868 (3.2%)	0 (0.0%)
Stabbing	2871 (2.1%)	1192 (2.1%)	29 (0.5%)
Crush	1065 (0.8%)	268 (0.5%)	76 (1.3%)
Shooting	328 (0.2%)	91 (0.2%)	2316 (38.9%)
Burn	91 (0.07%)	27 (0.05%)	3 (0.1%)
Blast	88 (0.07%)	50 (0.09%)	2926 (49.1%)
Other	1766 (1.3%)	634 (1.1%)	86 (1.4%)
Missing data	0 (0.0%)	0 (0.0%)	16 (0.3%)

JTTR, Joint Theatre Trauma Registry; TARN, Trauma Audit and Research Network.

models' sensitivity at given specificity points were calculated using 500-stratified bootstrap replicates.[29]

Patient and public involvement
Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS
Training dataset and primary outcome of interest
A total of 200 728 patients were captured by TARN over the 10-year period. After exclusions, the sample consisted of 193 261 patients, of which 21 878 patients (11.3%) fulfilled P1 criteria.

The model training dataset comprised 135 282 patients, with a median age of 64.3 years, in-hospital mortality of 5.7% and predominantly blunt injuries (97%), most commonly low falls (56.3%) (table 2). Patients within the internal test dataset (n=57979) were older (median age 70.9 years vs 64.3 years, respectively, p<0.001) and more often injured by a low-level fall (62.7% vs 56.3%, p<0.001) compared with patients within the model training dataset.

Model training and internal testing
In the test set, the BCD Triage Sieve demonstrated the greatest sensitivity at 68% with overtriage at 80.8% (table 3). Existing tools performed less well in the elders' subgroup compared with younger (16–64 years) adults, with sensitivity 5.8–14.6% lower and overtriage rates 11.5–33.2% higher among elders (online supplemental table 3).

Four hundred fifty-six models were developed, which, when applied to the internal test dataset, demonstrated greater sensitivity and AUC than all existing tools. Model selection was initially narrowed down to five decision tree models as candidates for primary triage tools and 29 XGB models as candidates for secondary triage tools (see online supplemental figure 2). A comprehensive list, including performance by age subgroups within the internal (TARN) testing and external validation (JTTR) datasets (described later), is detailed in online supplemental table 4A–C. Receiver operating curves demonstrating the performance of the novel primary and secondary tool candidate models when applied to the internal testing dataset are shown in figure 1.

Feature importance
The top 10 features (figure 2A), and their relative contribution in predicting P1 status (figure 2B) are presented. By far, the most important variable was breathing status (mean Shapley value 1.2), followed by presence of a chest injury and GCS Verbal score. Age >65 years was negatively predictive of P1 status. Any abnormal GCS Verbal or GCS Motor score contributed substantially in predicting P1 status (see figure 2B). The XGB method of

Table 3 Performance characteristics of existing triage tools and novel machine learning models among adult patients (16+ years) in the testing (TARN) dataset

	Sensitivity	Specificity	Undertriage	Overtriage	AUC
Existing tools					
BCD Triage Sieve	68.2 (66.9, 69.4)	69.5 (69.1, 69.9)	31.8 (30.6, 33.1)	80.8 (80.2, 81.3)	0.688 (0.682, 0.695)
CareFlight	39.9 (38.6, 41.2)	94.5 (94.3, 94.7)	60.1 (58.8, 61.4)	56.4 (55.0, 57.8)	0.672 (0.666, 0.679)
MPTT-24	48.4 (47.1, 49.7)	66.4 (66.0, 66.8)	51.6 (50.3, 52.9)	86.7 (86.2, 87.2)	0.574 (0.567, 0.581)
MSTART	54.9 (53.6, 56.2)	88.4 (88.1, 88.7)	45.1 (43.8, 46.4)	66.5 (65.5, 67.5)	0.717 (0.710, 0.723)
NARU Triage Sieve	43.0 (41.7, 44.3)	88.3 (88.1, 88.6)	57.0 (55.7, 58.3)	71.8 (70.9, 72.8)	0.657 (0.650, 0.663)
RAMP	37.1 (35.9, 38.4)	94.6 (94.5, 94.8)	62.9 (61.6, 64.1)	57.5 (56.1, 58.9)	0.659 (0.653, 0.665)
Models selected as candidates for novel primary and secondary triage tools					
Primary triage tool candidate (decision tree)	73.0 (71.8, 74.2)	73.9 (73.5, 74.3)	27.0 (25.8, 28.2)	77.0 (76.4, 77.7)	0.782 (0.775, 0.789)
Secondary triage tool candidate (XGB)	77.9 (76.8, 79.0)	73.1 (72.7, 73.5)	22.1 (21.0, 23.2)	76.4 (75.8, 77.0)	0.817 (0.810, 0.824)

Values shown are percentages (except for AUC), accompanied by 95% CIs.
*The best performing model using each method is shown. Both machine learning models and the triage tools were evaluated using the same TARN population (internal testing dataset).
AUC, area under the receiver operating curve; BCD, Battlefield Casualty Drills (UK Military); MPTT-24, Modified Physiological Triage Tool 24 (2017); NARU, National Ambulance Resilience Unit (current UK civilian triage tool); TARN, Trauma Audit and Research Network; XGB, eXtreme Gradient Boosting.

determining feature importance yielded similar rankings (online supplemental figure 3).

Primary and secondary triage tool candidate models

The decision tree model selected for clinical adaptation into a primary triage tool (figure 3) used three qualitative binary (yes/no) assessments (breathing status at scene, ability to obey commands, that is, GCS Motor score=6, and presence of a chest injury) to categorise patients as P1 or non-P1. This achieved 73.0% sensitivity, overtriage rate of 77.0% and AUC of 0.782 when applied to the internal testing dataset (see table 3).

The XGB model selected as a secondary triage tool (figure 4) combines four input variables: GCS Motor score, breathing status at scene, presence of chest injury and classification of injury as blunt or penetrating. This model achieved 77.9% sensitivity, overtriage of 76.4% and AUC of 0.817 when applied to the internal testing dataset (figure 1 and table 3). This has been adapted into an online interactive tool (accessible via link: https://ywxtriageapp.shinyapps.io/mltriage/).

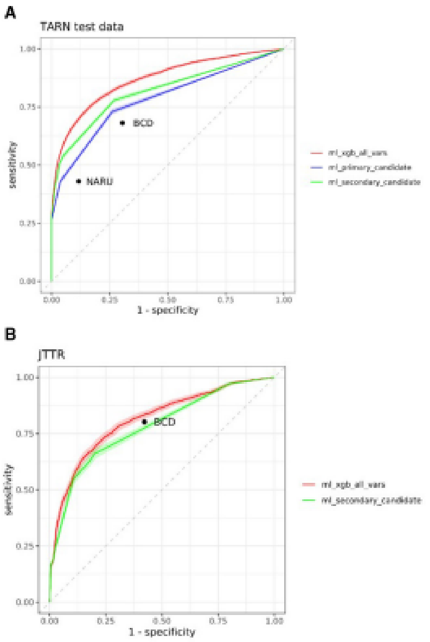


Figure 1 Performance of tool candidate models in the internal and external validation datasets compared with the Battlefield Casualty Drills (BCD) Triage Sieve (best performing existing tool) and the current UK tool, the National Ambulance Resilience Unit (NARU) Triage Sieve. Additionally, the performance of an XGB model using all 13 input variables is shown for comparison (see online supplemental material for more details). JTTR, Joint Theatre Trauma Registry; TARN, Trauma Audit and Research Network; XGB, eXtreme Gradient Boosting.

External validation of the secondary triage model (JTTR)

A total of 5956 JTTR patients met inclusion criteria (online supplemental figure 1). Median age was 24 years (IQR 21–28) and most were male (97.9%). Compared with patients in the TARN model training set, JTTR patients had lower mortality (4.6% vs 5.7%, p<0.001) and lower injury severity (median ISS 8 (IQR 2–17) vs median ISS 9 (IQR 9–16), p=0<0.001). A greater proportion of JTTR patients suffered penetrating trauma (81.7% vs 3.0%, p=0<0.001), with high prevalence of blast injury (49.1% vs 0.07%, p=0<0.001) and shooting (38.9% vs 0.2%, p=0<0.001) (see table 2). A total of 2046 (34.3%) JTTR patients had missing GCS Motor scores.

Given the high proportion of JTTR patients missing GCS Motor scores, as well as inability for decision trees to perform predictions when data are missing (unlike XGB and RF), application of the primary tool candidate model to JTTR patients would not reliably measure the model's external validity. Hence, this was not performed.

Performance of the models shortlisted as candidates for a secondary triage tool for JTTR patients is shown in online supplemental table 4B and model calibration is presented as online supplemental figure 4. The model selected as a secondary tool (XGB model, ID 37) achieved sensitivity of 97.6%, overtriage of 57.5% and AUC of 0.778 (figure 1). Secondary candidate models were evaluated on a subset of JTTR patients containing sufficient data to apply the BCD Triage Sieve (n=5455), thereby facilitating direct comparison (online supplemental table 5): the secondary tool candidate attained comparatively higher sensitivity (97.3% vs 80.2%), but had a higher overtriage rate (58.5% vs 47.4%).

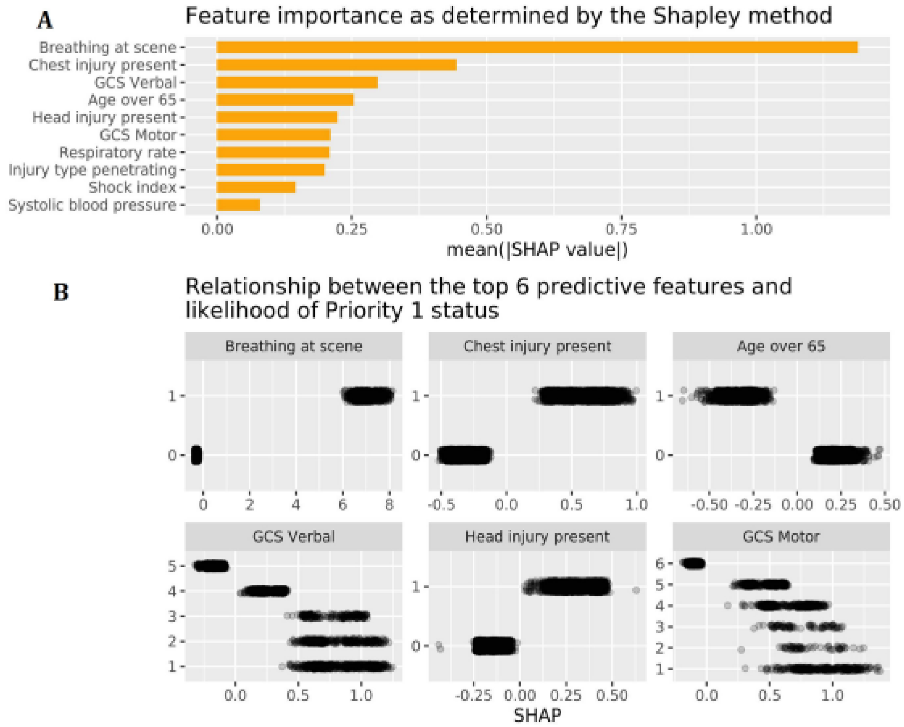


Figure 2 (A) Mean absolute Shapley value for the top 10 predictors. This is followed by the (B) Shapley values for the top six most important features (Shapley values are shown on the x axis, feature values are shown on the y axis). Large, positive Shapley values represent a greater contribution to the likelihood of P1 status. Negative Shapley values represent contributions to non-P1 status. Age over 65 years was found to be negatively predictive of P1 status. GCS Motor, motor subcomponent of the GCS; P1, priority 1.

DISCUSSION

We have developed MI triage tools based on machine learning models that outperform 10 existing international triage tools in predicting the need for time-critical interventions (P1 status) among adults. The best existing primary triage tool, the BCD Triage Sieve, demonstrated sensitivity of 68.2% and overtriage of 80.8% (AUC 0.688), while the selected machine learning primary triage tool achieved a sensitivity of 73% and overtriage of 77% (AUC 0.782). The model selected as a secondary MI triage tool achieved sensitivity of 77.9% and an overtriage rate of 76.4% (AUC 0.817). When externally validated, the secondary tool demonstrated excellent performance with sensitivity of 97.6% and overtriage of 57.5% (AUC 0.778). External validation of the primary tool was precluded by a lack of GCS subcomponent data within the UK combat casualty registry. A novel aspect of this exercise was including anatomical assessment of injuries as part of an MI triage tool and presence of a chest injury was found to be one of the most important variables. Our models serve as evidence-based alternatives to existing tools.

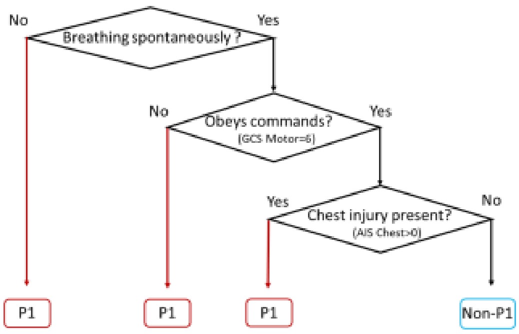


Figure 3 AIS, Abbreviated Injury Severity; P1, priority 1.

The models proposed are based entirely on qualitative assessments. Eliminating arithmetic calculations (RR and HR) from triage under challenging circumstances has been advocated by expert consensus.[19] The proposed four-variable secondary tool may also reduce triage time relative to the seven-step NARU and BCD Triage Sieve tools. In addition, decision support using portable device applications has established utility in the MI setting, exemplified by CitizenAID, which enables mutual aid by members of the general public.[30] Triage using a portable device could help to minimise interuser variability and human error.

Breathing status was the most important predictor of P1 status; this constitutes the opening step in several existing tools.[6] Our study concurs with the findings of Wallis and Carley, who determined that the GCS Motor component was strongly predictive of P1 status.[31] The finding that age >65 years is negatively associated with P1 status may be confounded by the predominantly

Figure 4 An interactive online application is demonstrated at https://ywxtriageapp.shinyapps.io/mltriage/. AIS, Abbreviated Injury Severity; GCS Motor, motor subcomponent of the GCS; ML, machine learning; P1, priority 1.

low-risk injury mechanism (low-level falls) in elders in our training dataset: hence, these patients are a poor surrogate for elders injured in an MI. Further work is required to develop effective trauma triage tools for elders, who differ in their physiology, and in whom presence of comorbidities and/or frailty is an important determinant of outcome.[13] Penetrating mechanism was also an important predictor of P1 status: MIs involving penetrating trauma have historically yielded larger proportions of P1 patients.[5]

A key strength of this study is use of a large sample of injured patients using prospective data collected by trained TARN coordinators.[17] The primary outcome measure chosen for this study is the only validated outcome measure for MI triage tool performance.[10] A further strength is that the proposed secondary triage tool has undergone blinded, external validation using the UK military's JTTR database. This provides estimates of the model's predictive capability overall, but importantly, also among patients with blast and penetrating mechanisms (under-represented in the TARN dataset) typical of terrorist attacks, the prevalent type of UK MI in recent years.[1] Selection of an XGB model as a secondary tool, which can make predictions in the context of some missing data, has avoided the possible bias which can result from multiple imputation. Importantly, based on the TARN patients included in our study, both novel tools would generate proportions of P1 casualties that fall within UK national mass casualty planning assumptions.[32] Notably, no UK or international guidance exists to define acceptable rates of undertriage and overtriage in the major incident setting.

Limitations of this study include use of retrospectively calculated AIS scores (incorporating CT and operative findings) during modelling in place of documented prehospital clinical assessment. While paramedics routinely conduct anatomical assessments during triage in singly injured patients using existing field triage tools and clinical assessment has proven effective in ruling out clinically significant chest injuries, some overtriage can be expected.[12 33] Clinicians have performed improvised anatomical-based secondary triage following two mass shooting incidents, with a subsequent low rate of undertriage.[5] Another limitation is the use of singly injured patients within a civilian trauma registry as surrogates for those injured in an MI; outcomes in the MI setting may be worse. Our models focus on predicting P1 status only: however, these patients are at greatest risk of preventable death. In current UK practice, a small proportion of P1 patients may be subsequently assigned P4/expectant status by a senior clinician at scene; this contrasts with practice elsewhere, where triage tools fulfil this role (eg, Australian CareFlight and US START tools).[6 32] Exclusion of P4 patients (<1% of the sample size) from the modelling process is unlikely to have impacted significantly on study findings. Application of models to the first recorded hospital physiology in JTTR may be biased by prehospital interventions; however, collection of prehospital physiological data during combat is particularly challenging.[28] The results of external validation in a military trauma population may have limited generalisability to the civilian setting. Further validation of our models in a true MI dataset or a prospective UK civilian database, including blast/penetrating trauma and burns, would provide further assurance of the models' performance. A further limitation is that we were unable to externally validate our proposed primary tool due to the paucity of prehospital vital signs (GCS) documented in the JTTR dataset.

In conclusion, using machine learning, we developed primary and secondary triage tools which differ from prior tools by incorporating anatomical assessment and have superior sensitivity and more favourable overtriage rates. Although the primary tool requires external validation among patients with injuries similar to those sustained in MI, the proposed secondary triage tool, which was externally validated, may be suitable for use in civilian hospital reception areas and in the military evacuation chain during MIs prior to or in conjunction with senior clinician triage using a portable device.

Author affiliations
[1]Centre for Computational Biology, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK
[2]Health Data Science Centre, University of Birmingham, Birmingham B15 2TT, UK
[3]NIHR Surgical Reconstruction Microbiology Research Centre, Edgbaston, UK
[4]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

[5]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK
[6]Academic Department of Military Surgery & Trauma, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B152WB, UK
[7]Emergency Department, Derriford Hospital, Plymouth, UK
[8]Academic Department of Military Emergency Medicine, Royal Centre for Defence Medicine, Mindelsohn Way, Edgbaston, Birmingham B15 2WB, UK
[9]UK Strategic Command, Northwood Headquarters, Northwood, UK
[10]Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[11]MRC Health Data Research UK (HDR UK), Birmingham, UK

Contributors NM and YX contributed equally to this study. NM and GG designed the study. YX, NM and SC accessed the TARN database, verified the underlying data and conducted analysis. JV and YX accessed JTTR data, JV verified the underlying data and YX conducted analysis. All authors contributed to data interpretation. NM (clinician) and YX (machine learning expert) wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approved the final version.

Funding This study is funded by the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre. GG acknowledges support from the NIHR Birmingham ECMC, Nanocommons H2020-EU (731032), MAESTRIA (grant agreement ID 965286) and the MRC Health Data Research UK (HDRUK/CFC/01), an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities.

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, the Medical Research Council, the Department of Health and Social Care, or the Ministry of Defence.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval The UK Health Research Authority Patient Information Advisory Group (Section 20) has granted ethical approval and waived the requirement for individual patient consent for research using anonymised TARN data. The Ministry of Defence (through its Medical Directorate) granted approval for the use of anonymised JTTR data.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. All data relevant to the study are included in the article or uploaded as supplemental information. De-identified patient data used for this study are proprietary to the Trauma Audit and Research Network (TARN), University of Manchester, and may be requested directly from TARN.

ORCID iD
Nabeela Malik http://orcid.org/0000-0003-0791-4158

REFERENCES

1 Dark P, Smith M, Ziman H, *et al*. Healthcare system impacts of the 2017 Manchester arena bombing: evidence from a national trauma registry patient case series and hospital performance data. *Emerg Med J* 2021;38:746–55.
2 Frykberg ER, Tepas JJ. Terrorist bombings. Lessons learned from Belfast to Beirut. *Ann Surg* 1988;208:569–76.
3 Hirsch M, Carli P, Nizard R, *et al*. The medical response to multisite terrorist attacks in Paris. *Lancet* 2015;386:2535–8.
4 Moran CG, Webb C, Brohi K, *et al*. Lessons in planning from mass casualty events in UK. *BMJ* 2017;359:j4765.
5 Turner CDA, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. *Crit Care* 2016;20:362.
6 Jenkins JL, McCarthy ML, Sauer LM, *et al*. Mass-casualty triage: time for an evidence-based approach. *Prehosp Disaster Med* 2008;23:3–8.
7 Lerner EB, Schwartz RB, Coule PL, *et al*. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. *Disaster Med Public Health Prep* 2008;2 Suppl 1:S25–34.
8 Group ALS. *Major incident medical management and support*. 3rd Edition ed. Wiley-Blackwell, 2011.
9 Jain TN, Ragazzoni L, Stryhn H, *et al*. Comparison of the Sacco triage method versus START triage using a virtual reality scenario in advance care paramedic students. *CJEM* 2016;18:288–92.
10 Malik NS, Chernbumroong S, Xu Y, *et al*. The BCD triage sieve outperforms all existing major incident triage tools: comparative analysis using the UK national trauma registry population. *EClinicalMedicine* 2021;36:100888.
11 Vassallo J, Smith J. Major incident triage and the evaluation of the triage sort as a secondary triage method. *Emerg Med J* 2019;36:281–6.
12 Trauma ACoSCo. *Resources for optimal care of the injured patient*. 2014.
13 Sammy I, Lecky F, Sutton A, *et al*. Factors affecting mortality in older trauma patients-a systematic review and meta-analysis. *Injury* 2016;47:1170–83.
14 Klug M, Barash Y, Bechler S, *et al*. A gradient boosting machine learning model for predicting early mortality in the emergency department triage: devising a nine-point triage score. *J Gen Intern Med* 2020;35:220–7.
15 Ramlakhan S, Saatchi R, Sabir L, *et al*. Understanding and interpreting artificial intelligence, machine learning and deep learning in emergency medicine. *Emerg Med J* 2022;39:380–5.
16 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:735–6.
17 The University of Manchester. The trauma audit and research network. n.d. Available: www.tarn.ac.uk
18 Lerner EB, McKee CH, Cady CE, *et al*. A consensus-based gold standard for the evaluation of mass casualty triage systems. *Prehosp Emerg Care* 2015;19:267–71.
19 Model uniform core criteria for mass casualty triage. *Disaster Med Public Health Prep* 2011;5:125–8.
20 Bazarian JJ, Eirich MA, Salhanick SD. The relationship between pre-hospital and emergency department Glasgow coma scale scores. *Brain Inj* 2003;17:553–60.
21 Cobas MA, De la Peña MA, Manning R, *et al*. Prehospital Intubations and mortality: a level 1 trauma center perspective. *Anesth Analg* 2009;109:489–93.
22 Jost E, Roberts DJ, Penney T, *et al*. Accuracy of clinical, laboratory, and computed tomography findings for identifying hollow viscus injury in blunt trauma patients with unexplained intraperitoneal free fluid without solid organ injury. *Am J Surg* 2017;213:874–80.
23 Domeier RM, Evans RW, Swor RA, *et al*. The reliability of prehospital clinical evaluation for potential spinal injury is not affected by the mechanism of injury. *Prehosp Emerg Care* 1999;3:332–7.
24 Vandromme MJ, Griffin RL, Kerby JD, *et al*. Identifying risk for massive transfusion in the relatively normotensive patient: utility of the prehospital shock index. *J Trauma* 2011;70:384–90.
25 Raita Y, Goto T, Faridi MK, *et al*. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23:64.
26 Soltan AAS, Kouchaki S, Zhu T, *et al*. Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Health* 2021;3:e78–87.
27 Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature Attribution for tree ensembles. *arXiv Preprint arXiv* 2018:180203888.
28 Hettiaratchy S, Tai N, Mahoney P, *et al*. UK's NHS trauma systems: lessons from military experience. *The Lancet* 2010;376:149–51.
29 Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
30 citizenAID A UK charity empowering the public to save lives. Available: https://www.citizenaid.org/ [Accessed 21 Jan 2022].
31 Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. *Emerg Med J* 2006;23:475–8.
32 NHS England Emergency Preparedness RaR. *Concept of operations for managing mass casualties*. Emergency Preparedness RaR, 2017.
33 Rodriguez RM, Anglin D, Langdorf MI, *et al*. NEXUS chest: validation of a decision instrument for selective chest imaging in blunt trauma. *JAMA Surg* 2013;148:940–6.

Triage in major incidents: development and external validation of novel machine-learning derived primary and secondary triage tools

Supplementary material

Additional details of machine learning modelling

An overview of study methodology and data processing is presented in Supplementary Figure 1, with a more detailed model development and selection strategy outlined in Supplementary Figure 2.

For the decision tree (also known as Recursive Partitioning And Regression Tree, RPART) method, a limit of a maximum tree depth of 3 was imposed for ease of interpretation.To guard against overfitting, we chose to tune the cost complexity and the tree depth parameter of the decision tree model. Effectively, the tree depth (distance from the root to a terminal node) represents the number of measurables needed in order to determine a triage category. However, we note that unlike triage tools conceived by human experts, it is possible to have the same variable used more than once to split the nodes, if the reuse of variables reduces classification error. A deep tree with many splits tends to overfit the data, and makes it difficult to adapt the model to a tool that can be implemented in practice.

Both random forest (RF) and gradient boosted tree (XGB) are popular machine learning algorithms with strong predictive power. RF is based on averaging an ensemble of trees and the idea of bagging, which lowers the prediction variance. Furthermore, instead of growing each tree using all variables, it randomly chooses a subset of variables at each split of the node in the tree, thereby forcing it to learn through all subsets of available variables. For XGB, the prediction target is estimated by sum-of-trees, and the model is built by successively fitting each tree to the residue of previously fitted trees with no single tree dominating the prediction, while regularizing the fit through multiplication by a scaling factor known as learning rate. In short, XGB estimates the target function by a sum of trees each of which explains a small and different portion of the target and no single tree dominates the prediction.

For the L1-regularized logistic regression model, the penalty parameter, specifying the amount of regularization, was tuned. We add a regularization term in logistic regression so that the solution is well-defined even if the data are perfectly linearly separable.

Initially, models were trained using all 13 input variables (summarised in Supplementary Table 1): the resulting models would be too complex for practical application as tools, but nonetheless act as a useful comparator for model performance (see Supplementary Figure 2 detailing the model building and selection strategy). The optimal hyperparameters that yield the best AUC were selected. For decision tree and logistic regression, a grid search was used; whereas for RF and XGB, random sampling of points in the parameter space was used to try to cover the space as uniformly as possible. For each model, having selected the hyperparameters, a final model was trained on the whole training set (70% of TARN data) and then evaluated on the remaining 30% hold-out data. Models developed using all 13 input variables yielded similar AUC values (range 0.862-0.868, see Supplementary Table 2), except for the decision tree model (AUC 0.782), which also exhibited lower specificity and higher over-triage than the other ML models. All models employing 13 variables attained sensitivity above 72%, exceeding that of the BCD Triage Sieve. Performance characteristics of models employing all 13 input variables were further evaluated by age subgroup (16-64 years and 65+ years (Supplementary Table 2). We note that for ML models evaluated on the 65+ group, while there is slight decrease in AUC compared to the 16-64 group, sensitivity is much worse, except for the decision tree model which has the best sensitivity (66.3%) among all models and triage tools. However, the price of this relatively high sensitivity of decision tree is a high over-triage rate (87.2%).

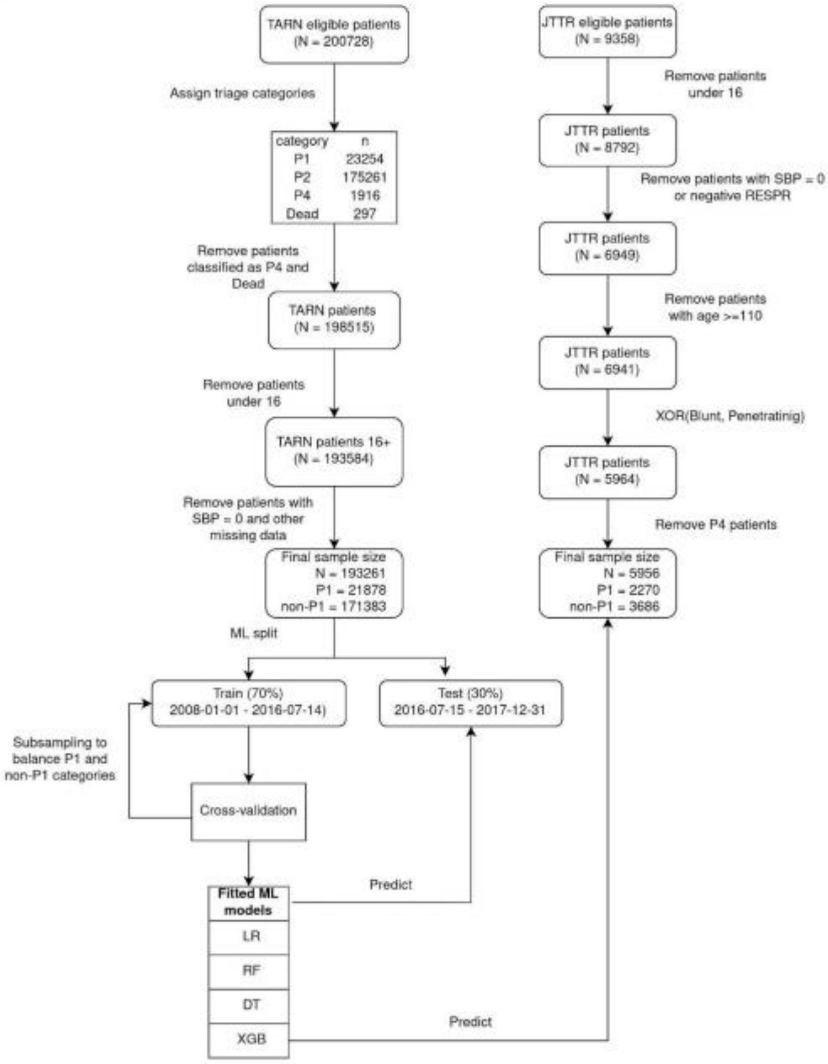
Existing triage tools were applied to the internal validation dataset to act as comparators to the models proposed as novel triage tools. To overcome the over-representation of elders (65+ years) within the TARN database (see Manuscript, Table 1), who also differ in their physiology to younger adults, tool performance was additionally tested in subgroups by age (16-64 years and 65+ years), as shown in Supplementary Table 3. Existing tools demonstrated lower sensitivity and higher over-triage rates amongst elders compared to younger adults (16-64 years).

We sought to combine the individual models in a weighted fashion by training a super model [1], in which weights are assigned to models based on their predictive power and the final predictions are driven by models with high weights. For the super model, a binomial likelihood maximization using the BFGS quasi-Newton optimization method was used, the model was fitted using the “SuperLearner” R package [2]. The weights are normalized and sum to one. The super model assigned coefficients (weights) to each individual model, along with the minimized risks. We note that the decision tree model was in fact excluded from the super model, since it has a weight of zero (risk 0.525). The XGB model has highest weight of 0.717 (risk 0.442). Random forest had the second highest weight (0.241, risk 0.454) whilst logistic regression had a low contribution to the overall super model (coefficient 0.041, risk 0.451). The AUC for the super model is 0.868.

The importance of individual features (input parameters) was also estimated using the XGB method (see Supplementary Figure 3). This method yielded similar rankings to those generated by the TREEShap method: breathing status contributed 36% of the total gain, followed by presence of chest injury (13%) and GCS verbal score (11%).

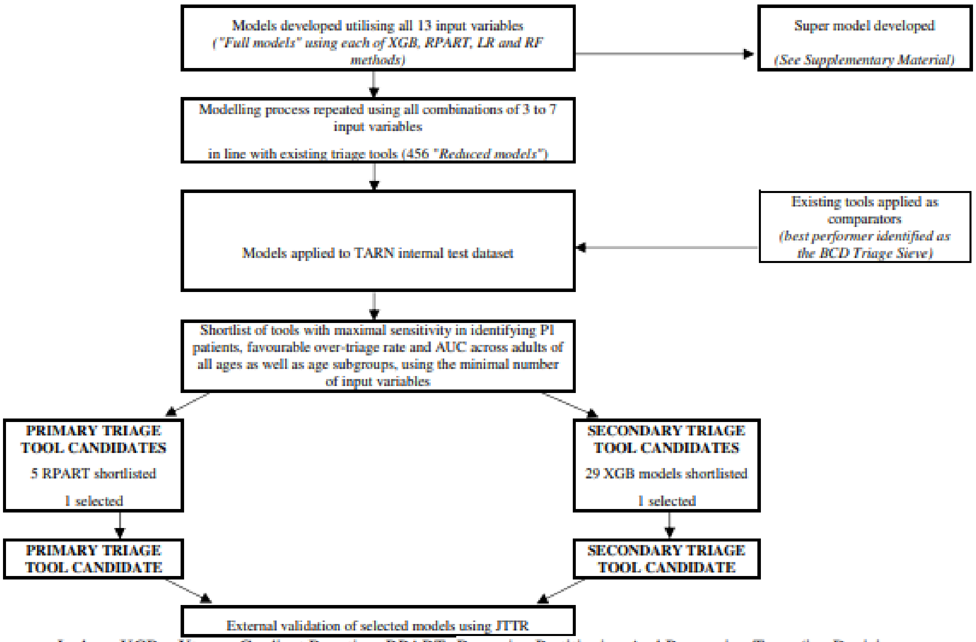
Secondary candidate models were subsequently evaluated on a smaller subset of JTTR patients (n=5455) for which there is complete data available to test the performance of the BCD Triage Sieve, thereby facilitating direct comparison (Supplementary Table 6). The secondary tool candidate (XGB 37) attained comparatively high sensitivity (97.3% vs 80.2%), although this was associated with an 11.1% increase in over-triage (58.5% vs 47.4%).

Supplementary Figure 1: Overview of study methodology and data processing



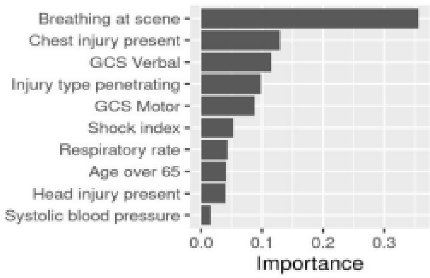
Ledger: Abbreviations: TARN= Trauma Audit and Research Network; JTTR= Joint Theatre Trauma Registry; SBP=Systolic Blood Pressure; XOR=Exclusive/or; LR=Logistic Regression; RF=Random Forest; DT=Decision Tree; XGB= Extreme Gradient Boost; RESPR=Respiratory rate

Supplementary Figure 2: Model building and selection strategy



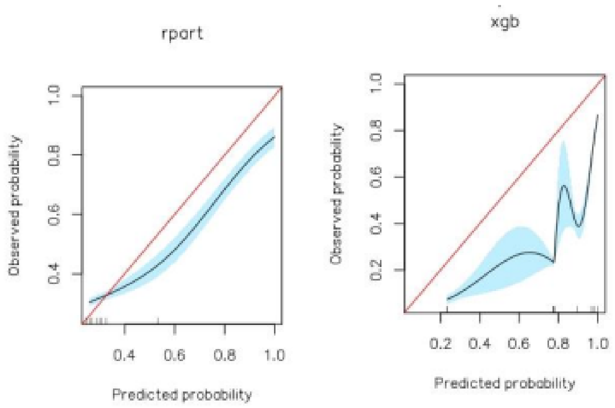
Ledger: XGB=eXtreme Gradient Boosting, RPART=Recursive Partitioning And Regression Trees (i.e. Decision Tree), LR=Logistic regression and RF=random forest, TARN=Trauma Audit and Research Network Registry, JTTR=Joint Theatre Trauma Registry, BCD Triage Sieve=Battlefield Casualty Drills Triage Sieve.

Supplementary Figure 3: Feature importance plot for the XGB model



Ledger: Importance of top 10 predictors for the XGB model as measured by the fractional contribution of each feature to the model based on the total gain of each feature's splits. High values represent more predictive features. Respiratory rate is measured in breaths per minute, Systolic blood pressure is measured in mmHg. Presence of chest and head injuries are denoted by a positive Abbreviated Injury Severity score.

Supplementary Figure 4: Calibration plot for models selected as candidate primary and secondary triage tools



Ledger: calibration plot for the candidate primary (left) and secondary (right) ML models, evaluated using JTTR data.

The calibration curve was estimated by natural splines using the R package gbm [3]. 95% confidence intervals covering 2 standard errors are demonstrated (blue).

For perfect calibration, the calibration curve would align with the 45-degree line (red). It can be seen that the secondary tool (XGB model) over-predicted risk, since the predicted P1 probabilities were greater than the observed probabilities across all patients. This is expected as the secondary tool candidate (XGB model) had high sensitivity but low specificity. In contrast, the calibration curve of the primary tool candidate (decision tree or rpart model) was smoother and the over-prediction was less extreme than XGB, reflecting the fact that the decision tree model had lower sensitivity and higher specificity than XGB.

Supplementary Table 1: Clinical parameters included as input variables for modelling

	Input variables
Physiological parameters*	Heart rate (beats per minute), Respiratory rate (breaths per minute) Systolic blood pressure (mmHg) Ability to breathe spontaneously** GCS Verbal component GCS Motor component GCS Eyes component Shock index***
Anatomical parameters	Presence (AIS>0) or absence (AIS=0) of injury in the following anatomical regions: Head Face Thorax Limb
Age	Age 65 and over (Binary – Yes or No)
Injury Mechanism*	Blunt or penetrating injury

Ledger: GCS=Glasgow Coma Score, AIS=Abbreviated Injury Score. *First recorded pre-hospital physiological parameters and injury mechanism were utilised. **All patients who underwent an advanced airway intervention in the pre-hospital environment were assumed to be unable to breathe. ***Shock index=heart rate/systolic blood pressure.

Supplementary Table 2: Performance of machine learning models utilising all 13 input variables in predicting P1 status amongst patients in the internal (TARN) testing dataset

	Method	Sensitivity	Specificity	Under-triage	Over-triage	AUC
All adults (16+ years)	ml_rpart	73.0 [71.8, 74.2]	73.9 [73.5, 74.3]	27.0 [25.8, 28.2]	77.0 [76.4, 77.7]	0.782 [0.775, 0.789]
	ml_rf	72.6 [71.4, 73.8]	86.0 [85.7, 86.3]	27.4 [26.2, 28.6]	64.5 [63.6, 65.4]	0.867 [0.861, 0.873]
	ml_xgb	72.7 [71.5, 73.9]	85.9 [85.6, 86.2]	27.3 [26.1, 28.5]	64.6 [63.7, 65.5]	0.868 [0.862, 0.874]
	ml_lr	72.2 [71.0, 73.4]	85.2 [84.9, 85.5]	27.8 [26.6, 29.0]	65.8 [64.9, 66.6]	0.862 [0.857, 0.868]
16-64 years subgroup	ml_rpart	76.0 [74.6, 77.3]	71.8 [71.2, 72.4]	24.0 [22.7, 25.4]	66.8 [65.8, 67.8]	0.794 [0.786, 0.803]
	ml_rf	81.9 [80.7, 83.1]	76.1 [75.5, 76.7]	18.1 [16.9, 19.3]	61.2 [60.2, 62.3]	0.877 [0.871, 0.884]
	ml_xgb	82.3 [81.0, 83.5]	75.7 [75.2, 76.3]	17.7 [16.5, 19.0]	61.5 [60.4, 62.6]	0.879 [0.872, 0.885]
	ml_lr	82.5 [81.2, 83.7]	74.7 [74.1, 75.3]	17.5 [16.3, 18.8]	62.5 [61.4, 63.5]	0.873 [0.866, 0.879]
65+ years subgroup	ml_rpart	66.3 [64.0, 68.6]	75.3 [74.8, 75.8]	33.7 [31.4, 36.0]	87.2 [86.5, 87.9]	0.746 [0.733, 0.759]
	ml_rf	51.7 [49.3, 54.1]	92.5 [92.2, 92.8]	48.3 [45.9, 50.7]	72.7 [71.1, 74.2]	0.806 [0.793, 0.818]
	ml_xgb	51.3 [48.9, 53.6]	92.6 [92.3, 92.9]	48.7 [46.4, 51.1]	72.5 [70.9, 74.0]	0.807 [0.795, 0.820]
	ml_lr	49.2 [46.8, 51.5]	92.2 [91.9, 92.5]	50.8 [48.5, 53.2]	74.3 [72.8, 75.8]	0.800 [0.787, 0.812]

Ledger: Results shown are percentages (except for AUC). The best performing model amongst all adults for each method is shown, including performance by age subgroup. Abbreviations: ml=machine learning, rpart= decision tree, rf=random forest, xgb= extreme gradient boosting, lr=logistic regression.

Supplementary Table 3: Performance characteristics of existing triage tools when applied to the internal validation dataset

Tool	Sensitivity	Specificity	Under-triage	Over-triage	AUC
All adults (16+ years)					
BCD Triage Sieve	68.2 [66.9, 69.4]	69.5 [69.1, 69.9]	31.8 [30.6, 33.1]	80.8 [80.2, 81.3]	0.688 [0.682, 0.695]
CareFlight	39.9 [38.6, 41.2]	94.5 [94.3, 94.7]	60.1 [58.8, 61.4]	56.4 [55.0, 57.8]	0.672 [0.666, 0.679]
JumpSTART	42.5 [41.2, 43.8]	92.1 [91.8, 92.3]	57.5 [56.2, 58.8]	63.7 [62.5, 64.9]	0.673 [0.666, 0.679]
MIMMS Triage Sieve	40.5 [39.2, 41.8]	92.0 [91.8, 92.3]	59.5 [58.2, 60.8]	64.9 [63.7, 66.1]	0.663 [0.656, 0.669]
MPTT	50.5 [49.2, 51.8]	62.4 [62.0, 62.8]	49.5 [48.2, 50.8]	87.5 [87.1, 87.9]	0.565 [0.558, 0.571]
MPTT-24	48.4 [47.1, 49.7]	66.4 [66.0, 66.8]	51.6 [50.3, 52.9]	86.7 [86.2, 87.2]	0.574 [0.567, 0.581]
MSTART	54.9 [53.6, 56.2]	88.4 [88.1, 88.7]	45.1 [43.8, 46.4]	66.5 [65.5, 67.5]	0.717 [0.710, 0.723]
NARU Triage Sieve	43.0 [41.7, 44.3]	88.3 [88.1, 88.6]	57.0 [55.7, 58.3]	71.8 [70.9, 72.8]	0.657 [0.650, 0.663]
RAMP	37.1 [35.9, 38.4]	94.6 [94.5, 94.8]	62.9 [61.6, 64.1]	57.5 [56.1, 58.9]	0.659 [0.653, 0.665]
START	51.8 [50.5, 53.2]	90.0 [89.7, 90.2]	48.2 [46.8, 49.5]	64.5 [63.5, 65.6]	0.709 [0.702, 0.716]
16-64 years subgroup					
BCD Triage Sieve	72.7 [71.2, 74.1]	64.8 [64.2, 65.5]	27.3 [25.9, 28.8]	72.4 [71.5, 73.3]	0.687 [0.680, 0.695]
CareFlight	42.7 [41.1, 44.3]	94.3 [94.0, 94.6]	57.3 [55.7, 58.9]	42.0 [40.2, 43.8]	0.685 [0.677, 0.693]
JumpSTART	45.5 [43.9, 47.0]	91.2 [90.9, 91.6]	54.5 [53.0, 56.1]	51.1 [49.4, 52.7]	0.684 [0.675, 0.692]
MIMMS Triage Sieve	43.0 [41.4, 44.6]	92.6 [92.2, 93.0]	57.0 [55.4, 58.6]	48.2 [46.5, 50.0]	0.678 [0.670, 0.686]
MPTT	52.3 [50.7, 53.9]	57.1 [56.4, 57.8]	47.7 [46.1, 49.3]	81.6 [80.9, 82.4]	0.547 [0.538, 0.555]
MPTT-24	50.5 [48.9, 52.1]	61.6 [60.9, 62.2]	49.5 [47.9, 51.1]	80.5 [79.7, 81.3]	0.560 [0.552, 0.569]
MSTART	57.6 [56.1, 59.2]	88.9 [88.5, 89.3]	42.4 [40.8, 43.9]	51.0 [49.6, 52.5]	0.733 [0.725, 0.741]
NARU Triage Sieve	47.1 [45.5, 48.7]	87.5 [87.0, 87.9]	52.9 [51.3, 54.5]	59.0 [57.6, 60.5]	0.673 [0.665, 0.681]
RAMP	39.6 [38.1, 41.2]	94.4 [94.1, 94.7]	60.4 [58.8, 61.9]	43.4 [41.5, 45.3]	0.670 [0.662, 0.678]
START	54.4 [52.8, 55.9]	90.7 [90.3, 91.1]	45.6 [44.1, 47.2]	48.1 [46.5, 49.6]	0.725 [0.717, 0.733]
65+ years subgroup					
BCD Triage Sieve	58.1 [55.7, 60.4]	72.6 [72.1, 73.1]	41.9 [39.6, 44.3]	89.6 [89.0, 90.2]	0.653 [0.642, 0.665]
CareFlight	33.7 [31.4, 36.0]	94.6 [94.4, 94.9]	66.3 [64.0, 68.6]	74.5 [72.6, 76.3]	0.642 [0.630, 0.653]
JumpSTART	35.8 [33.5, 38.1]	92.6 [92.3, 92.9]	64.2 [61.9, 66.5]	79.1 [77.6, 80.6]	0.642 [0.630, 0.653]
MIMMS Triage Sieve	34.9 [32.6, 37.2]	91.7 [91.3, 92.0]	65.1 [62.8, 67.4]	81.4 [80.0, 82.8]	0.633 [0.621, 0.644]
MPTT	46.5 [44.1, 48.9]	65.9 [65.4, 66.5]	53.5 [51.1, 55.9]	93.1 [92.6, 93.5]	0.562 [0.550, 0.574]
MPTT-24	43.7 [41.3, 46.1]	69.6 [69.1, 70.1]	56.3 [53.9, 58.7]	92.7 [92.2, 93.2]	0.567 [0.555, 0.579]
MSTART	48.9 [46.5, 51.3]	88.0 [87.7, 88.4]	51.1 [48.7, 53.5]	81.8 [80.6, 82.9]	0.685 [0.673, 0.697]
NARU Triage Sieve	33.7 [31.5, 36.0]	88.9 [88.6, 89.3]	66.3 [64.0, 68.5]	85.8 [84.7, 86.8]	0.613 [0.602, 0.625]
RAMP	31.6 [29.4, 33.8]	94.8 [94.6, 95.1]	68.4 [66.2, 70.6]	75.1 [73.2, 76.9]	0.632 [0.621, 0.643]
START	46.2 [43.8, 48.6]	89.5 [89.1, 89.8]	53.8 [51.4, 56.2]	80.7 [79.5, 81.9]	0.678 [0.666, 0.690]

Ledger: BCD Triage Sieve=Battlefield Casualty Drills Triage Sieve (UK Military), CareFlight (Australia), JumpSTART (US paediatric triage tool), MIMMS Triage Sieve=Major Incident Medical Management and Support Triage Sieve, MPTT=Modified Physiological Triage Tool (tool modelled in UK military casualties), MPTT-24 (modification of MPTT, 2017), START=Simple Triage and Rapid Treatment (US adult tool), MSTART=modified START, NARU Triage Sieve=National Ambulance Resilience Unit Triage Sieve (Current UK civilian tool), RAMP=Rapid Assessment of Mentation and Pulse (New York Fire Department).

Supplementary Tables 4A-C: See landscape format document

Supplementary Table 5: External validation of shortlisted models and the Battlefield Casualty Drills Triage Sieve (comparator) using the Joint Theatre Trauma Registry (n=5455)

Method	Sensitivity	Specificity	Under-triage	Over-triage	AUC
Comparator (best existing tool):					
BCD Triage Sieve	0.802 [0.784, 0.819]	0.578 [0.561, 0.595]	0.198 [0.181, 0.216]	0.474 [0.456, 0.492]	0.690 [0.678, 0.702]
Primary tool candidate models					
rpart_1	0.330 [0.310, 0.351]	0.892 [0.881, 0.902]	0.670 [0.649, 0.690]	0.360 [0.331, 0.390]	0.618 [0.606, 0.629]
rpart_3	0.330 [0.310, 0.351]	0.892 [0.881, 0.902]	0.670 [0.649, 0.690]	0.360 [0.331, 0.390]	0.618 [0.607, 0.630]
rpart_37	0.330 [0.310, 0.351]	0.892 [0.881, 0.902]	0.670 [0.649, 0.690]	0.360 [0.331, 0.390]	0.618 [0.607, 0.630]
rpart_52	0.479 [0.457, 0.501]	0.752 [0.737, 0.766]	0.521 [0.499, 0.543]	0.470 [0.447, 0.494]	0.611 [0.598, 0.624]
rpart_124	0.437 [0.415, 0.459]	0.879 [0.868, 0.890]	0.563 [0.541, 0.585]	0.322 [0.297, 0.348]	0.668 [0.656, 0.680]
Secondary tool candidate models					
xgb_1	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.755 [0.743, 0.768]
xgb_3	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.755 [0.743, 0.768]
xgb_37	0.973 [0.964, 0.979]	0.199 [0.186, 0.213]	0.027 [0.021, 0.036]	0.585 [0.571, 0.599]	0.780 [0.768, 0.792]
xgb_38	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.755 [0.743, 0.768]
xgb_41	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.782 [0.769, 0.795]
xgb_42	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.748 [0.734, 0.761]
xgb_43	0.675 [0.654, 0.695]	0.791 [0.777, 0.805]	0.325 [0.305, 0.346]	0.347 [0.326, 0.367]	0.776 [0.762, 0.790]
xgb_44	0.973 [0.964, 0.979]	0.199 [0.186, 0.213]	0.027 [0.021, 0.036]	0.585 [0.571, 0.599]	0.780 [0.768, 0.792]
xgb_53	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.748 [0.734, 0.761]
xgb_54	0.675 [0.654, 0.695]	0.791 [0.777, 0.805]	0.325 [0.305, 0.346]	0.347 [0.326, 0.367]	0.777 [0.763, 0.790]
xgb_121	0.973 [0.964, 0.979]	0.199 [0.186, 0.213]	0.027 [0.021, 0.036]	0.585 [0.571, 0.599]	0.780 [0.768, 0.792]
xgb_124	0.975 [0.966, 0.981]	0.174 [0.162, 0.187]	0.025 [0.019, 0.034]	0.592 [0.578, 0.606]	0.777 [0.764, 0.790]
xgb_125	0.973 [0.964, 0.979]	0.199 [0.186, 0.213]	0.027 [0.021, 0.036]	0.585 [0.571, 0.599]	0.778 [0.766, 0.790]
xgb_130	0.667 [0.646, 0.688]	0.788 [0.774, 0.802]	0.333 [0.312, 0.354]	0.353 [0.332, 0.374]	0.748 [0.734, 0.761]
xgb_131	0.674 [0.653, 0.695]	0.793 [0.779, 0.807]	0.326 [0.305, 0.347]	0.344 [0.324, 0.365]	0.774 [0.760, 0.788]
xgb_141	0.676 [0.655, 0.696]	0.792 [0.778, 0.805]	0.324 [0.304, 0.345]	0.346 [0.325, 0.367]	0.768 [0.754, 0.782]
xgb_154	0.965 [0.955, 0.972]	0.171 [0.159, 0.184]	0.035 [0.028, 0.045]	0.596 [0.582, 0.610]	0.695 [0.680, 0.709]
xgb_165	0.684 [0.663, 0.704]	0.792 [0.778, 0.805]	0.316 [0.296, 0.337]	0.343 [0.323, 0.364]	0.788 [0.774, 0.801]
xgb_166	0.676 [0.655, 0.696]	0.791 [0.777, 0.805]	0.324 [0.304, 0.345]	0.346 [0.326, 0.367]	0.770 [0.757, 0.784]
xgb_250	0.973 [0.964, 0.979]	0.199 [0.186, 0.213]	0.027 [0.021, 0.036]	0.585 [0.571, 0.599]	0.781 [0.769, 0.794]
xgb_251	0.972 [0.964, 0.979]	0.203 [0.190, 0.217]	0.028 [0.021, 0.036]	0.584 [0.570, 0.599]	0.792 [0.779, 0.804]
xgb_270	0.722 [0.702, 0.742]	0.701 [0.685, 0.716]	0.278 [0.258, 0.298]	0.416 [0.396, 0.435]	0.785 [0.772, 0.798]
xgb_271	0.670 [0.649, 0.691]	0.800 [0.786, 0.813]	0.330 [0.309, 0.351]	0.338 [0.318, 0.359]	0.770 [0.756, 0.784]
xgb_289	0.971 [0.962, 0.977]	0.195 [0.182, 0.209]	0.029 [0.023, 0.038]	0.587 [0.573, 0.601]	0.778 [0.765, 0.791]
xgb_311	0.721 [0.701, 0.741]	0.704 [0.689, 0.719]	0.279 [0.259, 0.299]	0.413 [0.393, 0.433]	0.781 [0.768, 0.795]
xgb_380	0.971 [0.962, 0.977]	0.196 [0.183, 0.210]	0.029 [0.023, 0.038]	0.587 [0.573, 0.601]	0.795 [0.782, 0.807]
xgb_392	0.976 [0.968, 0.982]	0.174 [0.162, 0.187]	0.024 [0.018, 0.032]	0.592 [0.578, 0.606]	0.800 [0.788, 0.812]
xgb_402	0.709 [0.688, 0.729]	0.732 [0.717, 0.747]	0.291 [0.271, 0.312]	0.394 [0.374, 0.414]	0.783 [0.769, 0.796]
xgb_417	0.976 [0.968, 0.982]	0.181 [0.168, 0.194]	0.024 [0.018, 0.032]	0.590 [0.576, 0.604]	0.799 [0.787, 0.812]

Ledger: *this is a reduced JTTR dataset for which a complete set of physiological data exists for application of the BCD Triage Sieve (the best performing existing tool, selected as a comparator) can be applied. To allow direct comparison, the models shortlisted as candidates for triage tools are applied to the reduced dataset. rpart=decision tree (Recursive Partitioning And Regression Trees), xgb= extreme gradient boosting.

References:

[1] Van der Laan, Mark J., Eric C. Polley, and Alan E. Hubbard. "Super learner." *Statistical applications in genetics and molecular biology* 6.1 (2007).

[2] Eric Polley, Erin LeDell, Chris Kennedy and Mark van der Laan (2019). SuperLearner: Super Learner Prediction. R package version 2.0-26. https://CRAN.R-project.org/package=SuperLearner

[3] Brandon Greenwell, Bradley Boehmke, Jay Cunningham and GBM Developers (2022). gbm: Generalized Boosted Regression Models. R package version 2.1.8.1. https://CRAN.R-project.org/package=gbm].

Supplementary Table 4A: Models shortlisted as candidates for primary and secondary tools: Description of model variables

<table>
<tr><th>Task ID</th><th>Model variables</th><th>Variables</th></tr>
<tr><td></td><td>Candidate primary tools (Decision Tree (RPART) models)</td><td></td></tr>
<tr><td>1</td><td>Breathing status at scene|Chest injury present|GCS Motor Score</td><td>3</td></tr>
<tr><td>3</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score</td><td>3</td></tr>
<tr><td>37</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type</td><td>4</td></tr>
<tr><td>52</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Respiratory rate</td><td>4</td></tr>
<tr><td>124</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|Respiratory rate</td><td>5</td></tr>
<tr><td></td><td>Candidate secondary tools (Extreme Gradient Boosting (XGB) models)</td><td></td></tr>
<tr><td>1</td><td>Breathing status at scene|Chest injury present|GCS Motor Score</td><td>3</td></tr>
<tr><td>3</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score</td><td>3</td></tr>
<tr><td>37</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type</td><td>4</td></tr>
<tr><td>38</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score</td><td>4</td></tr>
<tr><td>41</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Respiratory rate</td><td>4</td></tr>
<tr><td>42</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Head injury present</td><td>4</td></tr>
<tr><td>43</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Systolic blood pressure</td><td>4</td></tr>
<tr><td>44</td><td>Breathing status at scene|Chest injury present|Injury type|GCS Verbal Score</td><td>4</td></tr>
<tr><td>53</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Head injury present</td><td>4</td></tr>
<tr><td>54</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Systolic blood pressure</td><td>4</td></tr>
<tr><td>121</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|GCS Verbal Score</td><td>5</td></tr>
<tr><td>124</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|Respiratory rate</td><td>5</td></tr>
<tr><td>125</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|Head injury present</td><td>5</td></tr>
<tr><td>130</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score|Head injury present</td><td>5</td></tr>
<tr><td>131</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score|Systolic blood pressure</td><td>5</td></tr>
<tr><td>141</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Head injury present|Systolic blood pressure</td><td>5</td></tr>
<tr><td>154</td><td>Breathing status at scene|Chest injury present|Injury type|Respiratory rate|Head injury present</td><td>5</td></tr>
<tr><td>165</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Respiratory rate|Systolic blood pressure</td><td>5</td></tr>
<tr><td>166</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Head injury present|Systolic blood pressure</td><td>5</td></tr>
<tr><td>250</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|GCS Verbal Score|Head injury present</td><td>6</td></tr>
<tr><td>251</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|GCS Verbal Score|Systolic blood pressure</td><td>6</td></tr>
<tr><td>270</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score|Respiratory rate|Systolic blood pressure</td><td>6</td></tr>
<tr><td>271</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score|Head injury present|Systolic blood pressure</td><td>6</td></tr>
<tr><td>289</td><td>Breathing status at scene|Chest injury present|Injury type|GCS Verbal Score|Respiratory rate|Head injury present</td><td>6</td></tr>
<tr><td>311</td><td>Breathing status at scene|Chest injury present|GCS Verbal Score|Respiratory rate|Head injury present|Systolic blood pressure</td><td>6</td></tr>
<tr><td>380</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|GCS Verbal Score|Respiratory rate|Head injury present</td><td>7</td></tr>
<tr><td>392</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|Injury type|Respiratory rate|Head injury present|Systolic blood pressure</td><td>7</td></tr>
<tr><td>402</td><td>Breathing status at scene|Chest injury present|GCS Motor Score|GCS Verbal Score|Respiratory rate|Head injury present|Systolic blood pressure</td><td>7</td></tr>
<tr><td>417</td><td>Breathing status at scene|Chest injury present|Injury type|GCS Verbal Score|Respiratory rate|Head injury present|Systolic blood pressure</td><td>7</td></tr>
</table>

Supplementary Table 4B: Performance characteristics of models shortlisted as primary and secondary tool candidates

Task ID	Internal validation using TARN testing dataset (all adult patients)					External validation in JTTR (n=5956)				
	Sensitivity	**Specificity**	**Under-triage**	**Over-triage**	**AUC**	**Sensitivity**	**Specificity**	**Under-triage**	**Over-triage**	**AUC**
RPART:										
1	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·772 [0·765, 0·779]	34·0 [32·1, 36·0]	89·0 [87·9, 90·0]	66·0 [64·0, 67·9]	34·4 [31·7, 37·2]	0·622 [0·611, 0·633]
3	72·3 [71·1, 73·5]	74·5 [74·1, 74·8]	27·7 [26·5, 28·9]	76·9 [76·2, 77·5]	0·775 [0·768, 0·782]	34·0 [32·1, 36·0]	89·0 [87·9, 90·0]	66·0 [64·0, 67·9]	34·4 [31·7, 37·2]	0·623 [0·612, 0·634]
37	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·782 [0·775, 0·789]	49·8 [47·7, 51·9]	74·5 [73·0, 75·9]	50·2 [48·1, 52·3]	45·4 [43·3, 47·6]	0·616 [0·604, 0·629]
52	72·3 [71·1, 73·5]	74·5 [74·1, 74·8]	27·7 [26·5, 28·9]	76·9 [76·2, 77·5]	0·780 [0·773, 0·787]	34·0 [32·1, 36·0]	89·0 [87·9, 90·0]	66·0 [64·0, 67·9]	34·4 [31·7, 37·2]	0·623 [0·612, 0·634]
124	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·777 [0·770, 0·783]	46·1 [44·0, 48·2]	86·3 [85·2, 87·4]	53·9 [51·8, 56·0]	32·5 [30·2, 34·9]	0·671 [0·659, 0·683]
XGB:										
1	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·783 [0·776, 0·790]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·754 [0·742, 0·766]
3	72·3 [71·1, 73·5]	74·5 [74·1, 74·8]	27·7 [26·5, 28·9]	76·9 [76·2, 77·5]	0·796 [0·789, 0·803]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·754 [0·742, 0·766]
37	77·9 [76·8, 79·0]	73·1 [72·7, 73·5]	22·1 [21·0, 23·2]	76·4 [75·8, 77·0]	0·817 [0·810, 0·824]	97·6 [96·8, 98·2]	18·6 [17·4, 19·9]	2·4 [1·8, 3·2]	57·5 [56·2, 58·9]	0·778 [0·766, 0·790]
38	73·8 [72·6, 75·0]	73·6 [73·2, 73·9]	26·2 [25·0, 27·4]	77·1 [76·5, 77·7]	0·798 [0·792, 0·805]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·754 [0·742, 0·766]
41	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·808 [0·801, 0·815]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·779 [0·767, 0·792]
42	73·0 [71·8, 74·2]	73·9 [73·5, 74·3]	27·0 [25·8, 28·2]	77·0 [76·4, 77·7]	0·798 [0·791, 0·805]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·749 [0·736, 0·762]
43	70·0 [68·7, 71·2]	78·5 [78·2, 78·9]	30·0 [28·8, 31·3]	74·3 [73·6, 75·0]	0·809 [0·802, 0·816]	71·2 [69·3, 73·1]	74·0 [72·5, 75·4]	28·8 [26·9, 30·7]	37·2 [35·4, 39·1]	0·775 [0·762, 0·788]
44	77·3 [76·2, 78·4]	73·6 [73·3, 74·0]	22·7 [21·6, 23·8]	76·2 [75·6, 76·8]	0·830 [0·824, 0·837]	97·6 [96·8, 98·2]	18·6 [17·4, 19·9]	2·4 [1·8, 3·2]	57·5 [56·2, 58·9]	0·778 [0·766, 0·790]
53	72·3 [71·1, 73·5]	74·5 [74·1, 74·8]	27·7 [26·5, 28·9]	76·9 [76·2, 77·5]	0·802 [0·795, 0·809]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·749 [0·736, 0·762]
54	70·7 [69·5, 71·9]	77·7 [77·4, 78·1]	29·3 [28·1, 30·5]	74·7 [74·0, 75·4]	0·816 [0·810, 0·823]	71·2 [69·3, 73·1]	74·0 [72·5, 75·4]	28·8 [26·9, 30·7]	37·2 [35·4, 39·1]	0·776 [0·763, 0·789]
121	78·7 [77·6, 79·8]	72·7 [72·4, 73·1]	21·3 [20·2, 22·4]	76·5 [75·9, 77·1]	0·833 [0·827, 0·839]	97·6 [96·8, 98·2]	18·6 [17·4, 19·9]	2·4 [1·8, 3·2]	57·5 [56·2, 58·9]	0·778 [0·766, 0·790]
124	71·8 [70·6, 73·0]	81·9 [81·6, 82·2]	28·2 [27·0, 29·4]	70·3 [69·5, 71·1]	0·834 [0·827, 0·841]	97·8 [97·0, 98·3]	16·3 [15·1, 17·5]	2·2 [1·7, 3·0]	58·2 [56·8, 59·5]	0·775 [0·763, 0·787]
125	77·6 [76·5, 78·7]	73·6 [73·2, 74·0]	22·4 [21·3, 23·5]	76·2 [75·5, 76·8]	0·836 [0·829, 0·842]	97·6 [96·8, 98·2]	18·6 [17·4, 19·9]	2·4 [1·8, 3·2]	57·5 [56·2, 58·9]	0·778 [0·766, 0·790]
130	73·8 [72·6, 74·9]	73·6 [73·2, 74·0]	26·2 [25·1, 27·4]	77·1 [76·5, 77·7]	0·804 [0·797, 0·811]	70·5 [68·6, 72·4]	73·7 [72·2, 75·1]	29·5 [27·6, 31·4]	37·7 [35·9, 39·6]	0·749 [0·736, 0·762]
131	71·3 [70·1, 72·5]	77·7 [77·4, 78·1]	28·7 [27·5, 29·9]	74·6 [73·9, 75·3]	0·819 [0·812, 0·826]	71·2 [69·3, 73·0]	74·2 [72·7, 75·6]	28·8 [27·0, 30·7]	37·1 [35·2, 39·0]	0·771 [0·758, 0·783]
141	71·0 [69·8, 72·2]	77·9 [77·5, 78·2]	29·0 [27·8, 30·2]	74·6 [73·9, 75·2]	0·819 [0·813, 0·826]	71·3 [69·4, 73·2]	74·0 [72·6, 75·4]	28·7 [26·8, 30·6]	37·2 [35·3, 39·1]	0·768 [0·755, 0·781]
154	71·3 [70·1, 72·5]	78·0 [77·7, 78·4]	28·7 [27·5, 29·9]	74·4 [73·7, 75·0]	0·829 [0·823, 0·835]	96·7 [95·9, 97·4]	17·1 [15·9, 18·4]	3·3 [2·6, 4·1]	58·2 [56·8, 59·5]	0·702 [0·688, 0·715]
165	68·7 [67·4, 69·9]	81·7 [81·3, 82·0]	31·3 [30·1, 32·6]	71·5 [70·7, 72·3]	0·826 [0·820, 0·833]	72·0 [70·1, 73·9]	74·0 [72·6, 75·4]	28·0 [26·1, 29·9]	36·9 [35·1, 38·8]	0·784 [0·772, 0·797]
166	71·9 [70·7, 73·1]	76·2 [75·9, 76·6]	28·1 [26·9, 29·3]	75·7 [75·0, 76·3]	0·822 [0·815, 0·828]	71·3 [69·4, 73·2]	74·0 [72·5, 75·4]	28·7 [26·8, 30·6]	37·2 [35·3, 39·1]	0·772 [0·758, 0·785]
250	78·6 [77·5, 79·7]	73·0 [72·6, 73·4]	21·4 [20·3, 22·5]	76·4 [75·7, 77·0]	0·842 [0·835, 0·848]	97·6 [96·8, 98·2]	18·6 [17·4, 19·9]	2·4 [1·8, 3·2]	57·5 [56·2, 58·9]	0·779 [0·767, 0·790]
251	73·5 [72·3, 74·6]	80·1 [79·8, 80·5]	26·5 [25·4, 27·7]	71·8 [71·0, 72·5]	0·845 [0·839, 0·851]	97·5 [96·8, 98·1]	19·0 [17·7, 20·3]	2·5 [1·9, 3·2]	57·4 [56·1, 58·8]	0·788 [0·776, 0·801]
270	69·4 [68·1, 70·6]	81·6 [81·3, 82·0]	30·6 [29·4, 31·9]	71·3 [70·6, 72·1]	0·829 [0·822, 0·835]	75·4 [73·6, 77·2]	65·5 [63·9, 67·0]	24·6 [22·8, 26·4]	42·6 [40·8, 44·4]	0·782 [0·770, 0·794]
271	70·5 [69·3, 71·7]	78·3 [78·0, 78·7]	29·5 [28·3, 30·7]	74·3 [73·6, 75·0]	0·824 [0·817, 0·830]	70·8 [68·9, 72·7]	74·8 [73·4, 76·2]	29·2 [27·3, 31·1]	36·6 [34·7, 38·5]	0·771 [0·758, 0·784]
289	72·1 [70·9, 73·3]	82·6 [82·2, 82·9]	27·9 [26·7, 29·1]	69·4 [68·6, 70·2]	0·850 [0·844, 0·856]	97·4 [96·6, 98·0]	18·2 [17·0, 19·5]	2·6 [2·0, 3·4]	57·7 [56·3, 59·0]	0·779 [0·767, 0·791]
311	69·1 [67·8, 70·3]	81·5 [81·1, 81·8]	30·9 [29·7, 32·2]	71·6 [70·8, 72·4]	0·831 [0·825, 0·838]	75·3 [73·5, 77·1]	65·8 [64·3, 67·4]	24·7 [22·9, 26·5]	42·4 [40·6, 44·2]	0·781 [0·769, 0·794]
380	71·3 [70·1, 72·5]	83·9 [83·6, 84·2]	28·7 [27·5, 29·9]	68·0 [67·1, 68·8]	0·851 [0·845, 0·857]	97·4 [96·6, 98·0]	18·4 [17·1, 19·7]	2·6 [2·0, 3·4]	57·6 [56·3, 59·0]	0·793 [0·782, 0·805]
392	71·4 [70·2, 72·6]	84·5 [84·2, 84·8]	28·6 [27·4, 29·8]	67·1 [66·2, 67·9]	0·854 [0·848, 0·860]	97·9 [97·2, 98·4]	16·3 [15·1, 17·5]	2·1 [1·6, 2·8]	58·1 [56·8, 59·5]	0·798 [0·787, 0·810]
402	69·6 [68·4, 70·8]	81·6 [81·3, 81·9]	30·4 [29·2, 31·6]	71·3 [70·5, 72·0]	0·833 [0·827, 0·840]	74·2 [72·4, 76·0]	68·4 [66·9, 69·9]	25·8 [24·0, 27·6]	40·9 [39·0, 42·7]	0·781 [0·769, 0·794]
417	72·0 [70·8, 73·2]	83·7 [83·4, 84·0]	28·0 [26·8, 29·2]	68·0 [67·2, 68·8]	0·856 [0·850, 0·862]	97·9 [97·2, 98·4]	16·9 [15·7, 18·2]	2·1 [1·6, 2·8]	58·0 [56·6, 59·3]	0·799 [0·787, 0·811]

Ledger: Values shown are percentages (except AUC), accompanied by 95% confidence intervals· rpart=Decision tree, XGB=eXtreme Gradient Boosting

Supplementary Table 4C: Performance characteristics by age subgroup of models shortlisted as tool candidates using the internal validation (TARN) dataset

Task ID	16-64 years										65+ years									
	Sensitivity		Specificity		Under-triage		Over-triage		AUC		Sensitivity		Specificity		Under-triage		Over-triage		AUC	
RPART																				
1	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·783	[0·775, 0·791]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·738	[0·726, 0·751]
3	75·3	[73·9, 76·6]	72·4	[71·8, 73·0]	24·7	[23·4, 26·1]	66·6	[65·5, 67·5]	0·787	[0·778, 0·795]	65·7	[63·4, 67·9]	75·8	[75·4, 76·3]	34·3	[32·1, 36·6]	87·1	[86·4, 87·8]	0·741	[0·728, 0·754]
37	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·794	[0·786, 0·803]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·746	[0·733, 0·759]
52	75·3	[73·9, 76·6]	72·4	[71·8, 73·0]	24·7	[23·4, 26·1]	66·6	[65·5, 67·5]	0·792	[0·784, 0·800]	65·7	[63·4, 67·9]	75·8	[75·4, 76·3]	34·3	[32·1, 36·6]	87·1	[86·4, 87·8]	0·745	[0·732, 0·758]
124	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·788	[0·780, 0·796]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·742	[0·729, 0·754]
XGB																				
1	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·795	[0·787, 0·804]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·747	[0·734, 0·759]
3	75·3	[73·9, 76·6]	72·4	[71·8, 73·0]	24·7	[23·4, 26·1]	66·6	[65·5, 67·5]	0·809	[0·801, 0·817]	65·7	[63·4, 67·9]	75·8	[75·4, 76·3]	34·3	[32·1, 36·6]	87·1	[86·4, 87·8]	0·763	[0·750, 0·776]
37	82·6	[81·4, 83·8]	70·1	[69·4, 70·7]	17·4	[16·2, 18·6]	66·3	[65·3, 67·2]	0·839	[0·831, 0·847]	67·3	[65·0, 69·5]	75·1	[74·6, 75·6]	32·7	[30·5, 35·0]	87·1	[86·4, 87·8]	0·752	[0·739, 0·765]
38	76·8	[75·4, 78·1]	71·5	[70·9, 72·2]	23·2	[21·9, 24·6]	66·8	[65·8, 67·7]	0·811	[0·803, 0·819]	67·2	[64·9, 69·4]	74·9	[74·4, 75·4]	32·8	[30·6, 35·1]	87·3	[86·6, 88·0]	0·765	[0·752, 0·778]
41	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·821	[0·813, 0·829]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·766	[0·752, 0·779]
42	76·0	[74·6, 77·3]	71·8	[71·2, 72·4]	24·0	[22·7, 25·4]	66·8	[65·8, 67·8]	0·806	[0·798, 0·815]	66·3	[64·0, 68·6]	75·3	[74·8, 75·8]	33·7	[31·4, 36·0]	87·2	[86·5, 87·9]	0·778	[0·766, 0·791]
43	74·4	[73·0, 75·7]	74·0	[73·4, 74·6]	25·6	[24·3, 27·0]	65·4	[64·4, 66·4]	0·815	[0·807, 0·824]	60·0	[57·7, 62·4]	81·5	[81·1, 81·9]	40·0	[37·6, 42·3]	85·0	[84·1, 85·8]	0·760	[0·746, 0·774]
44	82·0	[80·8, 83·2]	70·6	[70·0, 71·2]	18·0	[16·8, 19·2]	66·0	[65·0, 67·0]	0·852	[0·845, 0·859]	66·7	[64·4, 68·9]	75·7	[75·2, 76·1]	33·3	[31·1, 35·6]	87·0	[86·3, 87·7]	0·769	[0·756, 0·781]
53	75·3	[73·9, 76·6]	72·4	[71·8, 73·0]	24·7	[23·4, 26·1]	66·6	[65·5, 67·5]	0·813	[0·805, 0·821]	65·7	[63·4, 67·9]	75·8	[75·4, 76·3]	34·3	[32·1, 36·6]	87·1	[86·4, 87·8]	0·780	[0·767, 0·792]
54	74·9	[73·5, 76·3]	73·6	[73·0, 74·2]	25·1	[23·7, 26·5]	65·6	[64·6, 66·7]	0·825	[0·817, 0·833]	61·2	[58·9, 63·5]	80·5	[80·1, 80·9]	38·8	[36·5, 41·1]	85·4	[84·5, 86·2]	0·773	[0·760, 0·786]
121	83·4	[82·2, 84·5]	69·8	[69·2, 70·4]	16·6	[15·5, 17·8]	66·2	[65·3, 67·2]	0·855	[0·848, 0·862]	68·1	[65·9, 70·3]	74·7	[74·2, 75·2]	31·9	[29·7, 34·1]	87·2	[86·5, 87·9]	0·770	[0·758, 0·783]
124	77·5	[76·1, 78·8]	78·8	[78·2, 79·3]	22·5	[21·2, 23·9]	59·7	[58·6, 60·8]	0·854	[0·847, 0·862]	59·1	[56·7, 61·4]	84·0	[83·6, 84·4]	40·9	[38·6, 43·3]	83·3	[82·3, 84·2]	0·772	[0·758, 0·785]
125	82·3	[81·1, 83·5]	70·5	[69·9, 71·2]	17·7	[16·5, 18·9]	66·0	[65·0, 66·9]	0·854	[0·846, 0·861]	67·2	[64·9, 69·4]	75·7	[75·2, 76·1]	32·8	[30·6, 35·1]	86·9	[86·2, 87·6]	0·784	[0·772, 0·797]
130	76·6	[75·3, 77·9]	71·6	[71·0, 72·2]	23·4	[22·1, 24·7]	66·7	[65·7, 67·7]	0·814	[0·806, 0·822]	67·3	[65·0, 69·5]	74·9	[74·4, 75·4]	32·7	[30·5, 35·0]	87·2	[86·5, 87·9]	0·783	[0·771, 0·795]
131	75·6	[74·2, 76·9]	73·7	[73·1, 74·3]	24·4	[23·1, 25·8]	65·4	[64·3, 66·4]	0·828	[0·820, 0·836]	61·7	[59·4, 64·0]	80·4	[80·0, 80·8]	38·3	[36·0, 40·6]	85·3	[84·5, 86·1]	0·775	[0·761, 0·788]
141	74·8	[73·4, 76·1]	73·7	[73·1, 74·3]	25·2	[23·9, 26·6]	65·6	[64·6, 66·6]	0·824	[0·816, 0·833]	62·4	[60·0, 64·7]	80·6	[80·2, 81·1]	37·6	[35·3, 40·0]	85·1	[84·2, 85·9]	0·788	[0·775, 0·800]
154	75·5	[74·1, 76·9]	77·1	[76·5, 77·7]	24·5	[23·1, 25·9]	62·2	[61·1, 63·3]	0·847	[0·839, 0·854]	61·8	[59·4, 64·1]	78·6	[78·2, 79·1]	38·2	[35·9, 40·6]	86·4	[85·6, 87·1]	0·780	[0·768, 0·793]
165	72·9	[71·5, 74·3]	78·3	[77·8, 78·9]	27·1	[25·7, 28·5]	61·7	[60·6, 62·8]	0·838	[0·830, 0·846]	59·2	[56·8, 61·5]	83·9	[83·5, 84·3]	40·8	[38·5, 43·2]	83·3	[82·4, 84·2]	0·781	[0·768, 0·794]
166	75·9	[74·5, 77·2]	72·2	[71·5, 72·8]	24·1	[22·8, 25·5]	66·5	[65·5, 67·5]	0·828	[0·820, 0·836]	62·8	[60·5, 65·1]	78·9	[78·5, 79·4]	37·2	[34·9, 39·5]	86·0	[85·2, 86·8]	0·790	[0·778, 0·803]
250	83·2	[82·0, 84·3]	70·1	[69·4, 70·7]	16·8	[15·7, 18·0]	66·1	[65·1, 67·1]	0·860	[0·853, 0·868]	68·3	[66·0, 70·4]	74·9	[74·4, 75·4]	31·7	[29·6, 34·0]	87·1	[86·4, 87·8]	0·789	[0·777, 0·801]
251	79·9	[78·6, 81·2]	75·2	[74·6, 75·8]	20·1	[18·8, 21·4]	62·7	[61·7, 63·8]	0·863	[0·856, 0·871]	58·9	[56·6, 61·3]	83·4	[83·0, 83·8]	41·1	[38·7, 43·4]	83·8	[82·8, 84·7]	0·783	[0·770, 0·796]
270	73·7	[72·3, 75·1]	78·2	[77·6, 78·7]	26·3	[24·9, 27·7]	61·6	[60·5, 62·7]	0·840	[0·833, 0·848]	59·6	[57·2, 61·9]	83·9	[83·5, 84·3]	40·4	[38·1, 42·8]	83·2	[82·2, 84·1]	0·783	[0·770, 0·796]
271	74·5	[73·1, 75·9]	74·6	[74·0, 75·2]	25·5	[24·1, 26·9]	64·9	[63·8, 65·9]	0·830	[0·822, 0·838]	61·6	[59·2, 63·9]	80·8	[80·3, 81·2]	38·4	[36·1, 40·8]	85·1	[84·3, 85·9]	0·791	[0·779, 0·804]
289	77·5	[76·1, 78·8]	80·2	[79·7, 80·7]	22·5	[21·2, 23·9]	58·1	[56·9, 59·2]	0·869	[0·862, 0·876]	60·2	[57·8, 62·5]	84·1	[83·7, 84·5]	39·8	[37·5, 42·2]	82·9	[81·9, 83·8]	0·797	[0·785, 0·810]
311	73·1	[71·6, 74·5]	78·1	[77·5, 78·6]	26·9	[25·5, 28·4]	61·9	[60·8, 63·0]	0·840	[0·832, 0·847]	60·1	[57·7, 62·4]	83·7	[83·3, 84·1]	39·9	[37·6, 42·3]	83·3	[82·3, 84·2]	0·797	[0·784, 0·809]
380	76·9	[75·5, 78·2]	81·6	[81·0, 82·1]	23·1	[21·8, 24·5]	56·5	[55·3, 57·7]	0·870	[0·863, 0·877]	58·7	[56·3, 61·0]	85·4	[85·1, 85·8]	41·3	[39·0, 43·7]	82·0	[80·9, 83·0]	0·798	[0·786, 0·810]
392	77·2	[75·9, 78·5]	80·8	[80·3, 81·3]	22·8	[21·5, 24·1]	57·4	[56·2, 58·6]	0·869	[0·862, 0·876]	58·3	[55·9, 60·6]	87·0	[86·6, 87·4]	41·7	[39·4, 44·1]	80·3	[79·2, 81·4]	0·801	[0·788, 0·813]
402	73·6	[72·2, 75·0]	78·2	[77·6, 78·7]	26·4	[25·0, 27·8]	61·6	[60·5, 62·8]	0·842	[0·834, 0·849]	60·6	[58·2, 62·9]	83·9	[83·5, 84·3]	39·4	[37·1, 41·8]	83·0	[82·0, 83·9]	0·797	[0·785, 0·810]
417	77·9	[76·5, 79·2]	80·2	[79·7, 80·8]	22·1	[20·8, 23·5]	57·9	[56·7, 59·0]	0·873	[0·866, 0·880]	58·8	[56·4, 61·1]	86·0	[85·6, 86·4]	41·2	[38·9, 43·6]	81·4	[80·3, 82·4]	0·803	[0·790, 0·815]

Ledger: Values shown are percentages (except AUC), accompanied by 95% confidence intervals· rpart=Decision tree, XGB=eXtreme Gradient Boosting

Chapter 3 Conclusions

Using tree-based machine learning methodology, a novel primary triage tool as well as a secondary triage tool that is applicable via a portable device, have been developed. Both tools outperform the BCD Triage Sieve amongst TARN patients, with reduced associated over-triage rates. The app-based secondary triage tool has withstood external validation amongst injured patients from the UK military's Joint Theatre Trauma Registry (JTTR); however, a paucity of pre-hospital data precluded external validation of Quick Triage amongst JTTR patients, necessitating further work.

CHAPTER 4: PENETRATING TRAUMA AND TERRORIST-RELATED MAJOR INCIDENTS

Introduction to Chapter 4

A review of the existing literature has revealed that in the UK, man-made major incidents comprising blast and penetrating trauma have occurred with increasing frequency in recent years and continue to constitute the greatest threat to national security. In Chapter 2, an objective and evidence-based system to define triage categories has been validated amongst adults and children from the UK national trauma registry (TARN). Existing international adult and paediatric MI triage tools have been evaluating using this proposed gold standard. This comparative analysis identified the BCD Triage Sieve (now adapted for civilian use at the NHS MITT) as the tool with the greatest sensitivity in predicting P1 status amongst both adults and children, surpassing the current UK adult (NARU Triage Sieve) and Paediatric (Paediatric Triage Tape) MI triage tools. However, this work assessed tool performance in the entire trauma registry population in which 97% of patients sustained blunt injury. It is not known, therefore, how these existing tools, or indeed the novel ML-derived Quick Triage, performs in patients with penetrating trauma. The aim of chapter 4 is to identify the best performing triage tool for use in the most prevalent types of UK major incident, i.e. terrorist-related major incidents, which are characterised by blast and penetrating trauma.

Manuscript in draft: Triage in terror-related major incidents and the introduction of Quick Triage, a novel tool for UK major incidents

Author contributions:

NM, JV, AB, DB, DK, MF and GVG designed the study. **NM**, SC and YX verified the underlying data and conducted analysis. All authors contributed to data interpretation. **NM** wrote the initial draft of the manuscript. All authors contributed to critical revisions of subsequent manuscript drafts and approve of the final version.

Triage in terror-related major incidents and the introduction of Quick Triage, a novel tool for UK major incidents

Abstract:

Introduction:

With an ongoing threat from terror-related and other major incidents (MI), evidence-based selection of triage tools for the UK setting is imperative. Triage tools must be highly sensitive in identifying patients in need of time-critical, life-saving interventions (Priority 1, or P1 casualties), whilst seeking to minimise over-triage. We introduce Quick Triage, a novel machine learning-derived tool, and evaluate its performance in clinically important subgroups within the Trauma Audit and Research Network (TARN) registry. NHS England has recommended use of Ten Second Triage (TST) by non-NHS first responders and Major Incident Triage Tool (MITT) for NHS first responders from April 2024; these and other prominent tools are used as comparators.

Methods:

TARN patients of all ages (January 2008-December 2017) were included and divided into subgroups based on age (<16 years, 16-64 years and elders aged 65+ years) and injury mechanism (blunt or penetrating). Quick Triage, TST, MITT, the UK's current tool NARU Triage Sieve, Battlefield Casualty Drills Triage Sieve, CareFlight, Modified Simple Triage and Rapid Treatment (START), and the Modified Physiological Triage Tool (MPTT-24) were retrospectively applied to patients' first recorded pre-hospital physiology. True P1 patients were identified using pre-defined, intervention-based criteria. Tool-assigned and intervention-based assignment of P1 status were compared.

Results:

195,709 adults (including 100,403 elders) were included with median injury severity score (ISS) of 9 (IQR 9-17) and in-hospital mortality of 7.0% (n=13,601). 11.3% fulfilled criteria for P1 status. 97.1% (n=190,048) sustained blunt injuries, including all elders. Amongst 4,962 children, median ISS was 9

(9-17) with mortality of 1.1% (n=53); blunt injuries predominated (95.4%). 27.1% (1,343) of children fulfilled P1 criteria.

In blunt-injured adults, Quick Triage was most sensitive (78.8%; elders 77.0%), followed by MITT (74.0%; elders 57.9%). TST attained sensitivity of just 50.5% (elders 37.9%). TST, validated in real-life patients for the first time, demonstrated exceptionally high rates of under-triage in blunt-injured P1 adults (43.5%), children (46.7%) and elders (62.1%). MITT's sensitivity for blunt-injured children was 17% greater than Quick Triage, however Quick Triage outperformed MITT by a similar margin in children with penetrating injuries, with 30% reduction in over-triage. Across the subgroups, Quick Triage had lower over-triage (0.5%-30.5%) relative to MITT, which demonstrated the highest over-triage (46.7-88.3%) amongst the tools studied.

Conclusion:

Quick Triage offers the most balanced performance in predicting the need for life-saving interventions across adults and children with any injury mechanism and is associated with significantly lower over-triage rates than MITT: we recommend its use as a single tool for all first responders in UK MIs in place of TST and MITT. Use of TST may be associated with significant harm due to substantial under-triage of blunt-injured adults, children and elders, a weakness not recognised prior to this study.

INTRODUCTION:

In the wake of the 2017 Manchester Arena bombing, much consideration has been given to how the best possible outcomes can be delivered to victims of UK major incidents (MIs)[1]. A key component of pre-hospital care is casualty triage using appropriate pre-determined triage tools[2,3]. The triage process seeks to ensure that casualties with life-threatening injuries are treated at scene ahead of the less critically injured, and that casualties are transferred to the appropriate medical facility within regionalised trauma networks[2 4]. A Priority 1 (P1) casualty is defined as one in need of time-critical, life-saving intervention[5-7]. It is vital that triage tools correctly identify P1 patients, since failure to do so (under-triage) is associated with direct harm and increased mortality resulting from delayed care or transfer to an inappropriate medical facility[8-10]. Some over-triage, or assignment of P1 status to non-critical patients, is inevitable if tools are to remain simple, however this is associated with the potential harm of overwhelming healthcare facilities, thus impairing the care of individual patients[10]. Whilst major incident tools must perform robustly across a range of injury mechanisms, the greatest loss of life on UK soil since World War 2 resulted from terrorist attacks during the London 7/7 bombings and 2017 Manchester Arena Bombing, making the validation of any potential triage tools amongst patients with blast and penetrating injuries a high priority[1,11-13]. Indeed, the 2023 National Risk Register highlights that the likeliest threat continues to be terrorist attacks, including those on public places and transport[14]. Whilst blast and penetrating mechanisms have dominated, recent terrorist attacks have included combinations of mechanisms, such as vehicle drive-through (i.e. blunt trauma) followed by dismounted knife attacks, as seen during the Westminster and London Bridge Attacks, or blast associated with building collapse[1,11,12,15]. Sadly, the Manchester Arena bombing illustrated the vulnerability of children to major incidents (17% of the UK population is aged <15 years) and planning must also take into account that 19.7% of UK adults are aged over 65[1,16]. With public expectations at their highest, coupled with a higher level of scrutiny than ever before, there is a need for robust evidence to inform national policymakers in their choice of triage tools[1].

UK major incident triage has traditionally involved a two-stage "Sieve and Sort" approach with a primary triage "sieve" applied at the scene of injury followed by a more detailed assessment (secondary triage) conducted within the relative safety of a Casualty Clearing Station, prior to onward transfer of patients[2,3,17]. This approach assumes that the secondary triage process offers additional triage accuracy, justifying the added time and complexity this entails. However, evidence suggests that the existing UK secondary triage tool, the Triage Sort, is poorly sensitive (15%) in detecting the need for life-saving intervention[2,18]. This, coupled with its complexity, may explain why it has fallen out of use. Until recently, it has been UK policy to use the NARU Triage Sieve as an adult (12+ years) primary tool followed by the Triage Sort, with use of the Paediatric Triage Tape (PTT) at scene in children under 12 years, followed by JumpSTART at hospital reception[2,17]. Substantial evidence has emerged that these four tools have suboptimal performance in identifying casualties needing life-saving interventions, as well as other important clinical outcomes[5,7,18,19]. Additionally, use of four tools during a single major incident is unnecessarily complex, adding to the cognitive burden of first responders, who must maintain situational awareness to ensure their safety and that of their patients[3,11,20,21]. Simplification of the triage process may minimise delays and improve triage accuracy. As such, NHS England has proposed that by June 2024, non-NHS first responders should use the ultra-simple, expert-derived Ten Second Triage (TST, Figure 1), alongside use of the NHS Major Incident Triage Tool (MITT, Figure 2) by NHS first responders, in patients of all ages[22,23]. This empowers non-NHS first responders to begin the triage process and administer life-saving interventions ahead of the arrival of NHS first responders, addressing the so-called "care gap" highlighted in the Manchester Arena Inquiry[1]. However there is no published evidence describing how effective TST is in discriminating P1 from non-P1 patients outside of a simulated setting[22]. MITT represents the civilian adaptation of the UK Military's BattleField Casualty Drills (BCD) Triage Sieve, following evidence that this has a relatively high sensitivity (56.7-75.7%) in predicting the need for life-saving interventions amongst adults, elders and children, as well as predicting in-patient mortality and injury severity amongst casualties of all ages[7,19,23]. Additionally, there is indirect evidence of its efficacy from the excellent trauma outcomes achieved when the BCD Triage Sieve has been used by soldiers during military operations in Afghanistan[24].

Quick Triage is a novel machine learning-derived major incident triage tool that was modelled to predict Priority 1 status within a UK-specific population of injured adults (Figure 3)[25]. Uniquely for the major incident context, the modelling process incorporated assessment of injury mechanism as well as anatomical assessment of injuries[25]. The resulting decision tree model included assessment of the ability to breathe spontaneously, followed by ability to obey commands (GCS Motor Score=6) and presence of a chest injury. Additionally, presence of a torso, junctional or head and neck penetrating injury was found to be strongly associated with P1 status[25]. In adapting this model into a practically applicable tool, assessment for and interventions to control catastrophic haemorrhage have been included: haemorrhage is a leading cause of death following injury and control of peripheral or junctional haemorrhage may be readily achieved using tourniquets and haemostatic dressings[26,27]. To prevent early deaths, two further simple interventions have been included: opening of the airway and placement of patients with airway compromise or altered mental status in the recovery position[4]. Additionally, in keeping with conventional tools, ability to walk has been included as a marker of non-severe injury, i.e. Priority 3 status[2,4]. Quick Triage has yet to be employed in practice and its performance in clinically important subgroups of the population has yet to be demonstrated. The aim of this study was to measure the performance of Quick Triage amongst paediatric, adult and elderly patients with blunt and penetrating injury from the UK national Trauma, Audit and Research Network (TARN) registry, in comparison with existing tools, to inform national policymakers.

Figure 1: Ten Second Triage (TST)

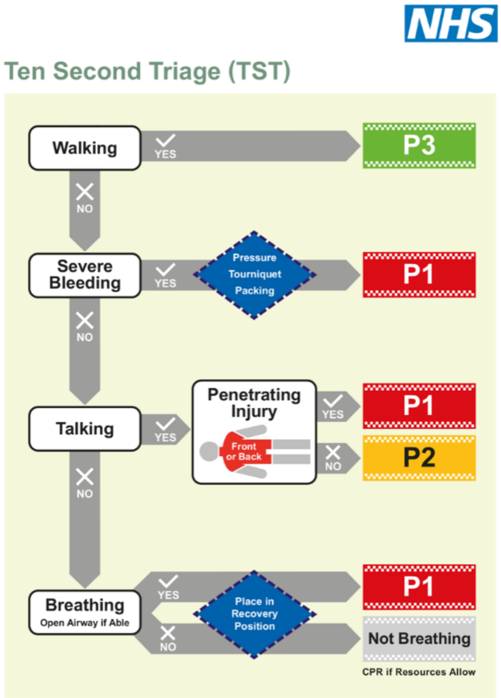


Figure 2: NHS Major Incident Triage Tool (MITT)

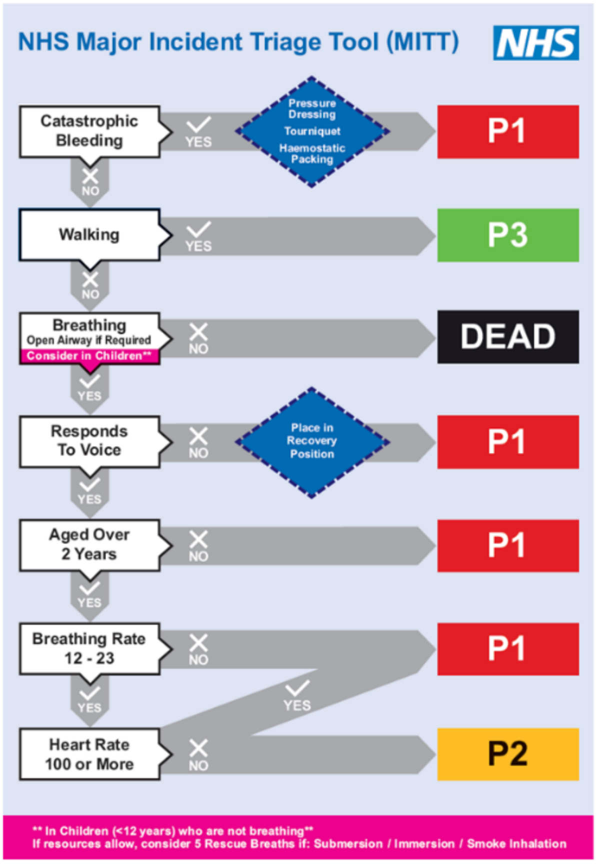
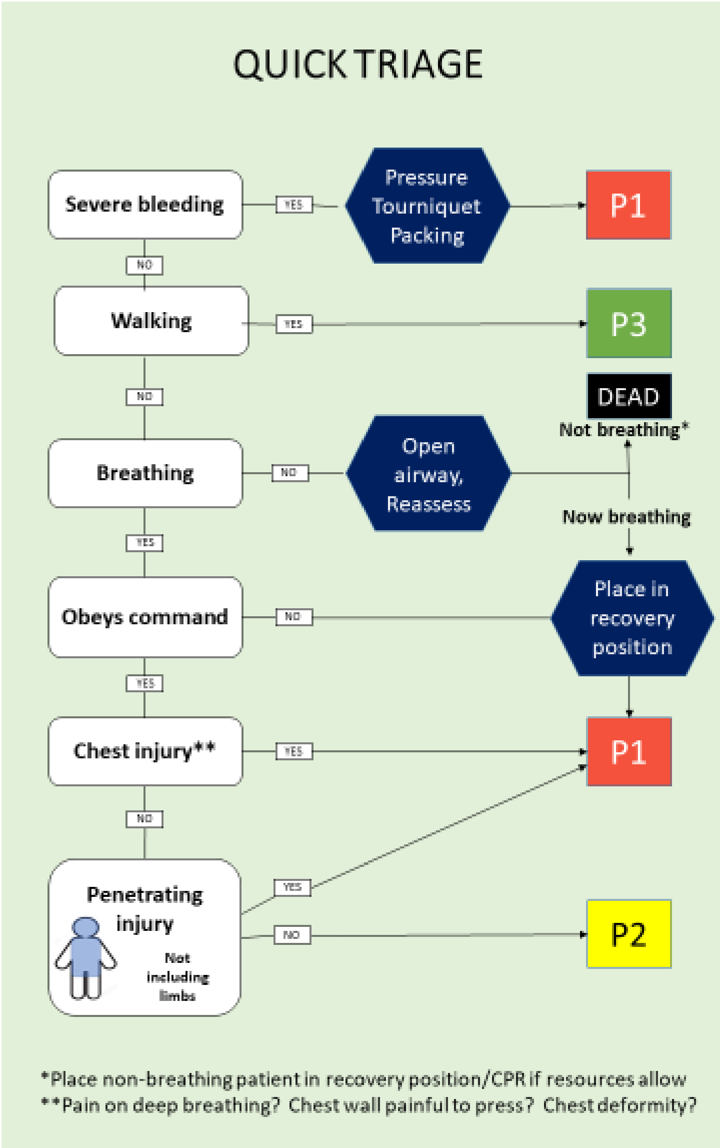


Figure 3: Quick Triage



METHODS:

Overview of study design

This study measures the performance of eight major incident triage tools in predicting P1 status using patients from the UK Trauma Audit and Research Network (TARN) registry as surrogates for those injured in a major incident. Each triage tool was applied to patients' pre-hospital physiology to determine whether the patient would have been designated P1 status. Patient records were reviewed to determine whether they required time-critical, life-saving interventions from a predefined list, allowing assignment of "actual" triage categories (P1, P2, P3, Expectant or Dead). Tool performance was assessed by comparing tool-predicted and intervention-based P1 status.

Study population

Prospectively recorded, anonymised data for TARN registry patients of all ages presenting to hospitals in England and Wales between 1 January 2008 and 31 December 2017 were included. Patients with incomplete pre-hospital physiological data required to apply the triage tools (respiratory rate, heart rate, capillary refill time, Glasgow Coma Score (GCS), and GCS Motor Component) were excluded. TARN constitutes the largest trauma registry in Europe, receiving data from all UK Major Trauma Centres and Trauma Units[28]. The registry includes injured patients fulfilling the following criteria: length of stay over 72 hours, intensive care (ICU) admission and/or in-hospital death. TARN does not include prehospital deaths19. Trained TARN co-ordinators record pre-hospital and in-hospital physiological, demographic, and outcome data contemporaneously into a web-based proforma.

Application of triage tools

To test tool performance, patients were categorised by age into children (<16 years), adults (16-64 years) and elders (65+ years), consistent with National Health Service configuration, and further divided into subgroups based on the presence of blunt and penetrating injuries. Quick Triage (Figure 1), TST (Figure 2), NHS MITT (Figure 3), its parent tool, the UK military's BCD Triage Sieve[7], and the current UK tool, the NARU Triage Sieve2, were transcribed into computer code and applied to first recorded pre-hospital physiology to determine whether patients were assigned P1 or non-P1. As comparators, CareFlight, used widely in Australia4 and MSTART4, used in several parts of North America were applied, as well as the MPTT-24[29], which was modelled in UK military casualties but has yet to be used in practice. Several assumptions were made to facilitate retrospective tool application, as described previously[7]. Since TARN does not record the pre-hospital suspected anatomical distribution of injuries, in the application of Quick Triage "presence of a chest injury" was assumed if the patient had a retrospective assignment of Abbreviated Injury Score greater than 0 (AIS Chest>0)[25,28]. For QT and TST, presence of a central penetrating injury was coded as "Injury type: penetrating" and "AIS>0" in any body region other than limbs. For TST, "Talking?" was denoted as a GCS Verbal score greater than or equal to 3[22].

Outcome measures:

The primary outcome measure was the ability of triage tools to predict P1 status, defined as the need for time-critical lifesaving intervention(s). Each patient was assigned a triage category (Dead, Expectant, P1, P2 or P3) based on a pre-defined system utilising Emergency Medical Services and hospital-based interventions described by Lerner et al, using equivalent TARN terminology (see Supplementary data table 1)[6]. This system was originally derived using literature review followed by expert consensus, and has been validated amongst UK trauma registry patients, demonstrating clinically meaningful and statistically significant differences in mortality, intensive care use, length of stay and injury severity between the triage categories[7]. TARN records the timing of hospital arrival and each intervention, allowing incorporation of this interval into the time-critical definitions constituting P1 status.

To further evaluate tool strengths and weaknesses, secondary outcome measures included prediction of in-hospital mortality and ISS>15. Although it does not directly correlate with resource requirements, which is of prime importance in the MI setting, ISS>15 is commonly used to define major trauma, to justify major trauma centre admission within UK regionalised inclusive trauma systems, and to assess the validity of field triage tools outside of the major incident context[9,28,30].

Data processing and analyses:

TARN data were received in SPSS Version 24.0 (Armonk NY: IBM Corp 2015) and processed using R software (Version 3.6, R Core Team, New Zealand, 2000). Non-parametric data are presented as median and interquartile range; categorical data as frequency and percent. Performance characteristics included sensitivity, specificity, under-triage (1-sensitivity), over-triage (1-positive predictive value), and Area Under the Receiver Operating Curve (AUROC). 95% confidence intervals were calculated using the Wilson Score with continuity correction for binomial proportions, and DeLongs Algorithm for comparing AUC curves. Paediatric patients excluded due to missing data were compared to included patients (see Supplementary data table 3 for statistical tests and results).

Ethical approval: The UK Health Research Authority Patient Information Advisory Group (Section 20) has granted ethical approval for research using anonymised TARN data.

Role of funding source: The funding source had no role to play in study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

RESULTS:

Characteristics of the study population and subgroups

TARN captured 215,632 adult patients. 19,923 (9.2%) patients were excluded due to incomplete physiological data, hence, 195,709 patients were included. Adult patient and injury characteristics are summarised in Table 1. Median injury severity was 9 (IQR 9-17) and in-hospital mortality was 7.0% (n=13,601). Approximately half of adults were aged 65+ years (n=100,403). The vast majority (97.1%) of adults sustained blunt injuries, including all those aged 65+ years. Low-level falls constituted the commonest mechanism (n=113,319, 57.9%) followed by vehicle collisions (n=41,590, 21.3%) and falls greater than 2m (n=25,194, 12.9%).

Table 1: Patient and injury characteristics (Adults aged 16+ years, n=195,709)

Characteristic	n (%)
Gender	
Male	104,019 (53.1%)
Female	91,690 (46.9%)
Missing data	0 (0.0%)
Injury Severity Score (ISS)	
Median (IQR)	9 (9–17)
ISS>15	62,402 (31.9%)
Missing data	0 (0.0%)
Age	
Median (IQR), years	66.2 (47.3–83.0)
Patients aged 16–64 years	95,306 (48.7%)
Patients aged 65+ years	100,403 (51.3%)
Missing data	0 (0.0%)
Discharge status	
Alive	182,107 (93.0%)
Dead	13,601 (7.0%)
Missing data	1 (0.0%)
Mode of injury	
Blunt	190,048 (97.1%)
Penetrating	5660 (2.9%)
Missing data	1 (0.0%)
Mechanism of injury	
Fall less than 2m	113,319 (57.9%)
Vehicle Incident/Collision	41,590 (21.3%)
Fall more than 2m	25,194 (12.9%)
Blow(s)	6827 (3.5%)
Stabbing	4105 (2.1%)
Other	2609 (1.3%)
Crush	1355 (0.7%)
Shooting	440 (0.2%)
Blast	142 (0.1%)
Burn	128 (0.1%)
Missing data	0 (0.0%)
Triage category	
Priority 1	22,144 (11.3%)
Priority 2	171,404 (87.6%)
Dead	282 (0.1%)
Expectant	1879 (1.0%)

2.9% (n=5660) of adults sustained penetrating injuries with a median age of 32.6 years (IQR 23-48) and a strong male preponderance (85.6%, n=4,845) (See Supplementary data table 2 for full results). In-hospital mortality was relatively low (3.1%, n=171) and median ISS was 9 (IQR 9-16). Penetrating injuries were most frequently attributable to stabbing (n=4105, 2.1%) or other penetrating injuries (n=1,115). A minority sustaining gunshot wounds (n=440, 0.2%).

Over the 10 year period, 15,133 children were identified; 10,171 (67.2%) were excluded due to incomplete pre-hospital physiological data19. Hence 4,962 children were included in the study. Included and excluded patients were compared (see Supplementary data table Table 4): excluded patients had comparable injury severity (median ISS 9 [IQR9-16] vs. median ISS 9 [IQR9-17], p<0.01 respectively) but higher mortality (3.1% vs. 1.1%, p<0.01) compared to included patients. Excluded patients were more likely to have sustained burns and falls below 2m.

Paediatric patient and injury characteristics are summarised in Table 2. Male patients constituted two thirds of the cohort (n=3447). Nearly half of the paediatric cohort were aged 12+ years, one quarter were aged 8 to 11 years (n=1281, 25.8%) and the remaining quarter were aged under 8 years. As seen in adults, median ISS was 9 (IQR 9-17). In-hospital mortality was low at 1.1% (n=53). Blunt injuries predominated (95.4%), with nearly half of all children injured in vehicle collisions (n=2459) and over a third from falls (n=1832). 4.6% (n=229) of children sustained penetrating injuries, most commonly due to stabbings (n=130, 2.6%), other penetrating injuries (n=83, 1.7%) and a very small minority (n=16, 0.3%) sustained gunshot wounds.

Table 2: Paediatric patient and injury characteristics (age <16 years, n=4,962)

Characteristic	n (%)
Gender	
Male	3447 (69.5%)
Female	1515 (30.5%)
Missing data	0 (0.0%)
Injury Severity Score (ISS)	
Median (IQR)	9 (9, 17)
ISS>15	1,617 (32.6%)
Missing data	0 (0.0%)
Age	
Median (IQR), years	11.9 (8.0, 14.2)
<16 years, n (%)	4962 (100.0%)
Of which, 12 to <16	2446 (49.3%)
8 to <12	1281 (25.8%)
4 to <8	768 (15.5%)
0 to <4	467 (9.4%)
Missing data	0 (0.0%)
Discharge status	
Alive	4909 (98.9%)
Dead	53 (1.1%)
Missing data	1 (0.0%)
Mode of injury	
Blunt	4733 (95.4%)
Penetrating	229 (4.6%)
Missing data	1 (0.0%)
Mechanism of injury	
Vehicle Incident/Collision	2459 (49.6%)
Fall less than 2m	1187 (23.9%)
Fall more than 2m	645 (13.0%)
Blow(s)	327 (6.6%)
Other	147 (3.0%)
Stabbing	130 (2.6%)
Crush	42 (0.9%)
Shooting	16 (0.3%)
Blast	5 (0.1%)
Burn	4 (0.1%)
Missing data	0 (0.0%)
Triage category	
Priority 1	1,343 (27.1%)
Priority 2	3,588 (72.3%)
Dead	2 (0.04%)
Expectant	29 (0.58%)

Designation of intervention-based triage categories

Amongst the adult (16+ years) population, 11.3% (n=22,144) of patients fulfilled criteria for P1 status. The majority of these required advanced airway intervention (n=18,890, 85.3%), and one fifth needed a chest drain within 2 hours (n=4,301). 15.5% (n=3,427) were designated P1 status based on requiring emergency lifesaving surgery. A list of the eight criteria constituting P1 status, with a breakdown of patient designations is presented in Supplementary data table 1. A minority of patients fulfilled criteria for the "Dead" category (n=282, 0.1%), with a further 1% (n=1,879) classed as "Expectant" based on the presence of non-survivable burns or catastrophic head injuries6. The vast majority (87%, n=171,402) of patients were designated P2 status, forming the largest triage category. In this TARN-eligible study population, no patients met criteria for Priority 3 (minimally injured) status28.

In the subgroup of adults with penetrating injuries, and in contrast to their blunt-injured counterparts, nearly half of patients fulfilled criteria for P1 status (n=2,639), the largest portion of whom needed life-saving surgery (44.3%, n=1,170) (See Supplementary data table 4 for a breakdown of P1 designations). A third (n=864) of P1 patients required chest drain placement within 2 hours and there was a relatively lower requirement for airway interventions (37.0%, n=976). A similar proportion arrived in ED with uncontrolled haemorrhage (12.4% in penetrating subgroup compared with 8.9% amongst all adults). In-hospital mortality amongst adults with penetrating trauma was relatively low (3.1%, n=171).

In the paediatric study population, 1343 (27.1%) patients met criteria for P1 status: three quarters (n=808) required advanced airway intervention, nearly a third (n=310) arrived in ED with uncontrolled haemorrhage and 15.7% (n=169) required life-saving surgery within 4 hours (see Supplementary data table 5 for further details). Less than 1% (n=31) of paediatric patients fulfilled criteria for "Dead" and "Expectant" categories and the vast majority (n=3588, 72·3%) of patients were designated P2. No patients met criteria for the P3 (minor) category.

Triage tool performance

(i) Prediction of P1 status (need for life-saving interventions)

Amongst adults aged 16-64 years with blunt trauma (Table 3), Quick Triage had the highest sensitivity (78.8% 95CI 78.1-79.5) and the greatest overall discriminatory capability reflected by AUC of 0.758 (95CI 0.754-0.762). Quick Triage offered a 6.3% sensitivity advantage over the next best performer, MITT, with an associated improvement in over-triage at 66.4% (95CI 65.9-66.9) compared with MITT's 74.0% (95CI 73.5-74.4). Both Quick Triage and MITT offer an improvement in sensitivity of 25% or greater compared with the currently utilised NARU Triage Sieve (Sensitivity 46.8% 95CI 46.0-47.7), although at the expense of slightly higher (8.5 to 16.1%) over-triage rates. TST attained sensitivity of 50.5%, under-triaging nearly half of all blunt injured adults needing life-saving interventions. CareFlight, NARU Triage Sieve and TST all demonstrated very high specificity (88.8-95.8%) but all three tools exhibited under-triage rates of over 50%. The highest over-triage rate was demonstrated by MPTT-24 (82.1% 95CI 81.7-82.5), which also had the lowest overall discriminatory capability (AUC 0.546 95CI 0.541-0.550).

Table 3: Tool performance in predicting the need for life-saving intervention (Priority 1 status) in adult patients (16-64 years) with blunt trauma (n=190,048)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	78.8	[78.1, 79.5]	72.8	[72.5, 73.1]	21.2	[20.5, 21.9]	66.4	[65.9, 66.9]	0.758	[0.754, 0.762]
Tools proposed by NHS England										
TST	50.5	[49.6, 51.3]	95.8	[95.7, 96.0]	49.5	[48.7, 50.4]	32.0	[31.1, 32.9]	0.732	[0.727, 0.736]
NHS MITT	72.5	[71.7, 73.2]	64.0	[63.7, 64.4]	27.5	[26.8, 28.3]	74.0	[73.5, 74.4]	0.683	[0.679, 0.687]
Other tools										
BCD Triage Sieve	71.2	[70.4, 71.9]	66.3	[66.0, 66.6]	28.8	[28.1, 29.6]	73.0	[72.6, 73.5]	0.687	[0.683, 0.692]
CareFlight	47.2	[46.4, 48.1]	93.0	[92.8, 93.1]	52.8	[51.9, 53.6]	46.1	[45.2, 47.0]	0.701	[0.696, 0.705]
MPTT-24	45.6	[44.7, 46.4]	63.5	[63.2, 63.9]	54.4	[53.6, 55.3]	82.1	[81.7, 82.5]	0.546	[0.541, 0.550]
MSTART	61.3	[60.4, 62.1]	89.2	[89.0, 89.4]	38.7	[37.9, 39.6]	50.2	[49.4, 50.9]	0.752	[0.748, 0.757]
NARU Triage Sieve	46.8	[46.0, 47.7]	88.8	[88.5, 89.0]	53.2	[52.3, 54.0]	57.9	[57.1, 58.7]	0.678	[0.673, 0.682]

In adults aged 16-64 years with penetrating trauma (Table 4), very high sensitivities were exhibited by Quick Triage (90.3% 95CI 89.1-91.4) and TST (88.5% 95CI 87.2-89.8). These were followed by MITT, with a moderate sensitivity of 68.4% (95CI 66.5-70.2). All three aforementioned tools substantially outperformed NARU Triage Sieve's sensitivity of 34.5% (95CI 32.7-36.4). CareFlight had the lowest sensitivity (22.6% 95CI 21.0-24.3) but the highest specificity (89.8% 88.6-90.9). Over-triage rates across all tools in this penetrating subgroup of adults consistently fell below 50%, unlike those seen amongst blunt injured adults (range 32.0-82.1%).

Table 4: Tool performance in predicting the need for life-saving intervention (Priority 1 status) in adult patients (16-64 years) with penetrating trauma (n=5,660)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	90.3	[89.1, 91.4]	27.3	[25.6, 29.0]	9.7	[8.6, 10.9]	46.2	[44.7, 47.8]	0.588	[0.578, 0.598]
Tools proposed by NHS England										
TST	88.5	[87.2, 89.8]	32.4	[30.6, 34.2]	11.5	[10.2, 12.8]	44.9	[43.4, 46.5]	0.605	[0.594, 0.616]
NHS MITT	68.4	[66.5, 70.2]	43.8	[42.0, 45.8]	31.6	[29.8, 33.5]	46.7	[45.0, 48.5]	0.561	[0.548, 0.574]
Other tools										
BCD Triage Sieve	66.2	[64.3, 68.1]	46.8	[44.9, 48.7]	33.8	[31.9, 35.7]	46.2	[44.4, 48.0]	0.565	[0.552, 0.578]
CareFlight	22.6	[21.0, 24.3]	89.8	[88.6, 90.9]	77.4	[75.7, 79.0]	32.5	[29.3, 35.8]	0.562	[0.552, 0.572]
MPTT-24	60.5	[58.6, 62.4]	44.6	[42.7, 46.5]	39.5	[37.6, 41.4]	49.4	[47.6, 51.2]	0.526	[0.512, 0.539]
MSTART	35.6	[33.8, 37.6]	83.7	[82.2, 85.1]	64.4	[62.4, 66.2]	32.8	[30.3, 35.4]	0.597	[0.585, 0.608]
NARU Triage Sieve	34.5	[32.7, 36.4]	77.5	[75.9, 79.1]	65.5	[63.6, 67.3]	41.0	[38.5, 43.6]	0.560	[0.548, 0.572]

Note: all adults with penetrating injuries fell within the 16-64 year subgroup.

Amongst elders (65+ years, see Table 5), all of whom sustained blunt injuries, Quick Triage demonstrated the greatest sensitivity of 77.0% (95CI 76.4-77.5), surpassing all other tools by a margin of 20%, with a relatively modest over-triage rate of 73.1% (95CI 72.7-73.4) and best overall AUC of 0.752 (95CI 0.749-0.755). This was followed with a 20% margin by MITT (sensitivity 57.9% 95CI 56.6-59.1), which had higher over-triage at 88.3% (95CI 87.9-88.6). High over-triage rates were seen with all tools (62.8-91.3%) in the elders subgroup. CareFlight, MPTT-24, MSTART, NARU Triage Sieve and TST all demonstrated sensitivity of less than 50%, with TST identifying little over one-third (sensitivity 37.9% 95CI 36.9-39.1) of elders needing life-saving interventions.

Table 5: Tool performance in predicting the need for life-saving intervention (Priority 1 status) in elders (65+ years, n=100,403)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	77.0	[76.4, 77.5]	73.4	[73.2, 73.6]	23	[22.5, 23.6]	73.1	[72.7, 73.4]	0.752	[0.749, 0.755]
Tools proposed by NHS England										
TST	37.9	[36.7, 39.1]	95.8	[95.6, 95.9]	62.1	[60.9, 63.3]	62.8	[61.6, 64.0]	0.668	[0.662, 0.674]
NHS MITT	57.9	[56.6, 59.1]	71.1	[70.8, 71.4]	42.1	[40.9, 43.4]	88.3	[87.9, 88.6]	0.645	[0.639, 0.651]
Other tools										
BCD Triage Sieve	56.7	[55.5, 57.9]	72.7	[72.4, 73.0]	43.3	[42.1, 44.5]	87.9	[87.5, 88.3]	0.647	[0.641, 0.653]
CareFlight	33.5	[32.3, 34.7]	93.4	[93.3, 93.6]	66.5	[65.3, 67.7]	74.7	[73.7, 75.6]	0.635	[0.629, 0.641]
MPTT-24	43.1	[41.9, 44.3]	69.9	[69.6, 70.2]	56.9	[55.7, 58.1]	91.3	[91.0, 91.6]	0.565	[0.559, 0.571]
MSTART	48.6	[47.4, 49.9]	88.5	[88.3, 88.7]	51.4	[50.1, 52.6]	78.2	[77.5, 78.8]	0.686	[0.679, 0.692]
NARU Triage Sieve	33.2	[32.1, 34.4]	89.6	[89.4, 89.8]	66.8	[65.6, 67.9]	82.5	[81.8, 83.1]	0.614	[0.609, 0.620]

Note: there were no elders with penetrating trauma within the TARN database.

In children with blunt injuries (Table 6), MITT had the greatest sensitivity at 84.7% (95CI 82.3-86.8), however this was associated with a relatively high over-triage rate of 73.2% (95CI 71.6-74.7) and AUC of 0.612 (95CI 0.599-0.626). Quick Triage had the second highest sensitivity, identifying two thirds of children needing life-saving intervention (sensitivity 67.1% 95CI 64.1-69.9), with a 30% lower over-triage rate than MITT and the highest AUC of 0.768 (95CI 0.753-0.784). As seen in blunt-injured adults, TST identified approximately half of blunt-injured children requiring life-saving intervention (sensitivity 53.3% 95CI 48.7-55.0), with similar performance to CareFlight (sensitivity 51.9% 95CI 48.7-55.0); both TST and CareFlight demonstrated high specificity above 94%.

Table 6: Tool performance in predicting the need for time-critical, life-saving intervention (Priority 1 status) in children (<16 years) with blunt trauma (n=4,733)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	67.1	[64.1, 69.9]	86.6	[85.4, 87.6]	32.9	[30.1, 35.9]	42.7	[39.9, 45.6]	0.768	[0.753, 0.784]
Tools proposed by NHS England										
TST	53.3	[50.1, 56.4]	95.6	[94.9, 96.2]	46.7	[43.6, 49.9]	23.5	[20.5, 26.9]	0.744	[0.729, 0.760]
NHS MITT	84.7	[82.3, 86.8]	37.8	[36.2, 39.3]	15.3	[13.2, 17.7]	73.2	[71.6, 74.7]	0.612	[0.599, 0.626]
Other tools										
BCD Triage Sieve	82.6	[80.1, 84.9]	41.9	[40.3, 43.5]	17.4	[15.1, 19.9]	72.3	[70.7, 73.9]	0.623	[0.609, 0.637]
CareFlight	51.9	[48.7, 55.0]	94.1	[93.2, 94.8]	48.1	[45.0, 51.3]	29.9	[26.6, 33.3]	0.730	[0.714, 0.746]
MPTT-24	53.8	[50.7, 56.9]	39.6	[38.0, 41.2]	46.2	[43.1, 49.3]	80.7	[79.1, 82.1]	0.467	[0.450, 0.484]
MSTART	62.6	[59.6, 65.6]	87.1	[86.0, 88.1]	37.4	[34.4, 40.4]	43.4	[40.5, 46.4]	0.749	[0.733, 0.764]
NARU Triage Sieve	57.8	[54.7, 60.8]	78.3	[77.0, 79.6]	42.2	[39.2, 45.3]	58.2	[55.6, 60.8]	0.681	[0.664, 0.697]

Tool performance in children with penetrating trauma closely mirrored that seen in adults (Table 7), with very high sensitivities demonstrated by Quick Triage (93.8%, 95CI 86.4-97.4) and TST (88.5%, 95CI 80.0-93.3), followed with some distance by MITT at 75%, (95CI 64.9-83.0). As in adults, over-triage rates were much lower amongst children in penetrating trauma (44.4-54.9%) compared with blunt-injured children (23.3-80.7%), in which wider variation but generally higher over-triage rates were noted.

Table 7: Tool performance in predicting the need for life-saving intervention (Priority 1 status) in children (<16 years) with penetrating trauma (n=229)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	93.8	[86.4, 97.4]	35.0	[27.2, 43.7]	6.2	[2.6, 13.6]	49.7	[42.2, 57.2]	0.644	[0.597, 0.691]
Tools proposed by NHS England										
TST	88.5	[80.0, 93.9]	46.0	[37.5, 54.7]	11.5	[6.1, 20.0]	46.5	[38.7, 54.6]	0.673	[0.620, 0.725]
NHS MITT	75.0	[64.9, 83.0]	40.9	[32.7, 49.6]	25.0	[17.0, 35.1]	52.9	[44.7, 61.0]	0.579	[0.519, 0.639]
Other tools										
BCD Triage Sieve	72.9	[62.7, 81.2]	44.5	[36.1, 53.2]	27.1	[18.8, 37.3]	52.1	[43.7, 60.3]	0.587	[0.526, 0.648]
CareFlight	16.7	[10.1, 26.0]	89.1	[82.3, 93.5]	83.3	[74.0, 89.9]	48.4	[30.6, 66.6]	0.529	[0.483, 0.574]
MPTT-24	66.7	[56.2, 75.8]	43.1	[34.7, 51.8]	33.3	[24.2, 43.8]	54.9	[46.4, 63.2]	0.549	[0.486, 0.612]
MSTART	31.2	[22.4, 41.6]	82.5	[74.9, 88.2]	68.8	[58.4, 77.6]	44.4	[31.2, 58.5]	0.569	[0.512, 0.625]
NARU Triage Sieve	38.5	[28.9, 49.1]	75.9	[67.7, 82.6]	61.5	[50.9, 71.1]	47.1	[35.2, 59.4]	0.572	[0.512, 0.633]

In all subgroups, the BCD Triage Sieve has near-identical performance to its derivative, MITT. MSTART had moderate sensitivity for adults and children with blunt trauma (61.3 %and 62.6%, respectively) and low sensitivity amongst adults and children with penetrating trauma (35.6% and 31.2%, respectively), although MSTART demonstrated high specificity of over 82% in all subgroups and relatively low over-triage rates (32.8-50.2%, with the exception of elders).

(ii) Prediction of in-hospital mortality (see Supplementary data tables 7A-E)

Amongst blunt-injured adults aged 16 to 64 years, MITT was most sensitive (85% 95CI 83.5-86.4) for in-patient mortality, followed closely by Quick Triage at 80.3% (95CI 78.7-81.9) and then MSTART at 76.6% (95CI 74.9-78.3). Amongst adults (16-64 years) with penetrating injuries, with the exception of MPTT-24 (sensitivity 29.3% 95CI 21.9-37.9), all tools demonstrated sensitivity of 84% and above, with Quick Triage, TST and MITT achieving sensitivities above 95%.

All tools demonstrated sensitivities of 85-100% for inpatient mortality amongst blunt-injured children, and sensitivities of 80-100% in children with penetrating injuries. The exception to this was MPTT-24, with poor sensitivity (7.3%-36.4%) for in-patient mortality within all subgroups except for elders (sensitivity 45.9% 95CI 44.9-46.8). Amongst elders, MITT achieved the highest sensitivity in predicting in-patient mortality at 50.8% (95CI 49.9-51.8), with all other tools performing worse: the lowest sensitivity was demonstrated by TST (24.5% 95CI 23.7-25.3) and CareFlight (26.9% 95CI 26.1-27.8).

(iii) Prediction of ISS greater than 15 (see Supplementary data tables 8A-E)

Amongst all adults, 62,402 (31.9%) had ISS>15, of whom approximately a quarter (n=15,058) met criteria for intervention-based P1 status. A third (n=7,086) of intervention-based P1 patients had an ISS≤15. In predicting ISS>15 amongst blunt injured adults (16-64 years), MITT demonstrated the greatest sensitivity of 85.0% (95CI 83.5-86.4) followed by Quick Triage at 80.3% (95CI 78.7-81.9). CareFlight, NARU Triage Sieve and START had similar sensitivity in the range of 69.1 to 76.6%. TST's sensitivity for ISS>15 was 51.0% (95%CI 49.0-53.1). Amongst adults (16-64 years) with penetrating trauma Quick Triage, TST, MITT (as well as parent tool BCD) all exhibited sensitivities of over 95% for ISS>15, closely followed by MSTART (88.7%), NARU Triage Sieve (84.2%) and CareFlight (78.2%). MPTT-24 had the lowest sensitivity for ISS>15 amongst adults aged 16-64, in both blunt (33.7% 95CI 31.8-35.6) and penetrating trauma (29.3% 95CI 21.9-37.9) subgroups. Amongst elders, all tools performed poorly at predicting ISS >15, ranging in sensitivity from 41.4% (Quick Triage) to the lowest at 18.3% (CareFlight).

Amongst children, one third of patients (n=1617) had ISS>15, half (n=849) of whom met criteria for intervention-based P1 status. One third (n=494) of intervention-based P1 patients had an ISS≤15. The tools performed very differently in predicting ISS>15 as compared with adults and demonstrated great variation between penetrating and blunt-injured paediatric subgroups. In the blunt injured subgroup, MITT (similar to its parent tool, BCD) demonstrated the highest sensitivity of 77.2% (95CI 75.1-79.3), followed at some distance by MPTT-24 (sensitivity 57.7% 95CI 55.2-60.1) and Quick Triage (sensitivity 57.4% 95CI 54.9-59.8). TST had the lowest sensitivity for severe injuries at 36.5% (95%CI 34.2-39.0). Amongst children with penetrating trauma, Quick Triage was very highly sensitive at 95.5% (95CI 86.4-98.8), followed with a 25% margin by MITT (sensitivity 69.7% (95CI 57.0-80.1) and then TST at 77.3% (95CI 65.0-86.3).

DISCUSSION

This study introduces Quick Triage, a novel machine-learning derived major incident triage tool, assessing its performance using national trauma registry data. Quick Triage outperforms existing prominent tools in predicting the need for life-saving interventions across all clinically important subgroups, with a reduction in over-triage of 0.5% to 30.5% relative to MITT. The exception to this is in blunt-injured children, where MITT has a 17% sensitivity advantage over Quick Triage, however, Quick Triage outperforms MITT by a similar margin in children with penetrating trauma. MITT, proposed by NHS England for use by NHS first responders from April 2024, performs moderately well (sensitivity 68.4%- 84.7%; elders 57.9%) in predicting the need for life-saving interventions, although it is associated with the highest rates of over-triage (46.7-88.3%) amongst the tools studied. Quick Triage and MITT have similarly strong performance in predicting in-patient mortality and ISS>15 amongst adults; both offer drastic improvements in identifying P1 patients relative to the current UK tool, the NARU Triage Sieve. TST, proposed by NHS England for use by non-NHS first responders from April 2024, has undergone its first validation using real patients, including children. TST demonstrated excellent performance in penetrating trauma but identifies only half of blunt-injured adults and children and approximately a third of elders requiring life-saving interventions: considerable limitations not recognised prior to this study.

Although TST has been field-tested using simulated casualties, this is its first validation amongst real-life casualties, permitting evaluation of its strengths and weaknesses[22]. There are several practical implications of study findings. The high sensitivity of TST for P1 status in penetrating trauma (88.5% in children, 88.1% in adults aged 16-65) suggests that TST-labelled P1 casualties can be transferred from scene without ill consequence or need for re-triage. In blunt-injured patients assigned P1 status by TST, a high specificity (>95%) suggests that these patients too may be transferred and that re-triage with an alternative tool is unlikely to change their tool-assigned triage category. The danger lies with blunt-injured patients in whom TST assigns P2 status. A significant proportion of these will be "true" P1s (TST under-triage rate of blunt-injured P1 adults is 43.5%, 46.7% in children and 62.1% in elders): these casualties may experience delayed care at scene, and transfer to an inappropriate hospital for definitive care if not re-triaged using a more sensitive tool. If NHS England's recommendation for non-medical first responders to use TST is implemented, it is imperative that all blunt-injured TST-labelled P2 patients undergo re-triage with a more sensitive tool prior to transfer; perhaps use of distinct tags highlighting the need for re-triage is warranted. This vulnerability associated with TST is compounded by several real-life reports that secondary triage is frequently omitted during MIs. Delays in establishing a CCS, bypass of an existing CCS and immense pressure on first responders to clear the scene of casualties rapidly, sometimes at the expense of triage accuracy, are commonplace in UK and international MIs[1,20,31].

By contrast, study findings suggest that Quick Triage offers more balanced performance across the various subgroups (sensitivity of 77.0-92.1%, elders 67.1%) ensuring that over two thirds of P1s of any age with any injury mechanism are correctly identified, negating the need for secondary triage. Quick Triage's lower over-triage rates relative to MITT may reduce the risk of overwhelming hospitals with non-critical patients[10]. Hence, we recommend Quick Triage as a single tool for all first responders at the scene of MIs. Allowing complete interchangeability between triage conducted by NHS and non-NHS first responders, this simplification of the triage process will likely relieve the existing cognitive burden on all first responders, facilitate collaboration at scene to save lives, and allow the rapid onward transfer of casualties without compromising triage accuracy[3,21]. Promotion of a shared mental model and minimising the duplication of efforts at scene by first responders across various organisations wholly embraces the principles outlined by the Joint Emergency Service Interoperability Programme (JESIP), addressing shortfalls identified repeatedly following the London 7/7 bombings and 2017 Manchester Area reports[1,3,13]. Quick Triage may be re-applied as required, for example during prolonged field care, at a CCS if desired or after the administration of a life-saving intervention (e.g. tourniquet application).

Triage of children in MIs is particularly emotive and challenging[1,11]. Often, uninjured children experience over-triage to paediatric trauma centres following MIs, which are more vulnerable to being overwhelmed than their more numerous adult counterparts[20]: removing the need to measure and calculate respiratory and heart rates (as demanded by PTT, JumpSTART and MITT) may reduce inter-user variability and promote objective triage. In this computed, retrospective application of tools, it is unclear whether MITT's advantage over Quick Triage in blunt-injured children (sensitivity 84.7% vs 67.1%) would be sustained in practice, given that MITT is a more complex tool to apply. Quick Triage's lower over-triage rates in both paediatric subgroups (3-30% reduction relative to MITT) has obvious practical advantages. Although not measured in this study, PTT and JumpSTART's performance has been evaluated in using the same trauma registry data and definition of P1, demonstrating inferior sensitivity to Quick Triage and MITT (PTT 44.8% and JumpSTART 35.6%)[19]. At the other extreme of age, Quick Triage demonstrates strength in predicting the need for life-saving interventions amongst elders, perhaps reflecting that Quick Triage was modelled using a TARN population, in which elders are well represented. Future work should determine how over-triage in elders can be reduced, and how other determinants of outcome including frailty and presence of co-morbidities can be practically incorporated into the triage process.

Consistent with reports following real-life MIs[32], we note a higher proportion of P1s overall amongst patients with penetrating compared with blunt trauma (46.6% vs 11.3% in adults), but additionally, there is a 2.5-fold higher requirement for life and limb-saving surgical interventions: this is particularly relevant for the hospital phase of planning for MIs[11,12,33]. Our study findings confer with Cicero and Cross, who demonstrated only a moderate sensitivity of START (very similar to MSTART) and CareFlight for in-patient mortality amongst 500,000 patients from the US National Trauma Data Bank, mortality, as well as confirming that all tools have lesser discriminatory capability amongst elders relative to younger adults[34]. In their retrospective application of tools to data collected on victims of the London 7/7 bombings, Challen et al also reported sensitivity of only 50% for both START and CareFlight in predicting P1 status, although their study was limited by a paucity of prehospital physiological data, as is often the case following MIs[35]. CareFlight has widely been shown to have high specificity, but suboptimal sensitivity of approximately 50%, in several studies[35-37]. More recently, Jerome et al retrospectively evaluated multiple tools using over 8000 patients of all ages from the Alberta Trauma Registry, demonstrating similarly strong performance of MITT and BCD Triage Sieve (sensitivity of 70%), whilst START demonstrated sensitivity of below 50% in this study population[5]. Jerome et al highlighted the limitations of using mortality as an outcome measure of MI tool performance, stating that resources must be directed to patients who are critically ill but highlighting that a proportion of patients may have unsurvivable injuries; this is particularly relevant amongst elders[5]. In contrast to our findings, Jerome et al found that MPTT, from which MPTT-24 was derived, outperformed MITT[5]. This may relate to the authors' interpretation of MPTT's treatment of the non-breathing patient (e.g. assuming that an airway manoeuvre was implicitly permitted; 61% of P1 patients required advanced airway intervention and MPTT was reported as having a sensitivity of 76%, despite the MPTT tool stipulating that non-breathing patients should be assigned "Dead")[5]. Another possible explanation is that MPTT involves calculation of a full GCS score (MPTT-24 does not); this may convey triage accuracy on retrospective evaluation of MPTT, but is felt to be too complex and time consuming to apply accurately in the MI setting[6,38]. In our study, MPTT-24 has demonstrated modest sensitivity for the need for life-saving interventions yet poor sensitivity for predicting in-hospital mortality and ISS>15, despite an author-led investigation demonstrating that patients under-triaged by MPTT had the lowest mortality when compared to the NARU and MIMMS Triage Sieve tools[39]. A possible explanation for this is that the MPTT-24's high over-triage rate of P1 casualties, not reported in that study, may have included those who died by chance, rather than due to diagnostic accuracy; another possible explanation is that a different, author-derived definition of P1 was utilised, limiting direct comparison of the results. Based on the findings of this and a previous study, MPTT-24 cannot be recommended for use in UK MIs. TST and CareFlight tools have a high specificity (93.7% and 92.8%, respectively) for predicting the need for life-saving intervention amongst blunt-injured adults aged 16-64 years. Coupled with the knowledge that both tools are highly simple to apply, these data suggest that they might be useful as primary triage tools, however their relatively high under-triage rates (43.5% and 56.7%, respectively) make them less desirable as standalone tools, mandating the need for secondary triage prior to patient transfer.

A key strength of this study is the evaluation of triage tool performance using a large number of injured UK casualties, including important injury mechanism and age subgroups, allowing a thorough evaluation not afforded by previous studies[1,7,12]. Use of singly-injured trauma registry patients to measure triage tool performance affords a completeness of physiological data that cannot ordinarily be obtained in real MIs5[,8,34,35,40]. Obvious limitations of this approach are that singly-injured patients are not wholly representative of MI casualties with regards to injury type, severity and outcome[15,31]. For example, in a London-based study, Olding et al illustrated that knife-injured MI patients tend to be older, more evenly spread across genders and may suffer more severe injuries than those injured singly outside of the MI setting, potentially reflecting the lethal intent of perpetrators during MIs[15]. Sadly, paediatric patients within the TARN registry appear to suffer from less complete data collection compared to adults, an important deficiency which must be addressed[28].

Quick Triage fulfils the majority of the Model Uniform Core Criteria for Mass Casualty Triage (agreed by several US professional prehospital societies and the American College of Surgeons), making it an eligible candidate for use in the United States[38,41]. To be fully compliant with MUCC criteria, it may preceded by "global sorting" (walking/waving/still) assessment[6] prior to individual patient assessment and should include chest decompression as a life-saving intervention[41]. Consideration of Quick Triage for use in the US is particularly relevant since MSTART appears to have suboptimal sensitivity, particularly in penetrating trauma, as seen in this and other studies[34]. However prior validation of Quick Triage using mass-shooting or US civilian gunshot wound databases would be prudent since TARN contains only small proportions of low velocity GSW victims, limiting generalisability of study findings[7,20]. More studies are required to assess validity of the ability to walk as a marker of the absence of significant injuries, with one air-crash evaluation suggesting that 17% of ambulatory patients harboured significant P2 (but importantly, not P1) defining injuries[40].

A significant weakness of this study is that retrospective, computed application of triage tools does not take into account human factors: some tools are more difficult to apply in practice then others, affecting triage accuracy[34]. Reports from previous international MIs (e.g. 9/11 terrorist attacks) highlight that tools involving arithmetic calculations (e.g. heart rate, respiratory rate) are rapidly abandoned by users in high-threat situations, with experts advising a move away from such tools[20,31,38]. The practical applicability of tools can only be evaluated following real-life events or simulation studies involving humans[31]. We propose that Quick Triage utilises qualitative input variables that can be reliably measured by any first responder under challenging circumstances[25]. However, the need to assess for presence of chest injuries, which forms a major component of pre-hospital triage tools regularly applied by paramedics (e.g. ACS COT Field Triage Tool)[30], may necessitate additional training amongst non-medical first responders. Blast-injured patients are under-represented in the study population: use of available military trauma registries may prove useful for this purpose, but not without limited generalisability to the civilian setting[27]. Patients with burns, who form a significant proportion of casualties following explosions[42], and CBRN-related injury are similarly underrepresented in TARN, warranting further academic inquiry, although the authors could not identify any CBRN-related databases in which tool performance may be assessed.

In conclusion, this study describes a novel machine-learning derived MI triage tool, Quick Triage, and describes its superior performance to existing tools across a range of clinically significant age and mechanism subgroups, using a nationally representative patient population. Strengths and weaknesses of TST and MITT have been evaluated, including relatively high over-triage rates associated with MITT, enabling care providers and policymakers to anticipate and therefore mitigate any potential harm associated with their use. Whilst equipping non-medical responders to administer life-saving interventions, NHS England's recommended use of TST is associated with a significant risk of under-triage in blunt injured adults, children and elders[22]. Using Quick Triage as a single tool for all first responders may drastically simplify the triage process, promote better collaboration amongst first responders at scene and eliminate the aforementioned risk to blunt-injured P1 patients, whilst reducing over-triage rates[3]. This addresses several deficiencies raised during previous UK MIs, including the "care gap" highlighted in the Manchester Arena Inquiry and several human factors considerations raised following UK and international MIs[3,21]. We therefore recommend that Quick Triage replaces NHS MITT immediately as the tool of choice for NHS first responders and recommend that field-testing amongst police and fire services is undertaken with a view to universal roll-out for all emergency services involved in UK major incidents.

REFERENCES

1. 2017 Report of the Public Inquiry into the Attack on Manchester Arena on 22nd May 2017. Manchester Arena Inquiry Volume 2: Emergency Response, 2022.

2. (NARU) NARU. Clinical Guidance: Medical Support Minimal Requirements for a Mass Casualty Incident. www.naru.org.uk; 2014.

3. Programme JESI. JESIP Joint Doctrine: The Interoperability Framework Edition 3. 2022. https://www.jesip.org.uk/news/joint-doctrine-edition-3/.

4. Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehospital and disaster medicine 2008; 23(1): 3-8.

5. Jerome D, Savage DW, Pietrosanu M. An assessment of mass casualty triage systems using the Alberta trauma registry. Cjem 2023; 25(8): 659-66.

6. Lerner EB, McKee CH, Cady CE, et al. 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 2015; 19(2): 267-71.

7. Malik NS, Chernbumroong S, Xu Y, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. EClinicalMedicine.

8. Alarhayem AQ, Myers JG, Dent D, et al. Time is the enemy: Mortality in trauma patients with hemorrhage from torso injury occurs long before the “golden hour”. The American Journal of Surgery 2016; 212(6): 1101-5.

9. Moran CG, Lecky F, Bouamra O, et al. Changing the System - Major Trauma Patients and Their Outcomes in the NHS (England) 2008–17. EClinicalMedicine 2018; 2: 13-21.

10. Frykberg ER, Tepas JJ, 3rd. Terrorist bombings. Lessons learned from Belfast to Beirut. Annals of surgery 1988; 208(5): 569-76.

11. Moran CG, Webb C, Brohi K, Smith M, Willett K. Lessons in planning from mass casualty events in UK. Bmj 2017; 359: j4765.

12. Aylwin CJ, König TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. The Lancet 2006; 368(9554): 2219-25.

13. Home Office HG. Coroner's inquests into the London bombings of 7 July 2005: review of progress, 2011.

14. Government H. National Risk Register 2023 edition. 2023.

15. Olding J, Olding C, Bew D, Fan K. Penetrating head & neck trauma - Epidemiology and injury characteristics in terror-related violence, interpersonal violence and deliberate self-harm at a level 1 trauma centre. Surgeon 2019; 17(3): 133-8.

16. Office of National Statistics. Population estimates 2021. 2022.

17. NHS England. NHS Clinical Guidelines for Major Incidents and Mass Casualty Events (Version 2). 2020.

18. Vassallo J, Smith J. Major incident triage and the evaluation of the Triage Sort as a secondary triage method. Emergency Medicine Journal 2019; 36(5): 281.

19. Malik NS, Chernbumroong S, Xu Y, et al. Paediatric major incident triage: UK military tool offers best performance in predicting the need for time-critical major surgical and resuscitative intervention. EClinicalMedicine 2021; 40.

20. Pepper M, Archer F, Moloney J. Triage in Complex, Coordinated Terrorist Attacks. Prehospital and disaster medicine 2019; 34(4): 442-8.

21. Thompson J, Rehn M, Lossius HM, Lockey D. Risks to emergency medical responders at terrorist incidents: a narrative review of the medical literature. Crit Care 2014; 18(5): 521.

22. England N. Major incident triage tools. In: NHS Emergency Preparedness RaR, editor. NHS England; 2023.

23. James V, Chris GM, Philip C, Jason S. New NHS Prehospital Major Incident Triage Tool: from MIMMS to MITT. Emergency Medicine Journal 2022; 39(11): 800.

24. Johnston AM, Henning JD. Outcomes of patients treated at Bastion Hospital Intensive Care Unit in 2009. J R Army Med Corps 2014; 160(4): 323.

25. Yuanwei X, Nabeela M, Saisakul C, et al. Triage in major incidents: development and external validation of novel machine learning-derived primary and secondary triage tools. Emergency Medicine Journal 2023: emermed-2022-212440.

26. Teixeira PGR, Brown CVR, Emigh B, et al. Civilian Prehospital Tourniquet Use Is Associated with Improved Survival in Patients with Peripheral Vascular Injury. Journal of the American College of Surgeons 2018; 226(5): 769-76.e1.

27. Eastridge BJ, Mabry RL, Seguin P, et al. Death on the battlefield (2001-2011): implications for the future of combat casualty care. The journal of trauma and acute care surgery 2012; 73(6 Suppl 5): S431-7.

28. The University of Manchester. The Trauma Audit and Research Network. Continuous. www.tarn.ac.uk.

29. Vassallo J, Smith J. 2 Major incident triage and the implementation of a new triage tool, the MPTT-24. Emergency Medicine Journal 2017; 34(12): A860-A1.

30. American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient, 2014.

31. Asaeda G. The day that the START triage system came to a STOP: observations from the World Trade Center disaster. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2002; 9(3): 255-6.

32. Turner CD, Lockey DJ, Rehn M. Pre-hospital management of mass casualty civilian shootings: a systematic literature review. Crit Care 2016; 20(1): 362.

33. Kaplowitz L, Reece M, Hershey JH, Gilbert CM, Subbarao I. Regional health system response to the Virginia Tech mass casualty incident. Disaster medicine and public health preparedness 2007; 1(1 Suppl): S9-13.

34. Cross KP, Cicero MX. Head-to-Head Comparison of Disaster Triage Methods in Pediatric, Adult, and Geriatric Patients. Annals of emergency medicine 2013; 61(6): 668-76.e7.

35. Challen K, Walter D. Major incident triage: comparative validation using data from 7th July bombings. Injury 2013; 44(5): 629-33.

36. Cicero MX, Santillanes GR, Cross KP, Kaji AH, Donofrio JJ. Prehospital Disaster Triage Does Not Predict Pediatric Outcomes: Comparing the Criteria Outcomes Tool to Three Mass-Casualty Incident Triage Algorithms. Prehospital and disaster medicine 2021; 36(5): 503-10.

37. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the Accuracy of Mass Casualty Triage Systems When Used in an Adult Population. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2019: 1-8.

38. Lerner EB, Cone DC, Weinstein ES, et al. Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. Disaster medicine and public health preparedness 2011; 5(2): 129-37.

39. Vassallo JM, Smith JE, Wallis LA. Investigating the effects of under-triage by existing major incident triage tools. European journal of emergency medicine : official journal of the European Society for Emergency Medicine 2019; 26(2): 139-44.

40. Postma IL, Weel H, Heetveld MJ, et al. Mass casualty triage after an airplane crash near Amsterdam. Injury 2013; 44(8): 1061-7.

41. Guideline N, E. Brooke Lerner DCC, Eric S. Weinstein, Richard B. Schwartz, Phillip L. Coule, Michael Cronin, Ian S. Wedmore, Eileen M. Bulger, Deborah Ann Mulligan, Raymond E. Swienton, Scott M. Sasser, Umair A. Shah, Leonard J. Weireter, Teri L. Sanddal, Julio Lairet, David Markenson, Lou Romig, Gregg Lord, Jeffrey Salomone, Robert O'Connor and Richard C. Hunt. Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. Disaster medicine and public health preparedness 2011; 5: 129-37.

42. de Ceballos JPG, Turégano-Fuentes F, Perez-Diaz D, Sanz-Sanchez M, Martin-Llorente C, Guerrero-Sanz JE. 11 March 2004: The terrorist bomb explosions in Madrid, Spain – an analysis of the logistics, injuries sustained and clinical management of casualties treated at the closest hospital. Critical Care 2004; 9(1): 104.

SUPPLEMENTARY DATA

Supplementary data table 1: Criteria defining Priority 1 status with a breakdown of adult Priority 1 patients fulfilling each criterion

Subcomponents of the Priority 1 triage category	n (% of P1 patients)
An advanced airway intervention (e.g. intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	18,890 (85.3%)
Chest tube placed within 2 hours of arrival at hospital	4,301 (19.4%)
Neurologic, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	3,427 (15.5%)
Arrived in the ED with uncontrolled haemorrhage	1,979 (8.9%)
Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	606 (2.7%)
IV vasopressors administered within 2 hours of arrival at hospital	361 (1.6%)
Patient who required EMS initiation of CPR (i.e. had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	182 (0.8%)
Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	3 (0.0%)
Total number of P1 patients	22,144 (100.0%)

Note: There is overlap between life-saving interventions (LSI): 78.0% (n=17,272) of P1 patients required one LSI, 15.9% (n=3,520) required two LSI, and 6.0% (n=1,328) required 3 or more LSI.

Supplementary data table 2: Patient and injury characteristics of subgroup of adults with penetrating injuries (n=5,660)

Characteristic	n (%)
Gender	
Male	4,845 (85.6%)
Female	815 (14.3%)
Missing data	0
Injury Severity Score (ISS)	
Median (IQR)	9 [9, 16]
Missing data	0
Age	
Median (IQR), years	32.6 [23, 48]
Missing data	0
Discharge status	
Alive	5,305 (93.7%)
Dead	171 (3.1%)
Missing data	184 (3.2%)
Mechanism of Injury	
Penetrating	5,660
Of which:	
Stabbing	4,066 (71.8%)
Other*	1,159 (20.5%)
Shooting	410 (7.2%)
Blast	25 (0.44%)
Missing data	0
Triage Category	
Priority 1	2639 (46.6%)
Priority 2	2971 (52.5%)
Dead	21 (0.37%)
Expectant	29 (0.51%)

Supplementary data table 3: Comparison of the characteristics of included versus excluded patients

Variable	Included patients	Excluded patients
Male gender, n (%)	3447 (69.5%)*	6847 (67.3%)
ISS, median (IQR)	9 (9, 17)*	9 (9, 16)
Age (years), median (IQR)	11.9 (8.0, 14.2)*	3.9 (1.6, 10.1)
Mortality, n (%)	53 (1.1%)	316 (3.1%)*
Injury type		
Blunt	4733 (95.4%)	9935 (97.7%)*
Penetrating	229 (4.6%)*	236 (2.3%)
Injury mechanism		
Vehicle incident/collision	2459* (49.6%)	2262 (22.2%)
Fall less than 2m	1187* (23.9%)	4827 (47.5%)
Blow(s)	327 (6.6%)	943 (9.3%)
Crush	42 (0.9%)	118 (1.1%)
Fall more than 2m	645* (13.0%)	869 (8.5%)
Other	147 (3.0%)	971* (9.6%)
Burn	4 (0.1%)	40* (0.4%)
Stabbing	130* (2.6%)	99 (1.0%)
Shooting	16 (0.3%)	26 (0.3%)
Blast	5 (0.1%)	16 (0.2%)

Note: Included patients were compared to those excluded with respect to clinical and demographic characteristics. Fisher's exact test was used to compare categorical variables (gender, mortality and mode of injury (blunt and penetrating)). Continuous variables were compared using a Two-sample Kolmogorov-Smirnov Test. Differences in injury mechanism were estimated using the Chi-square test, where results were significant, post-hoc tests were performed to generate a p value. P values were adjusted using the Bonferroni correction. A value of $p<0{\cdot}05$ was considered statistically significant.

Patients were excluded on the basis of insufficient pre-hospital physiological data required to apply the tools (see Methods). OR=Odds ratio. Percentages represent the proportion of patients within the included (or excluded) group with the characteristic described (e.g. 69.5 % of all included patients were of male gender). * indicates the group that has higher number of incidents than expected, and is statistically significant ($p<0.05$).

Supplementary data Table 4: Breakdown of Priority 1 assignments amongst adults with penetrating trauma

Subcomponents of the Priority 1 triage category	n (% of P1 patients)
Neurologic, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	1,170 (44.3%)
Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	209 (7.9%)
Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	2 (0.076%)
Chest tube placed within 2 hours of arrival at hospital	864 (32.7%)
An advanced airway intervention (e.g. intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	976 (37.0%)
IV vasopressors administered within 2 hours of arrival at hospital	9 (0.34%)
Arrived in the ED with uncontrolled haemorrhage	326 (12.4%)
Patient who required EMS initiation of CPR (i.e. had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	10 (0.38%)
Total number of P1 patients	2,639 (100%)

Note: All adults with penetrating trauma fell within the 16-64 years age group

Supplementary data Table 5: Breakdown of paediatric Priority 1 patients according to Priority 1-defining criteria

Subcomponents of the Priority 1 triage category	n (%)
An advanced airway intervention (e.g. intubation, LMA, surgical airway) performed in the pre-hospital setting or within 4 hours of arrival at hospital	808 (75.2%)
Arrived in the ED with uncontrolled haemorrhage	310 (28.9%)
Neurologic, vascular, or haemorrhage-controlling surgery to the head, neck or torso performed within 4 hours of arrival to hospital	169 (15.7%)
Chest tube placed within 2 hours of arrival at hospital	70 (6.5%)
Limb-conserving surgery performed within 4 hours of arrival at hospital on a limb that was found to be pulseless distal to the injury prior to surgery	23 (2.1%)
IV vasopressors administered within 2 hours of arrival at hospital	6 (0.6%)
Patient who required EMS initiation of CPR (i.e. had a cardiac arrest) during transport, in the ED, or within 4 hours of arrival at a hospital	1 (0.1%)
Escharotomy performed on a patient with burns within 2 hours of arrival at a hospital	1 (0.1%)
Total number of P1 patients	1343 (100.0%)

Note: There is overlap between life-saving interventions (LSI): 73.2% (n=1006) of P1 patients required one LSI, 21.8% (n=299) required two LSI, and 5.0% (n=69) required 3 or more LSI.

Supplementary Data Table 7: Tool performance in predicting in-patient mortality

7A: Adults (16-64 years) with blunt trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	80.3	[78.7, 81.9]	66.4	[66.0, 66.7]	19.7	[18.1, 21.3]	93.9	[93.6, 94.1]	0.733	[0.725, 0.742]
Tools proposed by NHS England										
TST	69.7	[67.8, 71.5]	90.6	[90.4, 90.8]	30.3	[28.5, 32.2]	83.2	[82.5, 83.9]	0.801	[0.792, 0.810]
NHS MITT	85.0	[83.5, 86.4]	59.8	[59.5, 60.1]	15.0	[13.6, 16.5]	94.5	[94.3, 94.8]	0.724	[0.717, 0.732]
Other tools										
BCD Triage Sieve	84.6	[83.1, 86.0]	62.0	[61.7, 62.3]	15.4	[14.0, 16.9]	94.3	[94.0, 94.5]	0.733	[0.726, 0.740]
CareFlight	69.1	[67.2, 71.0]	88.5	[88.3, 88.7]	30.9	[29.0, 32.8]	85.9	[85.2, 86.5]	0.788	[0.779, 0.797]
MPTT-24	33.7	[31.8, 35.6]	62.1	[61.8, 62.4]	66.3	[64.4, 68.2]	97.6	[97.5, 97.8]	0.479	[0.469, 0.488]
MSTART	76.6	[74.9, 78.3]	83.3	[83.1, 83.6]	23.4	[21.7, 25.1]	88.8	[88.3, 89.3]	0.800	[0.791, 0.808]
NARU Triage Sieve	72.0	[70.2, 73.8]	85.0	[84.7, 85.2]	28.0	[26.2, 29.8]	88.4	[87.9, 88.9]	0.785	[0.776, 0.794]

7B: Adults (16-64 years) with penetrating trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	97.7	[93.0, 99.4]	19.2	[18.1, 20.3]	2.3	[0.6, 7.0]	96.9	[96.3, 97.4]	0.585	[0.571, 0.599]
Tools proposed by NHS England										
TST	95.5	[90.0, 98.2]	22.7	[21.6, 23.9]	4.5	[1.8, 10.0]	96.8	[96.2, 97.3]	0.591	[0.573, 0.610]
NHS MITT	96.2	[91.0, 98.6]	38.8	[37.5, 40.2]	3.8	[1.4, 9.0]	96.0	[95.3, 96.6]	0.675	[0.658, 0.693]
Other tools										
BCD Triage Sieve	96.2	[91.0, 98.6]	41.5	[40.1, 42.8]	3.8	[1.4, 9.0]	95.8	[95.1, 96.5]	0.688	[0.671, 0.706]
CareFlight	78.2	[70.0, 84.7]	85.5	[84.5, 86.4]	21.8	[15.3, 30.0]	87.5	[85.1, 89.7]	0.818	[0.783, 0.854]
MPTT-24	29.3	[21.9, 37.9]	41.4	[40.0, 42.7]	70.7	[62.1, 78.1]	98.7	[98.2, 99.1]	0.353	[0.314, 0.393]
MSTART	88.7	[81.8, 93.3]	76.0	[74.8, 77.2]	11.3	[6.7, 18.2]	91.1	[89.4, 92.5]	0.824	[0.796, 0.851]
NARU Triage Sieve	84.2	[76.6, 89.7]	73.2	[71.9, 74.4]	15.8	[10.3, 23.4]	92.3	[90.8, 93.6]	0.787	[0.755, 0.819]

7C: Elders (65+ years)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	44.5	[43.6, 45.4]	74.6	[74.3, 74.9]	55.5	[54.6, 56.4]	82.2	[81.7, 82.6]	0.595	[0.591, 0.600]
Tools proposed by NHS England										
TST	24.5	[23.7, 25.3]	95.9	[95.8, 96.1]	75.5	[74.7, 76.3]	57.3	[56.1, 58.5]	0.602	[0.598, 0.606]
NHS MITT	50.8	[49.9, 51.8]	71.8	[71.5, 72.1]	49.2	[48.2, 50.1]	81.7	[81.3, 82.2]	0.613	[0.609, 0.618]
Other tools										
BCD Triage Sieve	49.7	[48.8, 50.7]	73.4	[73.1, 73.7]	50.3	[49.3, 51.2]	81.2	[80.7, 81.6]	0.616	[0.611, 0.621]
CareFlight	26.9	[26.1, 27.8]	94.1	[93.9, 94.2]	73.1	[72.2, 73.9]	63.9	[62.9, 65.0]	0.605	[0.601, 0.609]
MPTT-24	45.9	[44.9, 46.8]	71.0	[70.7, 71.3]	54.1	[53.2, 55.1]	83.6	[83.2, 84.0]	0.584	[0.579, 0.589]
MSTART	35.7	[34.8, 36.6]	88.9	[88.7, 89.1]	64.3	[63.4, 65.2]	71.6	[70.8, 72.3]	0.623	[0.618, 0.627]
NARU Triage Sieve	29.4	[28.6, 30.3]	90.4	[90.2, 90.6]	70.6	[69.7, 71.4]	72.5	[71.6, 73.3]	0.599	[0.595, 0.604]

7D: Paediatric patients (<16 years) with blunt trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	95.1	[87.3, 98.4]	76.4	[75.2, 77.6]	4.9	[1.6, 12.7]	93.4	[91.8, 94.7]	0.858	[0.834, 0.882]
Tools proposed by NHS England										
TST	93.9	[85.7, 97.7]	86.6	[85.6, 87.6]	6.1	[2.3, 14.3]	89.1	[86.5, 91.2]	0.903	[0.876, 0.929]
NHS MITT	100.0	[94.4, 100.0]	33.6	[32.2, 34.9]	0.0	[0.0, 0.06]	97.4	[96.8, 97.9]	0.668	[0.661, 0.675]
Other tools										
BCD Triage Sieve	96.3	[88.9, 99.1]	37.3	[35.9, 38.7]	3.7	[0.9, 11.1]	97.4	[96.7, 97.9]	0.668	[0.647, 0.690]
CareFlight	85.4	[75.4, 91.9]	85.5	[84.5, 86.5]	14.6	[8.1, 24.6]	90.7	[88.3, 92.6]	0.855	[0.816, 0.893]
MPTT-24	7.3	[3.0, 15.8]	40.1	[38.7, 41.5]	92.7	[84.2, 97.0]	99.8	[99.5, 99.9]	0.237	[0.208, 0.266]
MSTART	95.1	[87.3, 98.4]	77.8	[76.6, 79.0]	4.9	[1.6, 12.7]	93.0	[91.4, 94.4]	0.865	[0.840, 0.889]
NARU Triage Sieve	92.7	[84.2, 97.0]	71.8	[70.5, 73.1]	7.3	[3.0, 15.8]	94.6	[93.2, 95.7]	0.822	[0.793, 0.851]

7E: Paediatric patients (<16 years) with penetrating trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	100	[67.9, 100.0]	24.3	[18.9, 30.6]	0	[0.0, 32.1]	93.9	[89.0, 96.7]	0.622	[0.593, 0.650]
Tools proposed by NHS England										
TST	100.0	[67.9, 100.0]	33.3	[27.3, 40.0]	0.0	[0.0, 32.1]	93.1	[87.7, 96.3]	0.667	[0.636, 0.698]
NHS MITT	100.0	[67.9, 100.0]	36.0	[29.8, 42.8]	0.0	[0.0, 32.1]	92.8	[87.2, 96.2]	0.680	[0.649, 0.712]
Other tools										
BCD Triage Sieve	100.0	[67.9, 100.0]	39.2	[32.8, 46.0]	0.0	[0.0, 32.1]	92.5	[86.6, 96.0]	0.696	[0.664, 0.728]
CareFlight	81.8	[47.8, 96.8]	90.1	[85.2, 93.5]	18.2	[3.2, 52.2]	71.0	[51.8, 85.1]	0.860	[0.738, 0.981]
MPTT-24	36.4	[12.4, 68.4]	37.8	[31.5, 44.6]	63.6	[31.6, 87.6]	97.2	[92.5, 99.1]	0.371	[0.219, 0.523]
MSTART	100.0	[67.9, 100.0]	80.6	[74.7, 85.5]	0.0	[0.0, 32.1]	79.6	[66.1, 88.9]	0.903	[0.877, 0.929]
NARU Triage Sieve	81.8	[47.8, 96.8]	72.5	[66.1, 78.2]	18.2	[3.2, 52.2]	87.1	[76.5, 93.6]	0.772	[0.649, 0.895]

Supplementary Data Table 8: Tool performance in predicting in-patient mortality

8A: Adults (16-64 years) with blunt trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	80.3	[78.7, 81.9]	66.4	[66.0, 66.7]	19.7	[18.1, 21.3]	93.9	[93.6, 94.1]	0.733	[0.725, 0.742]
Tools proposed by NHS England										
TST	51.0	[49.0, 53.1]	95.4	[95.2, 95.5]	49.0	[46.9, 51.0]	76.8	[75.6, 77.9]	0.732	[0.722, 0.742]
NHS MITT	85.0	[83.5, 86.4]	59.8	[59.5, 60.1]	15.0	[13.6, 16.5]	94.5	[94.3, 94.8]	0.724	[0.717, 0.732]
Other tools										
BCD Triage Sieve	84.6	[83.1, 86.0]	62.0	[61.7, 62.3]	15.4	[14.0, 16.9]	94.3	[94.0, 94.5]	0.733	[0.726, 0.740]
CareFlight	69.1	[67.2, 71.0]	88.5	[88.3, 88.7]	30.9	[29.0, 32.8]	85.9	[85.2, 86.5]	0.788	[0.779, 0.797]
MPTT-24	33.7	[31.8, 35.6]	62.1	[61.8, 62.4]	66.3	[64.4, 68.2]	97.6	[97.5, 97.8]	0.479	[0.469, 0.488]
MSTART	76.6	[74.9, 78.3]	83.3	[83.1, 83.6]	23.4	[21.7, 25.1]	88.8	[88.3, 89.3]	0.800	[0.791, 0.808]
NARU Triage Sieve	72.0	[70.2, 73.8]	85.0	[84.7, 85.2]	28.0	[26.2, 29.8]	88.4	[87.9, 88.9]	0.785	[0.776, 0.794]

8B: Adults (16-64 years) with penetrating trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	97.7	[93.0, 99.4]	19.2	[18.1, 20.3]	2.3	[0.6, 7.0]	96.9	[96.3, 97.4]	0.585	[0.571, 0.599]
Tools proposed by NHS England										
TST	95.5	[90.0, 98.2]	22.7	[21.6, 23.9]	4.5	[1.8, 10.0]	96.8	[96.2, 97.3]	0.591	[0.573, 0.610]
NHS MITT	96.2	[91.0, 98.6]	38.8	[37.5, 40.2]	3.8	[1.4, 9.0]	96.0	[95.3, 96.6]	0.675	[0.658, 0.693]
Other tools										
BCD Triage Sieve	96.2	[91.0, 98.6]	41.5	[40.1, 42.8]	3.8	[1.4, 9.0]	95.8	[95.1, 96.5]	0.688	[0.671, 0.706]
CareFlight	78.2	[70.0, 84.7]	85.5	[84.5, 86.4]	21.8	[15.3, 30.0]	87.5	[85.1, 89.7]	0.818	[0.783, 0.854]
MPTT-24	29.3	[21.9, 37.9]	41.4	[40.0, 42.7]	70.7	[62.1, 78.1]	98.7	[98.2, 99.1]	0.353	[0.314, 0.393]
MSTART	88.7	[81.8, 93.3]	76.0	[74.8, 77.2]	11.3	[6.7, 18.2]	91.1	[89.4, 92.5]	0.824	[0.796, 0.851]
NARU Triage Sieve	84.2	[76.6, 89.7]	73.2	[71.9, 74.4]	15.8	[10.3, 23.4]	92.3	[90.8, 93.6]	0.787	[0.755, 0.819]

8C: Elders (65+ years)

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	41.4	[40.8, 41.9]	78.3	[78.0, 78.6]	58.6	[58.1, 59.2]	55.3	[54.7, 55.9]	0.598	[0.595, 0.602]
Tools proposed by NHS England										
TST	16.2	[15.8, 16.7]	97.9	[97.8, 98.0]	83.8	[83.3, 84.2]	23.7	[22.7, 24.8]	0.570	[0.568, 0.573]
NHS MITT	39.7	[39.1, 40.2]	73.1	[72.8, 73.5]	60.3	[59.8, 60.9]	61.5	[61.0, 62.1]	0.564	[0.561, 0.567]
Other tools										
BCD Triage Sieve	38.3	[37.8, 38.9]	74.8	[74.4, 75.1]	61.7	[61.1, 62.2]	60.9	[60.3, 61.4]	0.565	[0.562, 0.569]
CareFlight	18.3	[17.9, 18.8]	96.0	[95.9, 96.2]	81.7	[81.2, 82.1]	33.8	[32.8, 34.8]	0.572	[0.570, 0.574]
MPTT-24	39.3	[38.7, 39.8]	72.7	[72.3, 73.0]	60.7	[60.2, 61.3]	62.2	[61.6, 62.7]	0.560	[0.557, 0.563]
MSTART	24.7	[24.2, 25.2]	90.7	[90.5, 91.0]	75.3	[74.8, 75.8]	47.0	[46.2, 47.8]	0.577	[0.574, 0.580]
NARU Triage Sieve	19.0	[18.5, 19.4]	91.3	[91.0, 91.5]	81.0	[80.6, 81.5]	52.1	[51.2, 53.0]	0.551	[0.549, 0.554]

8D: Paediatric patients (<16 years) with blunt trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	57.4	[54.9, 59.8]	91.6	[90.6, 92.5]	42.6	[40.2, 45.1]	22.5	[20.2, 25.0]	0.745	[0.732, 0.758]
Tools proposed by NHS England										
TST	36.5	[34.2, 39.0]	96.2	[95.5, 96.9]	63.5	[61.0, 65.8]	17.0	[14.3, 20.0]	0.664	[0.652, 0.676]
NHS MITT	77.2	[75.1, 79.3]	38.2	[36.5, 39.9]	22.8	[20.7, 24.9]	61.4	[59.7, 63.1]	0.577	[0.564, 0.590]
Other tools										
BCD Triage Sieve	74.6	[72.4, 76.7]	42.4	[40.7, 44.2]	25.4	[23.3, 27.6]	60.5	[58.7, 62.2]	0.585	[0.572, 0.599]
CareFlight	37.5	[35.1, 39.9]	95.3	[94.5, 96.0]	62.5	[60.1, 64.9]	19.9	[17.1, 22.9]	0.664	[0.652, 0.676]
MPTT-24	57.7	[55.2, 60.1]	40.3	[38.6, 42.1]	42.3	[39.9, 44.8]	67.2	[65.5, 69.0]	0.490	[0.475, 0.505]
MSTART	45.6	[43.2, 48.1]	87.7	[86.5, 88.9]	54.4	[51.9, 56.8]	34.8	[32.0, 37.6]	0.667	[0.653, 0.680]
NARU Triage Sieve	45.4	[43.0, 47.9]	78.8	[77.3, 80.2]	54.6	[52.1, 57.0]	48.0	[45.4, 50.7]	0.621	[0.607, 0.635]

8E: Paediatric patients (<16 years) with penetrating trauma

	Sensitivity		Specificity		Undertriage		Overtriage		AUC	
Quick Triage	95.5	[86.4, 98.8]	30.5	[23.8, 38.2]	4.5	[1.2, 13.6]	64.8	[57.3, 71.7]	0.63	[0.587, 0.673]
Tools proposed by NHS England										
TST	77.3	[65.0, 86.3]	35.3	[28.2, 43.1]	22.7	[13.7, 35.0]	67.9	[60.0, 75.0]	0.563	[0.500, 0.626]
NHS MITT	69.7	[57.0, 80.1]	35.9	[28.8, 43.8]	30.3	[19.9, 43.0]	69.9	[61.9, 76.9]	0.528	[0.461, 0.595]
Other tools										
BCD Triage Sieve	69.7	[57.0, 80.1]	40.1	[32.7, 48.0]	30.3	[19.9, 43.0]	68.5	[60.2, 75.8]	0.549	[0.482, 0.616]
CareFlight	24.2	[14.9, 36.6]	91.0	[85.4, 94.7]	75.8	[63.4, 85.1]	48.4	[30.6, 66.6]	0.576	[0.520, 0.633]
MPTT-24	57.6	[44.8, 69.4]	37.7	[30.4, 45.6]	42.4	[30.6, 55.2]	73.2	[65.0, 80.2]	0.477	[0.406, 0.547]
MSTART	37.9	[26.5, 50.7]	82.6	[75.8, 87.9]	62.1	[49.3, 73.5]	53.7	[39.7, 67.2]	0.603	[0.537, 0.668]
NARU Triage Sieve	31.8	[21.2, 44.6]	70.7	[63.0, 77.3]	68.2	[55.4, 78.8]	70.0	[57.7, 80.1]	0.512	[0.446, 0.579]

Chapter 4 Conclusions

Quick Triage has balanced, favourable performance amongst adults, children and elders with blunt and penetrating trauma from the UK national trauma registry, with lower over-triage rates than the NHS MITT and BCD Triage Sieve. This offers the possibility of a single tool for use amongst patients of all ages by all NHS and non-NHS first responders, allowing complete interchangeability, simplification of the prehospital triage process and likely more rapid and accurate triage of major incident casualties in future UK and international major incidents.

A key limitation of Chapter 4 is the use of TARN patients with penetrating trauma as surrogates for those injured in UK MIs. TARN patients with trauma consist largely of young men involved in interpersonal violence involving knives, which has limited generalisability to MI patients with bladed, blast or ballistic injuries. However, this approach has enabled the examination of tool performance in a relatively large number of patients with pre-hospital and outcome data which has a high degree of completeness. Another limitation is that Quick Triage has been developed using patients from the TARN database, and future work should include validation of this tool in a geographically and temporally distinct patient population, ideally using data from real and contemporary UK MIs. Simulation studies to determine inter- and intra-user variability in the application of Quick Triage, NHS MITT and TST would also be beneficial in informing policy-makers in their choice of triage tool for use in future MIs.

CHAPTER 5: DISCUSSION

Discussion

Major incidents are characterised by a mismatch between the immediate needs of patients and the resources available to treat them. In the UK, major incidents comprising blast and penetrating trauma continue to constitute the greatest threat to national security[1-3]. Triage tools are crucial in ensuring that finite healthcare resources are allocated to maximise overall survival. Existing tools have largely been developed based on expert opinion and implemented without formal validation. A major challenge in furthering the science of major incident management is lack of consensus and standardisation of end-points which best define Priority 1 (P1) status and other triage categories (Chapter 1). In this thesis, the literature surrounding major incident triage was reviewed and modern, evidence-based definitions of triage categories (Lerner's criteria) were selected. A Priority 1 patient is defined as one needing one of eight time-critical life-saving interventions, directly reflecting patient resource requirements. When validated amongst adults and children from the UK national trauma registry, Lerner's triage category definitions correlate well with casualty distribution within UK regionalised systems of trauma care (Chapter 2). Hence it is recommended that Lerner's criteria should become the gold standard in reporting triage tool performance in future UK major incidents. These intervention-based triage category definitions were compared with the triage category assignments of ten existing international triage tools (Chapter 2, publications number 1 and 2). This comparative analysis demonstrated that the current UK national triage tool, the NARU Triage Sieve, is poorly sensitive in identifying P1 patients who require time-critical, life-saving interventions, and was surpassed in sensitivity by several tools, the best of which was the UK military's BCD Triage Sieve (Chapter 2, publication number 1). Similarly, in children from the national trauma registry, the current UK tools, the Paediatric Triage Sieve and JumpSTART, demonstrated suboptimal performance and were substantially outperformed by several tools, with the BCD Triage Sieve again demonstrating the greatest sensitivity (Chapter 2, publication number 2). This has directly informed national policy, resulting in the removal of previous existing tools and a recommendation for the NHS Major Incident Triage Tool (MITT), the civilian adaptation of the BCD Triage Sieve, to become a single tool for NHS first responders to apply to all adults and children in future UK major incidents[4]. Subsequently, tree-based machine learning methodology has been employed to develop a novel primary triage tool, Quick Triage, as well as a secondary triage tool that is applicable via a portable device, both of which outperform the BCD Triage Sieve amongst TARN patients, with reduced associated over-triage rates (Chapter 3, publication number 3). The app-based secondary triage tool has withstood external validation amongst injured patients from the UK military's Joint Theatre Trauma Registry (JTTR); however, a paucity of pre-hospital data precluded external validation of Quick Triage amongst JTTR patients, necessitating further work. Quick Triage has balanced, favourable performance amongst adults, children and elders with blunt and penetrating trauma from the UK national trauma registry, with lower over-triage rates than the NHS Major Incident Triage Tool (MITT) (Chapter 4). This offers the possibility of a single tool for use amongst patients of all ages by all NHS and non-NHS first responders, allowing complete interchangeability, simplification of the prehospital triage process and likely more rapid and accurate triage of major incident casualties in future UK and international major incidents.

With a paucity of literature regarding the performance of MI tools and a lack of standardisation of end-points used to measure the effectiveness of tools, coupled with limited generalisability of existing studies to the UK setting, this thesis had the broader aim of determining the optimal method of prehospital triage in UK major incidents.

More specifically, the psrticular aims of this thesis were:

1. To determine an objective, practically applicable and resource-based system to define major incident triage categories, and to validate this using a nationally representative sample of adults and children

2. To compare the performance of existing international tools in adults, older adults and children across injuries of all mechanisms using a nationally representative sample of injured patients, in order to determine the best performing existing tool

3. To identify the best performing existing triage tool for use in the most prevalent type of UK major incident, i.e. terrorist-related major incidents, which are characterised by blast and penetrating trauma

4. To use machine learning to develop novel primary and secondary triage tools for the UK civilian setting and to validate these externally using a geographically distinct population of injured patients

Multiple existing major incident triage tools were identified (Chapter 1), including several adult primary triage tools START, mSTART, MIMMS Triage Sieve, NARU Triage Sieve, MPTT, MPTT-24, SALT, ASAV, CareFlight, BCD Triage Sieve, two paediatric-specific primary triage tools (JumpSTART and PTT) as well as two points-based triage tools, the MIMMS Triage Sort (a secondary triage tool) and the Sacco Triage Method[5-10]. Several challenges in evaluating the performance of MI triage tools have been recognised. It is clear that randomised controlled trials of triage tool performance in MIs are both impractical, given the unpredictable nature of MIs, but also unethical, with all those involved focussing on the provision of care whilst the resources available to treat patients. As such, in order to inform practice, policy makers and care providers must rely on assimilating the findings of other types of study design, each of which is prone to particular biases (Chapter 1). Post-event evaluations, which fully assimilate the human factors challenges of working in thi s setting, often have small numbers of patients involved and a single predominating injury mechanism[11,12]. Often a single tool has been used, preventing the comparison of multiple tools. Incomplete prehospital data collection in this setting limits meaningful retrospective application of triage tools as a means of assessing triage tool performance[13]. Some studies evaluate tool performance by applying triage tools to patients attending hospitals following injuries incurred in regular day-to-day practice allowing larger patient numbers encompassing different mechanisms of injury to be included[14,15]. However, the findings of such studies may have limited generalisability due to application of the tools within a hospital rather than pre-hospital setting and the type and severity of injuries amongst patient populations may not be representative of patients in real MIs.

A method growing in popularity is the retrospective application of tools to high-quality prospectively maintained trauma registries, which eliminates information bias and allows very large numbers of patients, including injured children, to be included[16-18]. The present work employed this method, utilising 195,000 adults and 4,962 children from the UK national trauma registry (Chapter 2, publications number 1 and 2). The strengths of this approach were that computed application of tools can be assessed independently of human error, allowing the innate discriminatory capability of tools to be assessed[16,19,20]. TARN is the largest trauma registry in Europe, and unlike its US-based counterpart the National Trauma Data Bank, submission of patient data is mandatory for all trauma-receiving hospitals, giving rise to a nationally representative sample of patients[16,21]. Even so, amongst the children studied, nearly two thirds of children lacked the complete physiological data required to apply triage tools; this highlights an important shortcoming which TARN must address[21]. Like all trauma registries, TARN has very specific inclusion criteria, including a length of stay of over 72 hours following injury[16,21]. This criterion in particular leads to an over-representation of elders in the study sample, the resulting bias was anticipated and has been mititgated by analysing tool performance separately in young (16-64 years) and older (65+ years) adults. The subsequent use of this TARN database in the modelling of a novel triage tool (Quick Triage) has unsurprisingly resulted in a tool which demonstrated a 20% improvement in sensitivity in predicting the need for life-saving interventions amongst elders when compared with the best existing tool, the NHS MITT (Chapter 3, publication number 3 and Chapter 4).

An obvious limitation of the retrospective application of tools to registry data is that some tools are more complex to apply in practice than others, and are more prone to inter-user variability in their application. Examination of this effect is best assessed by comparing one to two tools in a simulated setting, either virtually or using model patients, to assess how quickly and accurately tools can be applied in practice[22,23]. Clearly, there is abundant evidence from real-world events that tools involving arithmetic calculation are readily abandoned by first responders in the face of an ongoing threat in terror-related major incidents, as described clearly following the 9/11 New York World Trade Centre events[11]. However, it is not uncommon for identical "lessons learnt" to be noted in consecutive major incidents, without these lessons being fully implemented in practice[1,2,24]. The development of our novel Quick Triage tool (Chapter 4) has taken into account several human factors considerations raised by previous post-incident reports and the US Centre for Disease control's Model Uniform Core Criteria[11,25]. However, as with any novel tool, there is a requirement to undertake simulated studies involving human users, particularly amongst non-medically trained first responders, to determine whether the favourable performance in a computed setting is also demonstrated in practice.

Regarding the endpoints used to measure triage tool performance, much variation is seen between studies, which previously limited direct comparison between studies and meaningful assimilation of results (Chapter 1). In keeping with the observations of Baxt and Upenieks in 1990 and several subsequent studies, we have again demonstrated the lack of full correlation between patients' injury severity score and the resources required to treat them (Chapter 2, publications number 1 and 2). In the context of MI triage, matching finite resources with those that need them the most is of the utmost priority. In our adult study, only 24.1% (n=15,058) of patients met criteria for intervention-based P1 status whilst 75.9% (n=47,344) were non-P1 (Chapter 2, publication number 1). One third (n=7086) of intervention-based adult P1 patients had an ISS≤15. Nevertheless, reporting prediction of ISS>15 allows direct comparison of MI triage tools to routinely used field triage tools and is a familiar end-point amongst trauma clinicians and policymakers, making it a useful secondary outcome measure of triage tool performance[26,27]. Although mortality is the simplest outcome to measure, prediction of in-patient mortality as a measure of tool performance may be confounded by the presence of patients with non-survivable injuries and again, is imperfect as a sole measure of trauma tool performance, particularly amongst elders[17,26]. However, the utility of tools which fail to assign P1 status to a large proportion of patients who go on to die must be questioned (Chapter 2, publication number 1 and Chapter 4). The present work has demonstrated this to be true of the MPTT and MPTT-24 tools, which prior to the present work, had only been evaluated by its authors[6,7]. In this thesis, the literature surrounding major incident triage was reviewed (Chapter 1) and a modern, evidence-based system of defining triage categories was validated amongst UK national trauma registry patients (Chapter 2, publication number 1). This revealed clinically and statistically meaningful differences in mortality, ICU usage, length of stay and injury severity, which correlated well with casualty distribution within UK regionalised systems of trauma care[26]. For example, P1 casualties were found to have a mortality of 12.5% and ICU requirement of over 50%, compared with 5% and 5% respectively, amongst P2 patients. This aligns well with the tendency for P1 casualties to be distributed to a major trauma centre, whereas the resource requirements of a P2 patient are likely to be met within the trauma unit setting[24,28]. An Utstein-style template to standardise the reporting of the acute medical response in disasters has previously highlighted that there were no common or standardised definitions of each triage category[29]. We recommend Lerner's criteria as the gold standard in measuring triage tool performance in subsequent UK and international major incidents (Chapter 2, publications number 1 and 2). Standardisation of primary outcome measurement across multiple MIs will enable direct comparison to be made between studies, allowing continued refinement of policy, practice and service delivery[29]. Following our use of Lerner's criteria as a gold standard, our study design has been replicated in a Canadian trauma registry and it is anticipated that use of standard will grow elsewhere, facilitating refinements in practices internationally[17].

In Chapter 3, we describe the use of tree-based machine learning methods to model novel primary and secondary triage tools. Machine learning has several advantages over traditional approaches in the modelling of novel triage tools. In particular, tree-based machine learning models have demonstrated utility in clinical risk stratification, with the ability to capture non-linear interactions between input variables, a key notable strength when compared with traditional statistical methods. Decision trees have the particular advantage of mirroring the binary (yes/no or present/absent) bifurcating format of contemporary triage tools, whilst more complex tree-based methods random forest and eXtreme Gradient Boosting (XGB) methods lend themselves to app development (Chapter 3). External validation of machine learning-derived clinical decision support tools is an essential step prior to clinical implementation. For this purpose, patients from the UK military's Joint Theatre Trauma Registry (JTTR) of combat casualties were utilised, with the advantage that blast and penetrating mechanisms of injury, which predominate in recent UK major incidents, are well represented. We describe the development of a novel primary triage tool (Chapter 3, publication number 3), later adapted into the clinically applicable "Quick Triage." Unfortunately, due to lack of GCS subcomponent data within the JTTR databse, Quick Triage could not accurately be externally validated using JTTR. We also describe the development of an XGB-based secondary tool applicable via a portable device, which demonstrated excellent external validity amongst JTTR patients.

National policy regarding major incident triage falls within the remit of NHS England's Emergency Preparedness, Resilience and Response Clinical Reference Group (NHS EPRR CRG). This group created a "Major Incident Triage Task and Finish Group" into which Nabeela Malik was invited (2021-present) as a subject matter expert and academic advisor. Based on the findings of the adult and paediatric comparative analyses of triage tools (Chapter 2, publications 1 and 2), the NARU Triage Sieve, Paediatric Triage Tape and JumpSTART are no longer recommended for use. In line with recommendations from these studies, the BCD Triage Sieve was identified as the most appropriate tool to take forward as a single tool for patients of all ages, with the UK Surgeon General granting permission for a civilian adaptation of this tool, now named the NHS Major Incident Triage Tool (MITT). NHS England has recommended that from April 2024, all NHS first responders are to utilise

NHS MITT as the prehospital triage tool of choice. Additionally, NHS England sought to address the "care gap" described in the report of the 2017 Manchester Arena bombing and equip non-NHS first responders with a triage tool, in order that the triage process may begin and basic life-saving interventions administered ahead of the arrival of trained healthcare providers, in line with JESIP principles[2,4]. NHS England has recommended Ten Second Triage (TST), an ultra-simple, expert-derived qualitative tool previously evaluated solely in a simulated setting, for use by non-NHS first responders (i.e. fire and police)[4]. The findings of Chapter 4 have been disseminated amongst members of the NHS EPRR CRG, including TST's potential to under-triage half of all blunt-injured P1 adult, paediatric and elderly patients if re-triage with a more sensitive tool is not conducted prior to casualty transfer. Quick Triage's superiority to MITT across the majority of clinically important subgroups, with substantially lower associated over-triage, has also been conveyed to policymakers ahead of attaining peer-review. External validation of Quick Triage has also been completed (not included here), with favourable results, and is undergoing peer review. It is hoped that field-testing of Quick Triage will ensue, with the possibility of incorporation into national policy at a later stage as a single tool for use in patients of all ages and by all first responders (Chapter 4). This may further simplify the triage process, allow full interchangeability and a shared mental model between first responders, and likely lead to more rapid and accurate patient triage with better clinical outcomes including survival following UK major incidents. Within the context of an ongoing threat from terror-related major incidents on public places, which often includes but is not limited to penetrating trauma, simplification and refinement of the triage process is of great relevance at present[3].

There are important limitations in generalising the findings obtained in a TARN dataset to MI patient populations. The prevalent mechanisms of injury in recent UK MIs have been blast and penetrating trauma, which are markedly underrepresented in the TARN database. Only 142 (0.1%) of the 195,709 TARN patients were injured in blasts. Penetrating injuries constitute only 3% of the entire TARN database. Whilst this was mitigated by performing a subgroup analysis of the 5,660 patients with penetrating trauma (Chapter 4), it is clear that these patients, who are largely young males injured with knives during interpersonal violence, differ qualitatively from patients injured in penetrating MIs, who are older, include both genders equally and seem to demonstrate an anatomical injury pattern consistent with lethal intent, such as a greater proportion of injuries to the neck[30]. In future, any UK or Western European healthcare-related data captured on victims of MIs should be made available for analysis with regards to triage assignments and patient outcomes. The paucity of data captured in such events, which limits analysis, may be augmented by the use of technology such as wearable sensors or live video footage of interventions performed and observations recorded. The use of military databases are not without limitation, particularly as they consist of physically and medically fit young males, and are also limited in their generalisability to the UK civilian MI setting.

Another potential source of bias in the study is the use of Lerner et al's definitions of P1 and other triage categories, and the observation that these align well with casualty distribution plans within the UK. Firstly, only 11% of the TARN population met P1 status according to the need for the eight time-critical life-saving interventions, including only 24.1% of the 62,402 patients with ISS>15 (Chapter 1). In regionalised systems of care, ISS>15 is seen as a justification to warrant MTC care, and these data suggest that a much greater proportion of the TARN population deserve MTC care. Hence, the elevated over-triage rates seen with our existing tools and Quick Triage may not be as alarming as they appear. There is much heterogeneity amongst trauma units (TUs) within the UK. Assuming that all of these can adequately meet the needs of the sicker of the P2 patients, who may require major resuscitative and surgical interventions described by Lerner et al, but not within the two- and four-hour cutoffs suggested, cannot be ascertained. There is a requirement for further work, ideally commissioned by policy-makers and specifically for the UK setting, to evaluate with formal modelling and MI casualty data, whether the time-frames employed by Lerner's P1 criteria are appropriate for use in the UK MI setting.

Regarding future work in the field of major incident triage, there is a constant need to review and refine triage tools in line with changing population demographics and evolving threats resulting in variation in the types of major incidents which may arise, in line with government's National Risk Register[3]. Furthermore, It is not yet known which existing tool would perform best in a burns or CBRN-related major incident. TARN does not specifically include burns patients unless they have concurrent injuries, hence this renders TARN an ineffective database with which to make this enquiry[21]. Further work should include measurement of triage tools' performance using a burns-specific database, such as the UK National Burn Injury Database[31]. Such work could, and should, utilise Lerner's definitions of major incident categories in order to standardise practice and allow the direct comparison of results with those pertaining to other types of major incidents (Chapter 2). Lerner's criteria include burns-specific procedures as part of triage category definitions, including assignment of P1 to patients needing escharotomy and assignment of the Expectant category to patients with 90% burns, as well as covering more general aspects such as the need for airway interventions and volume resuscitation[32]. As mentioned previously, comparison of Quick Triage's practical applicability to that of NHS MITT is an important consideration that is directly relevant to informing national policy: a simulation study comparing the two tools is an urgent research priority (Chapter 4). Another area in which further work is urgently required is to develop trauma triage tools specific to elders, including practical incorporation of biological rather than chronological age and potentially involving frailty scoring. Biomarkers, ideally those that utilise saliva (non-invasive measurement), may have a role in future triage, particularly if able to detect occult shock and traumatic brain injury[33]. Lastly, further research is required to establish the practical role of technology to augment the triage process, such as the use of wearable sensors to take real-time physiological measurements and track patient movements[34]. Whilst several potential techniques have been described theoretically, implementation studies to prove their safety and efficacy in the practical setting would prove beneficial[34].

In conclusion, this final chapter draws together the key findings of this thesis, identifying the under-performance of the current UK adult and paediatric major incident triage tools when evaluated in a nationally representative sample of patients, with a recommendation to change to the BCD Triage Sieve. This has been incorporated into national policy as the NHS MITT4 from April 2024, superseding the underperforming NARU Triage Sieve, Paediatric Triage Tape and JumpSTART tools. Triage category definitions according to Lerner's criteria have been validated using national trauma registry data and the resulting resource requirements are shown to align well with current casualty distribution plans in the UK. These definitions are recommended as a gold standard for future UK and international major incidents; standardisation will facilitate advances in the field by allowing direct comparison between studies. A simple, qualitative novel machine learning-derived triage tool, Quick Triage, has been developed using TARN data. Internal validation results suggest a higher sensitivity and lower over-triage rates across clinically important subgroups (age and injury mechanism) compared with MITT. This presents the possibility of a single tool for all age groups, for use by all emergency first responders in UK major incidents. Future work should involve external validation of Quick Triage, particularly amongst non-NHS first responders, prior to consideration for more widespread use. A significant limitation of computed, retrospective application of tools to trauma registry data is that it fails to encompass human factors and the speed and reliability with which tools can be applied; practical simulation studies are required to compare Quick Triage to MITT and TST, in order to fully inform practice and policymaking.

References

1. Moran CG, Webb C, Brohi K, Smith M, Willett K. Lessons in planning from mass casualty events in UK. Bmj 2017; 359: j4765.

2. 2017 Report of the Public Inquiry into the Attack on Manchester Arena on 22nd May. Manchester Arena Inquiry Volume 2: Emergency Response, 2022.

3. HM Government. National Risk Register 2023 edition. 2023.

4. NHS England. Major incident triage tools. In: NHS Emergency Preparedness RaR, editor. NHS England; 2023.

5. Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. Prehospital and disaster medicine 2008; 23(1): 3-8.

6. Vassallo J, Beavis J, Smith JE, Wallis LA. Major incident triage: Derivation and comparative analysis of the Modified Physiological Triage Tool (MPTT). Injury 2017; 48(5): 992-9.

7. Vassallo J, Smith J. 2 Major incident triage and the implementation of a new triage tool, the MPTT-24. Emergency Medicine Journal 2017; 34(12): A860-A1.

8. Hodgetts TJ, Hanlan CG, Newey CG. Battlefield First Aid: a simple, systematic approach for every soldier. J R Army Med Corps 1999; 145(2): 55-9.

9. T J Hodgetts JH, I Maconochie, C Smart. Paediatric triage tape. Pre-hospital Immediate Care 1998; 2: 155-9.

10. Sacco WJ, Navin DM, Fiedler KE, Waddell RK, 2nd, Long WB, Buckman RF, Jr. Precise formulation and evidence-based application of resource-constrained triage. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2005; 12(8): 759-70.

11. Asaeda G. The day that the START triage system came to a STOP: observations from the World Trade Center disaster. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2002; 9(3): 255-6.

12. Aylwin CJ, König TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. The Lancet 2006; 368(9554): 2219-25.

13. Challen K, Walter D. Major incident triage: comparative validation using data from 7th July bombings. Injury 2013; 44(5): 629-33.

14. Wallis LA, Carley S. Comparison of paediatric major incident primary triage tools. Emergency Medicine Journal 2006; 23(6): 475.

15. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the Accuracy of Mass Casualty Triage Systems When Used in an Adult Population. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2019: 1-8.

16. Cross KP, Cicero MX. Head-to-Head Comparison of Disaster Triage Methods in Pediatric, Adult, and Geriatric Patients. Annals of emergency medicine 2013; 61(6): 668-76.e7.

17. Jerome D, Savage DW, Pietrosanu M. An assessment of mass casualty triage systems using the Alberta trauma registry. Cjem 2023; 25(8): 659-66.

18. Vassallo J, Smith J, Bouamra O, Lecky F, Wallis LA. The civilian validation of the Modified Physiological Triage Tool (MPTT): an evidence-based approach to primary major incident triage. Emergency Medicine Journal 2017; 34(12): 810.

19. Malik NS, Chernbumroong S, Xu Y, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. eClinicalMedicine 2021; 36.

20. Malik NS, Chernbumroong S, Xu Y, et al. Paediatric major incident triage: UK military tool offers best performance in predicting the need for time-critical major surgical and resuscitative intervention. EClinicalMedicine 2021; 40.

21. The University of Manchester. The Trauma Audit and Research Network. Continuous. www.tarn.ac.uk.

22. Cone DC, Serra J, Burns K, MacMillan DS, Kurland L, Van Gelder C. Pilot test of the SALT mass casualty triage system. Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors 2009; 13(4): 536-40.

23. Jain TN, Ragazzoni L, Stryhn H, Stratton SJ, Della Corte F. Comparison of the Sacco Triage Method Versus START Triage Using a Virtual Reality Scenario in Advance Care Paramedic Students. CJEM 2016; 18(4): 288-92.

24. Joint Emergency Service Interoperability Programme. JESIP Joint Doctrine: The Interoperability Framework Edition 3. 2022. https://www.jesip.org.uk/news/joint-doctrine-edition-3/.

25. Guideline N, E. Brooke Lerner DCC, Eric S. Weinstein, Richard B. Schwartz, Phillip L. Coule, Michael Cronin, Ian S. Wedmore, Eileen M. Bulger, Deborah Ann Mulligan, Raymond E. Swienton, Scott M. Sasser, Umair A. Shah, Leonard J. Weireter, Teri L. Sanddal, Julio Lairet, David Markenson, Lou Romig, Gregg Lord, Jeffrey Salomone, Robert O'Connor and Richard C. Hunt. Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. Disaster medicine and public health preparedness 2011; 5: 129-37.

26. Malik NS, Chernbumroong S, Xu Y, et al. The BCD Triage Sieve outperforms all existing major incident triage tools: Comparative analysis using the UK national trauma registry population. EClinicalMedicine.

27. Baxt WG, Upenieks V. The lack of full correlation between the Injury Severity Score and the resource needs of injured patients. Annals of emergency medicine 1990; 19(12): 1396-400.

28. National Ambulance Resilience Unit (NARU). Clinical Guidance: Medical Support Minimal Requirements for a Mass Casualty Incident. www.naru.org.uk; 2014.

29. Debacker M HI, Dhondt E, Rockenschaub G, Rüter A, Codreanu T, Koenig KL, Schultz C, Peleg K, Halpern P, Stratton S, Della Corte F, Delooz H, Ingrassia PL, Colombo D, Castrèn M. Utstein-Style Template for Uniform Data Reporting of Acute Medical Response in Disasters. PLOS Currents Disasters 2012; 1.

30. Olding J, Olding C, Bew D, Fan K. Penetrating head & neck trauma – Epidemiology and injury characteristics in terror-related violence, interpersonal violence and deliberate self-harm at a level 1 trauma centre. The Surgeon 2019.

31. Association BB. National Burn Injury Database. Continuous.

32. Lerner EB, McKee CH, Cady CE, et al. 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 2015; 19(2): 267-71.

33. Toman E, Harrisson S, Belli T. Biomarkers in traumatic brain injury: a review. J R Army Med Corps 2016; 162(2): 103-8.

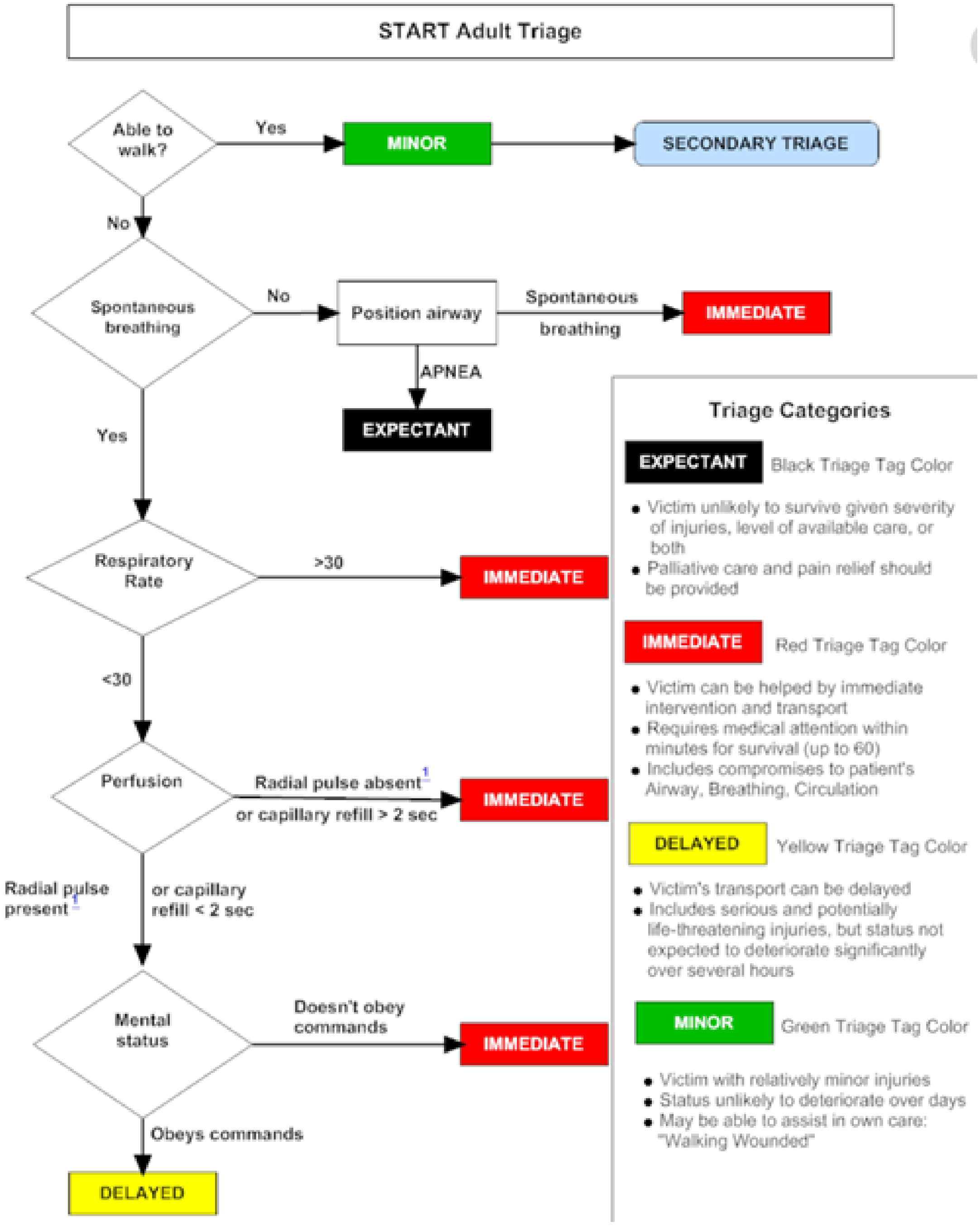
34. Ozella L, Gauvin L, Carenzo L, et al. Wearable Proximity Sensors for Monitoring a Mass Casualty Incident Exercise: Feasibility Study. Journal of medical Internet research 2019; 21(4): e12251.

APPENDIX: EXISTING MAJOR INCIDENT TRIAGE TOOLS

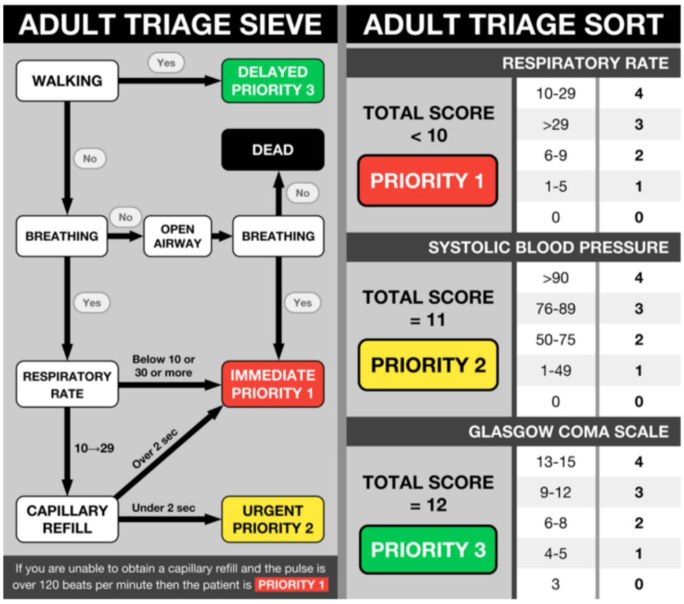
EXISTING MAJOR INCIDENT TRIAGE TOOLS

This appendix includes, in chronological order of introduction, the images of various international major incident triage tools included (and fully referenced) throughout this thesis.

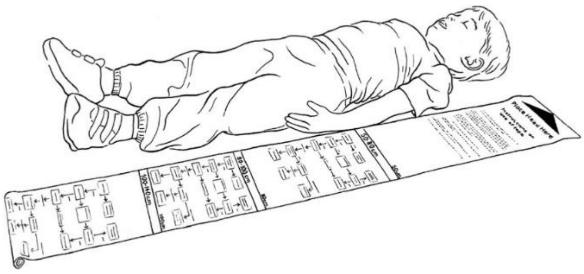
1. Simple triage and Rapid Treatment (START) Triage (United States, 1983):



2. The MIMMS Adult Triage Sieve and Sort (UK, 1998)

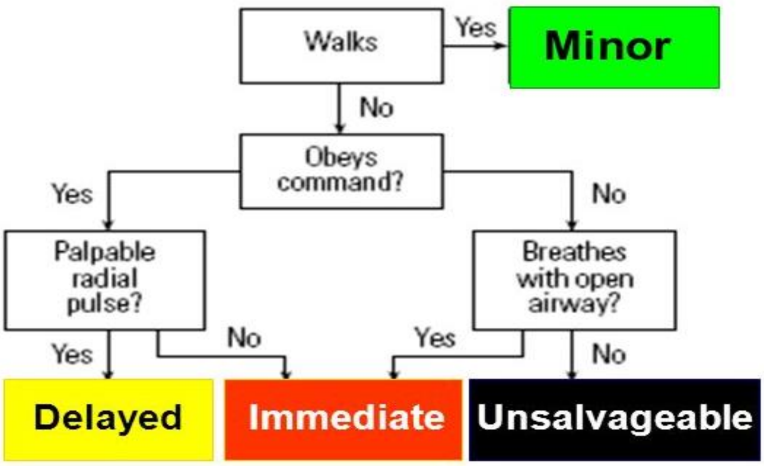


3. Paediatric Triage Tape (PTT) (UK, 1998)

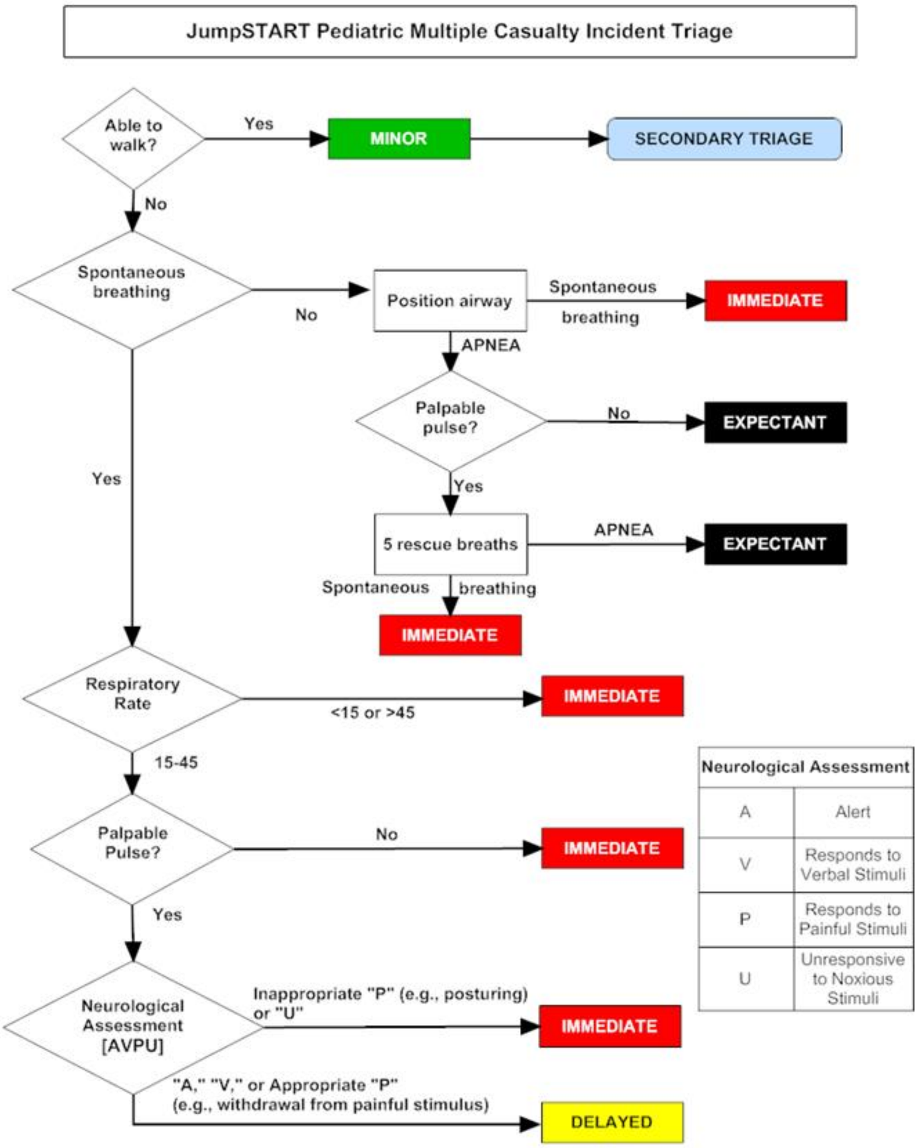


Ledger: The tape is placed next to the child from the head end. The child is then triaged using the algorithm next to the child's feet.

4. CareFlight (Australia, 2001)



5. JumpSTART (United States, 2001)



Use JumpSTART if the Patient appears to be a child.

Use an adult system, such as START, if the patient appears to be a young adult.

Triage Categories

EXPECTANT Black Triage Tag Color

- Victim unlikely to survive given severity of injuries, level of available care, or both
- Palliative care and pain relief should be provided

DELAYED Yellow Triage Tag Color

- Victim's transport can be delayed
- Includes serious and potentially life-threatening injuries, but status not expected to deteriorate significantly over several hours

IMMEDIATE Red Triage Tag Color

- Victim can be helped by immediate intervention and transport
- Requires medical attention within minutes for survival (up to 60)
- Includes compromises to patient's Airway, Breathing, Circulation

MINOR Green Triage Tag Color

- Victim with relatively minor injuries
- Status unlikely to deteriorate over days
- May be able to assist in own care: "Walking Wounded"

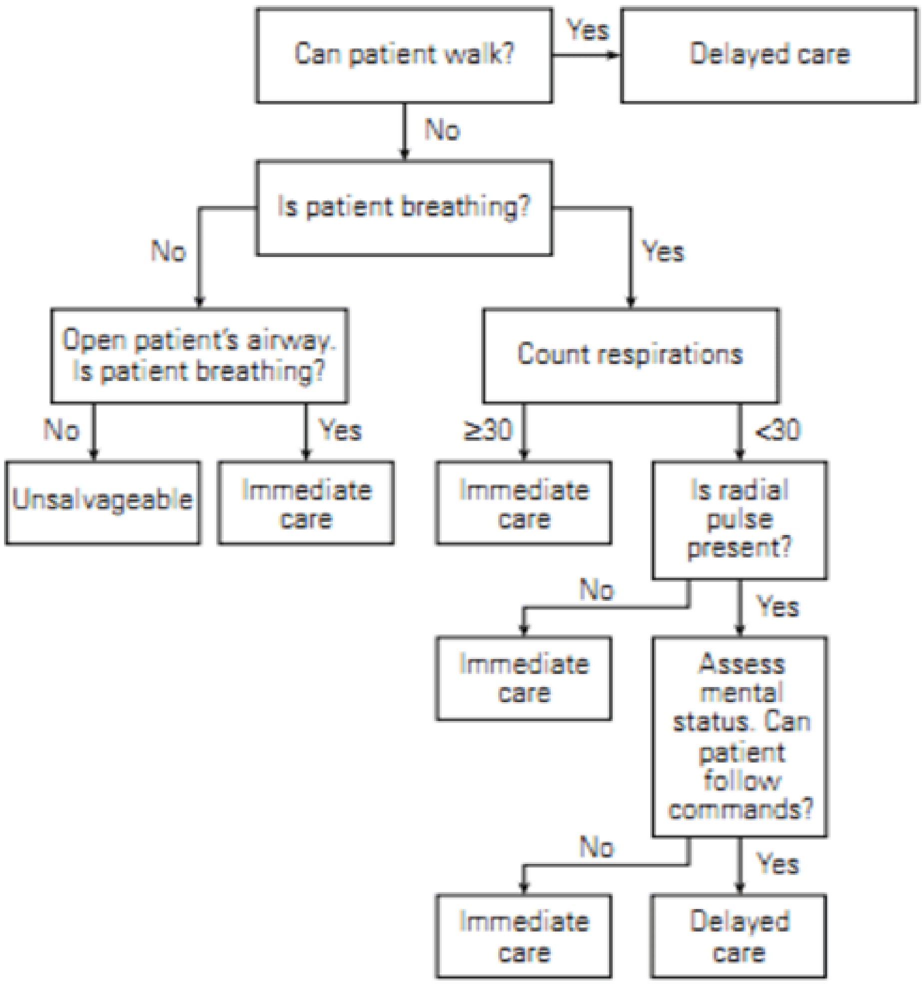
6. **Sacco Triage Method (United States, 2005)**

ThinkSharp **Sacco Score** R + P + M +/- A

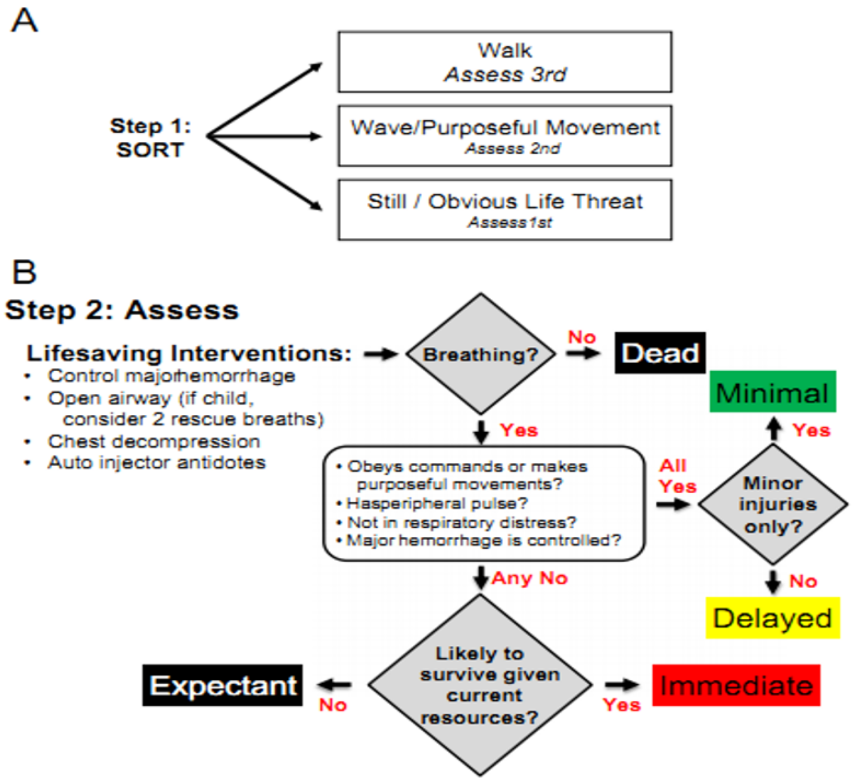
1 min (60 seconds)	0	1	2	3	4
R	0	1-9*	36+	25-35	10*-24
P	0	1-40	41-60*	121+	61*-120
M	No Response	Extension/ Flexion	Withdraws	Localizes	Obeys Commands
Adjustment: Age:	0-7 +2	8-14 +1	15-54 0	55-74 -2	75+ -3

* - measure must be verified

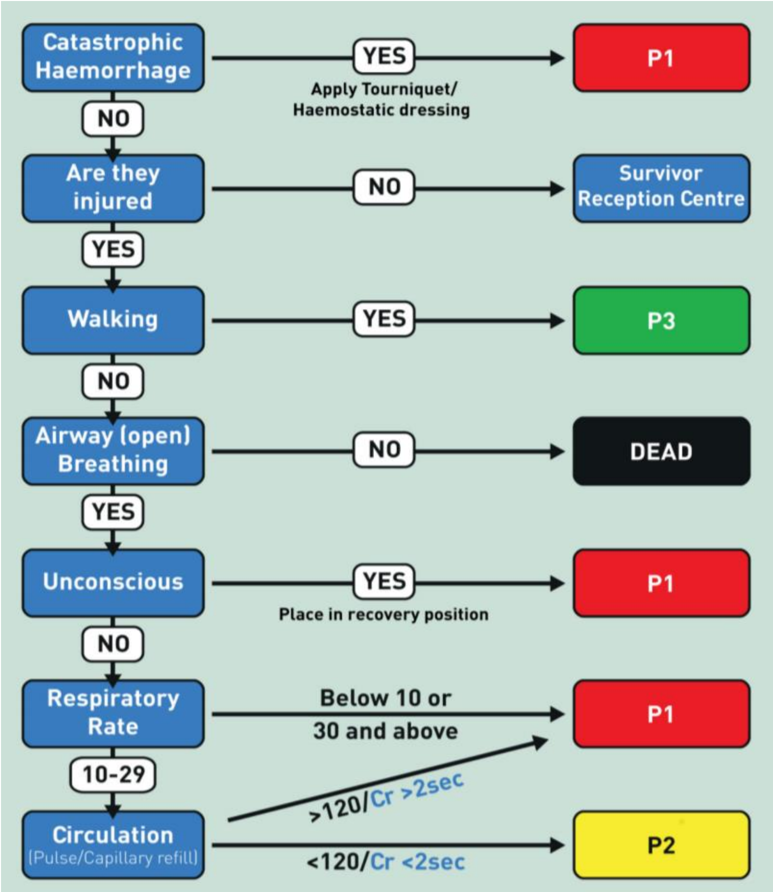
7. **The Modified START (MSTART) Triage (United States, 2006):**



8. **Sort, Assess, life-saving intervention, Treatment/Transport (SALT) (United States, 2008)**

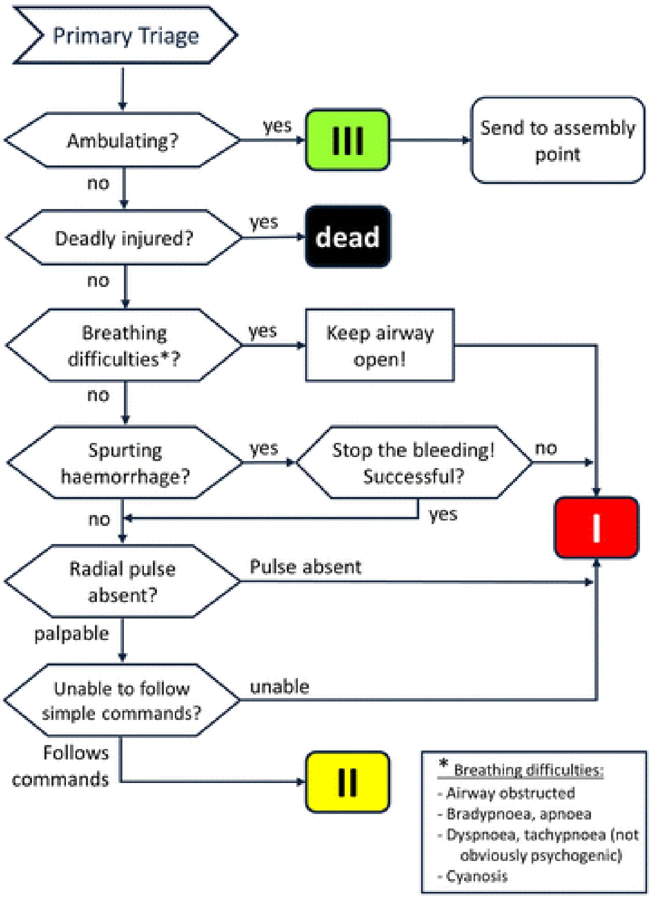


9. **The National Ambulance Resilience Unit (NARU) Triage Sieve (UK, 2013):**

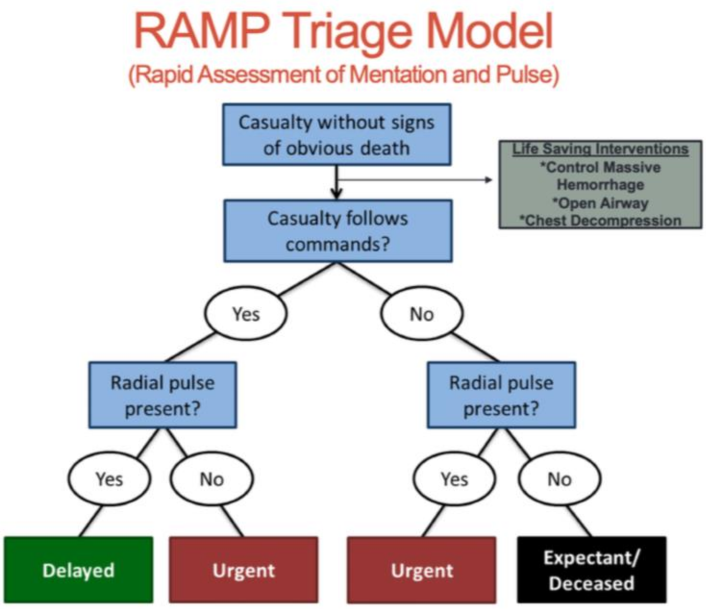


Catastrophic Haemorrhage
YES
P1
Apply Tourniquet/
Haemostatic dressing
NO
Are they injured
NO
Survivor Reception Centre
YES
Walking
YES
P3
NO
Airway (open) Breathing
NO
DEAD
YES
Unconscious
YES
P1
Place in recovery position
NO
Respiratory Rate
Below 10 or
30 and above
P1
10-29
>120/Cr >2sec
Circulation
(Pulse/Capillary refill)
<120/Cr <2sec
P2

10. The Amberg-Schwandorf Algorithm for Primary Triage (ASAV) (Germany, 2013)

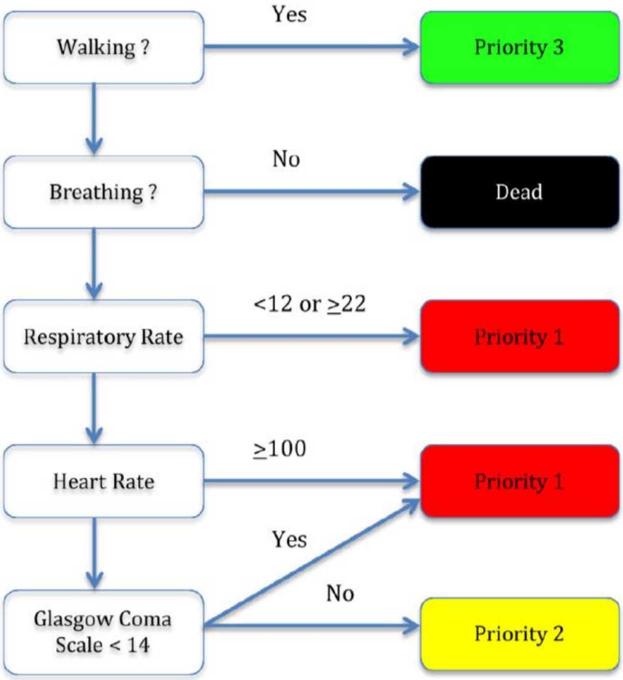


11. Rapid Assessment of Mentation and Pulse (RAMP) (United States, 2016):

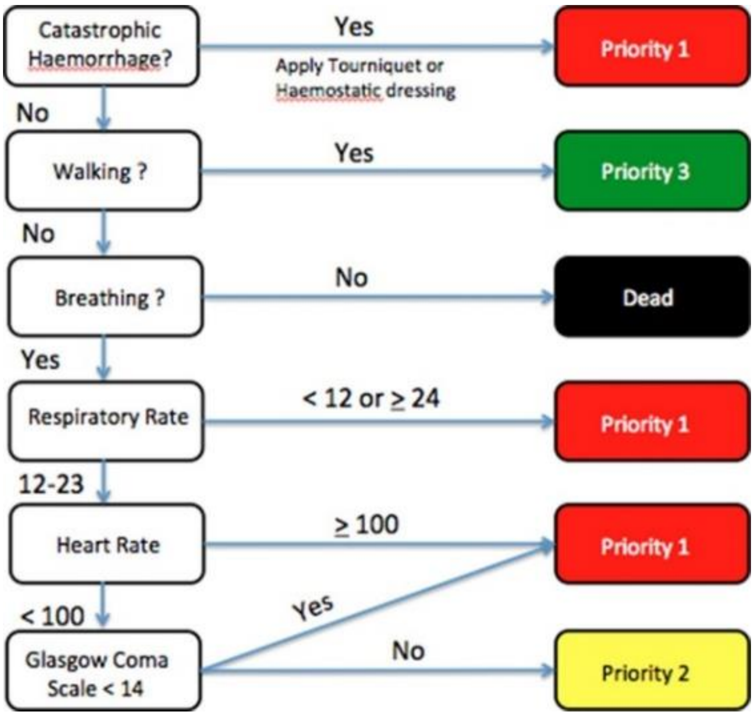


12. The Modified Physiological Triage Tool (MPTT) and MPTT-24 (UK, 2017-2018)

MPTT (Vassallo, Injury 2017):

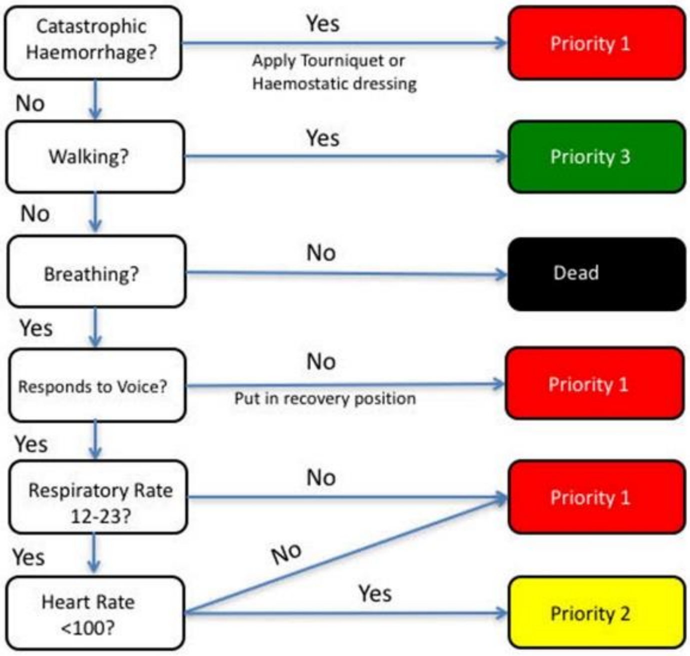


MPTT-24 (Vassallo, EMJ 2017):



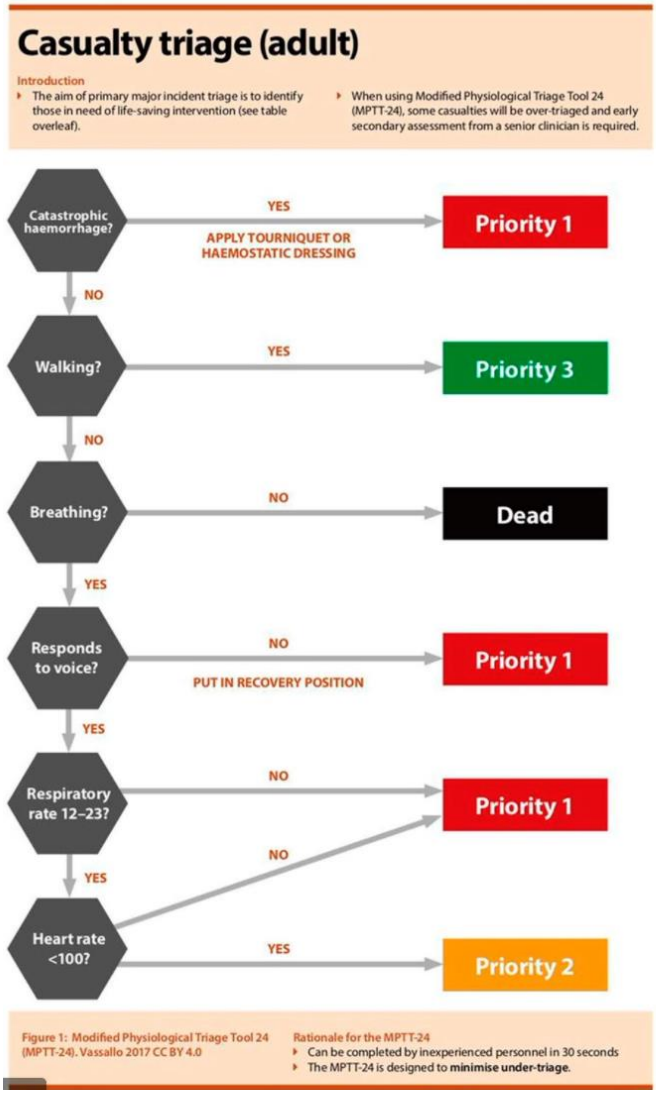
Ledger: The use of "catastrophic haemorrhage?" has been incorporated as the opening step and the upper respiratory rate threshold has been revised to 24 (from 22 in the MPTT). This version was first published alongside its external validation, hence it has been regarded as the definitive version for the purpose of this thesis. Note that GCS is used to assess mental status, and this features as the last step in the tool.

The MPTT-27 (Vassallo, J RAMC 2017):



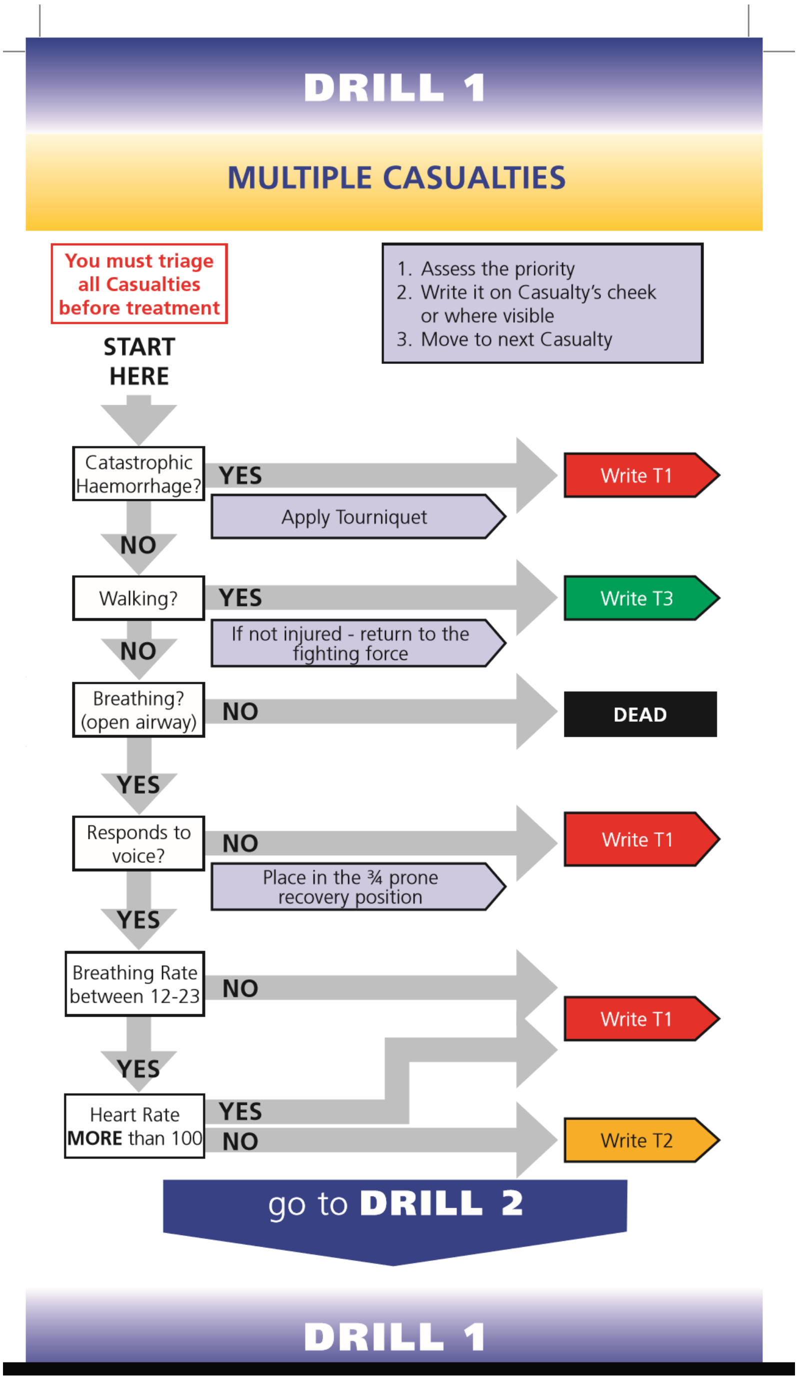
Ledger: In this version, mental status is assessed using the AVPU component "Responds to voice?" and its relative position within the tool has moved up, to precede the respiratory rate and heart rate estimations.

MPTT-24 (NHS Clinical guidelines for major incidents and mass casualty events) (UK, 2018)

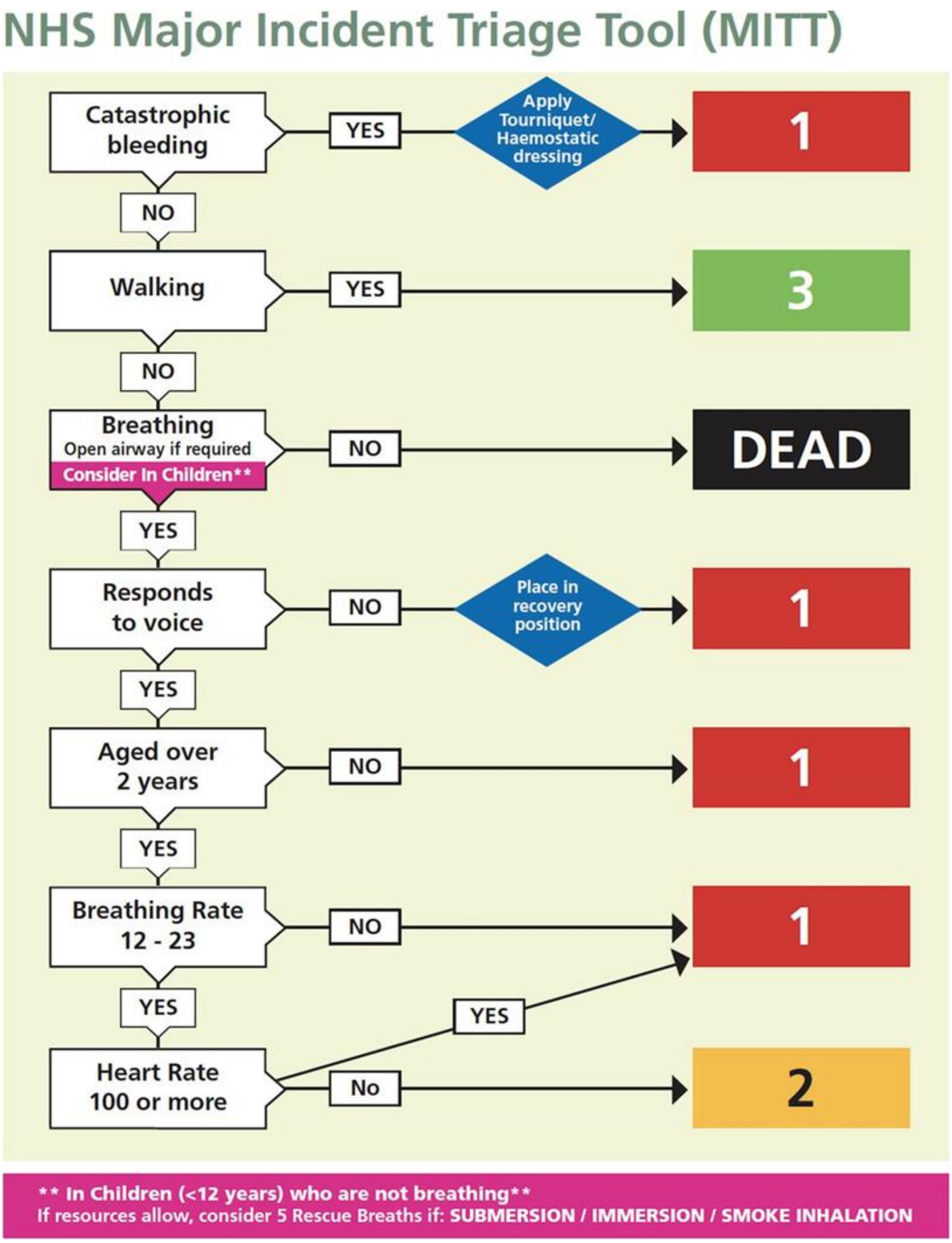


Ledger: This version appears to be identical to the BattleField Casualty Drills Triage Sieve (2018), with the exception that it does not include an airway intervention in the case of a non-breathing patient.

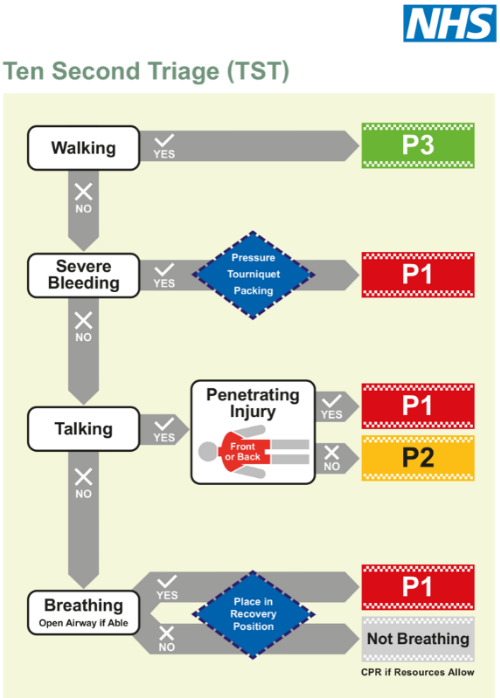
13. The Battlefield Casualty Drills Triage Sieve (UK, 2018)



14. NHS Major Incident Triage Tool (MITT) (UK, 2022)



15. Ten Second Triage (TST) (UK, 2022)



16. Quick Triage (UK, 2023)