

AWARD NUMBER: W81XWH-19-2-0047

TITLE: Exploitation of Bayesian Networks for Clinical Decision Support on the Battlefield

PRINCIPAL INVESTIGATOR: Col Nigel Tai

CONTRACTING ORGANIZATION: Queen Mary University London

REPORT DATE: OCTOBER 2023

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE
OCTOBER 2023

2. REPORT TYPE
Annual Report

3. DATES COVERED
15Sept2022 - 14SEPT2023

4. TITLE AND SUBTITLE
Exploitation of Bayesian Networks for Clinical Decision Support on the Battlefield

5a. CONTRACT NUMBER

W81XWH-19-2-0047

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR
Colonel Nigel RM Tai

E-Mail: nigel.tai@nhs.net

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AND ADDRESS(ES)

Queen Mary University London
327 Mile End Road London
E1 4NS
UK

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command

Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Purpose: The purpose of this research is to optimize the care of battlefield trauma patients through the development of Bayesian-network (BN) machine learning-powered clinical decision-support (CDS) tools. **Scope:** The scope of the research encompasses the refinement of existing BNs, and development and prototyping of new BNs designed for pre-hospital, en-route, and deployed healthcare facility stages of care, such that CDS prototypes are available for piloting and assessment in future, real-world clinical studies. **Major Findings - Year 4 Data.** Transfer of US DoD data to Barts Health/QMUL Data Safe Haven Environment - achieved with Data Sharing Agreement in place. **Model generation.** 1. Transfusion prediction model novel structure developed, reflecting non-compressible versus compressible haemorrhage. Novel outcome variables (predicted outcome of between 1 and 21+ units at incremental time segments of 0-2, 2-6, 6-12 and 12-24hrs) featured. Parameter learning completed. Trained on 995 patients. Probabilistic confusion matrix constructed to assess model performance as traditional performance metrics (AUROC) unsuited. Model performance assessed as Excellent or good in 81.4% of predictions. External validation awaited. 2. Survival-MIL model developed with TBI fragment AUROC at 0.93 (0.92-0.95). Final model structure (TAKI, TBI, TIC) developed. Overall performance of initial model AUROC 0.81 (0.77-0.85) - model refinement in progress. **Software + system engineering.** Back-end, front-end and manager API test suites are completed using relevant test packages for python and JavaScript languages. The server development is ongoing: handling continuous variables in the prediction models (front-end adjustments for data input and back-end adjustments for model inference), explanation (integrating the explanation generation capabilities to the BNDS server), and uncertainty (dealing with uncertainty fragments in our developed models with the correct data input behaviour as tested in the prototype usability study), and incorporating feedback from the prototype usability study to improve UI/UX. The development of the BN-explanation environment is ongoing. **Conclusion** The project is proceeding toward its endpoint of delivering a suite of validated prototype clinical decision support tools that will enhance the care of injured service men and women.

15. SUBJECT TERMS Trauma, Prediction, Coagulopathy, Mortality, Limb Salvage, Transfusion, Machine Learning, Clinical Decision Support, Military Surgery

16. SECURITY CLASSIFICATION OF"**a. REPORT**

U

b. ABSTRACT

U

c. THIS PAGE

U

**17. LIMITATION
OF ABSTRACT**

UU

**18. NUMBER
OF PAGES**

43

**19a. NAME OF
RESPONSIBLE PERSON**
USAMRDC**19b. TELEPHONE NUMBER**
(include area code)

<u>TABLE OF CONTENTS</u>	<u>Page</u>
1. Introduction	5
2. Keywords	5
3. Accomplishments	5
4. Impact	6
5. Changes/Problems	8
6. Products	9
7. Participants & Other Collaborating Organizations	12
8. Special Reporting Requirements	13
9. TECHNICAL NARRATIVE	14

INTRODUCTION:

The research subject concerns optimizing the care of battlefield trauma patients through the purposeful development and validation of accurate, clinically credible prognostic tools that will support informed and personalized treatment decisions by quantifying clinically relevant risks. The scope of the research encompasses the development and prototyping of tools designed for pre-hospital, en-route, and deployed healthcare facility stages of care, such that these prototypes are available for piloting and assessment in a clinical context.

1. KEYWORDS:

Trauma, Prediction, Coagulopathy, Mortality, Limb Salvage, Transfusion, Machine Learning, Clinical Decision Support, Military Surgery

2. ACCOMPLISHMENTS:

The scope of this project is to use machine-learning techniques established by the funded research group (COMputer Battlefield Assistance in Trauma care and Injury Decision-support: COMBAT-AID) to enable military trauma clinicians to provide precision-medicine on the battlefield. The date of Project Start was 15 September 2019.

- **Data.** All remaining Data Sets acquired (US Joint Theatre System data required for model validation obtained and transferred to Barts Health/ QMUL Safe Haven data).
- **Model generation 1.** The transfusion prediction model has been developed and found to perform well. The final training/test UK dataset includes 995 patients (547 of which have received at least one unit of blood transfusion within 24 hours from arrival). A novel model structure, separating compressible from non-compressible haemorrhage, was developed, incorporating junctional haemorrhage. Novel outcome variables (predicted outcome of between 1 and 21+ units at incremental time segments of 0-2,2-6,6-12 and 12-24hrs) featured. Parameter learning was completed and the model was trained on 995 patients. A probabilistic confusion matrix was constructed to assess model performance as traditional performance metrics (AUROC) were deemed unsuited to this. Model performance was duly assessed as Excellent or Good in 81.4% of predictions. External validation is awaited from newly received US data.
- **Model development 2.** The Survival-MIL model was developed to include a TBI fragment with performance AUROC at 0.93 (0.92-0.95). The TBI model was integrated with the TAKI and TIC models to provide a single architecture. Overall performance of the comprehensive model was moderate AUROC 0.81 (0.77-0.85) – model refinement in progress with expectation of improved performance.
- **Explanation.** Back-end, front-end and manager API test suites are completed using relevant test packages for python and JavaScript languages. The server development is ongoing: handling continuous variables in the prediction models (front-end adjustments for data input and back-end adjustments for model inference), explanation (integrating the explanation generation capabilities to the BNDS server), and uncertainty (dealing with uncertainty fragments in our developed models with the correct data input behaviour as tested in the prototype usability study), and incorporating feedback from the prototype usability study to improve UI/UX. The development of the BN-explanation environment is ongoing
- **End user testing.** Determined the most effective methods and metrics to evaluate the usability of CDS systems for medical emergencies. Optimized graphical displays of probability distributions for better clinician interpretation through usability studies with trauma clinicians. Completion of a prospective usability evaluation of the AI Risk Prediction System (UK Health Research Authority/Research Ethics Committee approval reference 22/HRA/2324). During this study, the system underwent iterative modifications, informed by the evidence generated from end-user feedback, to optimize system performance. Trauma clinicians were also asked about the potential barriers to adoption of an AI clinical decision support system in trauma care.

What opportunities for training and professional development has the project provided?

Three PhD students are supervised by Col N Tai, Prof K Brohi, Dr William Marsh, Mr Zane Perkins, and Dr Evangelina Kyrimi. All 3 students are registered to Queen Mary University London. Supervisors and students enrol in QMUL programmes to foster their continuing development https://www.qmul.ac.uk/doctorscollege/?page_id=49. These opportunities are a requirement for satisfactory study and supervision.

How were the results disseminated to communities of interest?

Via presentations at national and international academic fora and publication in relevant journals.

What do you plan to do during the next reporting period to accomplish the goals?

1. Complete prototype interface linked to all developed models.
2. Complete mortality model development and external validation on US Joint Theatre Trauma System data.
3. Complete blood transfusion prediction model.
4. Complete Overflight model.

3. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Artificial intelligence and Machine Learning are part of the modern discourse when considering advances in improved medical care, yet many approaches lack the ability to integrate patient data, clinical expertise and evidence-derived knowledge in deriving prognoses that can support decision-making. Published results from our foundational work, contributing to the projects goals, confirm that accurate, Bayesian Network-powered models can be produced that, by incorporating expert knowledge and data-sets, out-perform standard prediction tools with regard to the prediction of traumatic coagulopathy and risk of limb loss in trauma. However, generation of accurate predictions means little if clinicians are reluctant to use the tool, if they do not trust the result or both. We have advanced the discipline by developing a user-configurable interface powered by a 6 level system architecture that marries inputs, computation, outputs and evidence. We have also learnt how the ability of a tool to explain its reasoning in making a prediction is or is not valued by clinicians, and how this influences trust in the model. We have used an novel approach to predicting mortality that incorporates dynamic variables, accounting for the effect of interventions, which provides predictions tailored to the stage of care across a combat area – one that outperforms established mortality prediction tools. And we have made significant in-roads in to the development of a system that predicts likely consumption of blood transfusion resource, based on what the individual patient requires rather than what the clinician is likely to prescribe.

What was the impact on other disciplines?

The research outputs will contribute to the development of clinical decision support tools applicable to non-trauma, non-military patients. The modelling techniques developed are impactful for other areas of human expert decision-making where support is required, as the approaches used are applicable in Law, Engineering, Defence and the Financial sector.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nil to report

Actual or anticipated problems or delays and actions or plans to resolve them

Year 3 Staffing issues:

1. Tasks assigned to Computer Science Post-Doctoral staff were remapped to supervisors on account of the Post-Doctoral staff gaining another position within the university (April 2023 onwards)
2. One staff member returned from parental leave.
3. A second no-cost project extension of six months (bringing the end of project date to March 2024) was granted to mitigate progression issues.

Changes that had a significant impact on expenditures

Nil Significant to report. Expenditure for 2022-2023 as below.

	2022	2023	2023	2023	
	Oct-Dec	Jan-March	Apr-June	July-Sept	
Staff costs	\$91,400	\$55,319	\$40,586	\$26,906	
Stipend	\$5,688	\$5,688	\$5,688	\$5,688	
PhD fees	\$0	\$0	\$0	\$0	
Other non-staff costs	\$0	\$265	\$461	\$1,216	
Overheads	\$13,737	\$13,737	\$13,737	\$13,737	
Subcontracting costs	\$0	\$0	\$0	\$0	
TOTAL	\$110,826	\$75,009	\$60,472	\$47,547	\$293,853
					TOTAL

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nil Significant to report.

Significant changes in use or care of vertebrate animals

Nil Significant to report.

Significant changes in use of biohazards and/or select agents

Nil Significant to report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Papers

Accepted by Peer-Reviewed Journals:

- 1 Durrands T, Murphy M, Wohlgemut JM, De'Ath HD, Perkins ZB. Diagnostic accuracy for identification of life-threatening torso injuries: A Meta-analysis. BJS 2023 (accepted)
- 2 Marsden MER, Kellett S, Bagga R, Wohlgemut JM, Lyon R, Perkins ZB, Gillies K, Tai N. Understanding pre-hospital blood transfusion decision-making for injured patients: an interview study. Emergency Medical Journal (accepted 01/08/2023)
- 3 Wohlgemut JM, Pisirir E, Kyrimi E, Stoner RS, Marsh W, Perkins NRM, Tai NRM. Methods Used to Evaluate Usability of Mobile Clinical Decision Support Systems for Healthcare Emergencies: A Systematic Review and Qualitative Synthesis. JAMIA Open 2023 (accepted 26/6/2023)
- 4 Pisirir E, Wohlgemut JM, Kyrimi E, Stoner RS, Perkins ZB, Tai NRM, Marsh DWR. A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models. IEEE Xplore 2023. (accepted March 2023)
- 5 Wohlgemut JM, Marsden MER, Stoner RS, Pisirir E, Kyrimi E, Grier G, Christian M, Hurst T, Marsh W, Tai NRM, Perkins ZB. Diagnostic Accuracy of Clinical Examination to Identify Life- and Limb-Threatening Injuries in Trauma Patients. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 2023;31:18.

Submitted to Peer-Reviewed Journals:

1. Kyrimi et al, "A Scoping Review, Novel Taxonomy and Catalogue of Implementation Frameworks for Clinical Decision Support Systems" BMC Medical Informatics and Decision Making, submitted 2023.
2. Wohlgemut et al, "Identification of Major Hemorrhage in Trauma Patients in the Pre-Hospital Setting: Diagnostic Accuracy and Impact on Outcome" Trauma Surgery Acute Care Open, submitted 14/7/23
3. Kyrimi et al, "Updating and recalibrating causal probabilistic models on a new target population" Journal of Biomedical Informatics, resubmitted 6/7/23
4. Kyrimi et al, "Counterfactual reasoning using Bayesian networks as a healthcare governance tool" Artificial Intelligence in Medicine, submitted 5/6/23
5. Tandle et al, "Enhancing the Clinical Relevance of Haemorrhage Prediction Models in Trauma" Military Medical Research, submitted 2023.

Books or other non-periodical, one-time publications.

Other publications, conference papers and presentations.

International **Oral** Presentations:

- June 26-29th 2023 **A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models.** Pisirir E, **Wohlgemut JM**, Kyrimi E, Stoner RS, Perkins ZB, Tai NRM, Marsh DWR. The 11th IEEE International Conference on Healthcare Informatics, Houston, TX, USA
- May 9th 2023 **Usability of an Artificial Intelligence Clinical Decision Support System for Trauma: A Pilot Study (Wohlgemut JM, Pisirir E, Kyrimi E, Stoner RS, Marsh W, Perkins ZB, Tai NRM)** The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia
- May 8th 2023 **Diagnostic accuracy of Pre-hospital Clinical Assessment of Major Traumatic Haemorrhage (Wohlgemut JM, Pisirir E, Stoner RS, Kyrimi E, Grier G, Christian M, Marsh W, Perkins ZB, Tai NRM)** The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia
- Dec 8th 2022 **Critical decision analysis of pre-hospital major haemorrhage protocol activation (Wohlgemut JM, McMullen H, Marsden M, Pisirir E, Stoner RS, Kyrimi E, Marsh W, Perkins ZB, Tai NRM)** The London Trauma Conference 2022, London, UK

International **Poster** Presentations:

- May 17-19th 2023 **Critical decision analysis of pre-hospital major haemorrhage protocol activation (Wohlgemut JM, McMullen H, Marsden M, Pisirir E, Stoner RS, Kyrimi E, Marsh W, Perkins ZB, Tai NRM)** Association of Surgeons of Great Britain & Ireland (ASGBI) 2023 Congress, Harrogate, UK
- Sep 21-24th 2022 **Machine Learning to Aid Decision Making in Patients who may Benefit from Prehospital Tranexamic Acid (Wohlgemut JM, Almuwallad AO, Kyrimi E, Rossetto A, Stoner RS, Cole E, Marsh W, Perkins ZB, Tai NRM, Davenport R)** American Association of Surgery for Trauma, 81st Annual Meeting, Chicago, Illinois, USA
- Sep 12-15th 2022 **Incorporating Clinician Accuracy and Uncertainty of Prehospital Injury Diagnosis Improves Clinical Validity of a Decision Support System Without Compromising Performance of a Bayesian Network to Predict Trauma-Induced Coagulopathy (Wohlgemut JM, Kyrimi E, Rossetto A, Stoner R, Pisirir E, Marsh W, Perkins Z, Tai N)** Military Health Services Research Symposium 2022, Kissimmee, Florida, USA

- **Website(s) or other Internet site(s)**

- **Technologies or techniques**

The algorithms are novel but the underlying technology is not. The methods to explain model outputs are novel (see appendix).

- **Inventions, patent applications, and/or licenses**

Nil significant to report

- **Other Products**

Nil significant to report

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Name	Principle Role	Project time	Contribution	Other funding support	Notes
Nigel Tai	PI	13.3%	Overall leadership functions; review meetings, hiring and staff establishment	MoD employee	Honorary QMUL and NHS contracts
William Marsh	CI	13.3%	Leadership esp wrt Computer Science aspects; review meetings and staff establishment.	QMUL employee	
Zane Perkins	Investigator	20%	Project supervision and academic management of Clinical PhD students; review meetings; hiring and staff establishment	NHS employee	Part funded by DoD grant
Evangelina Kyrimi	Post-Doctoral Research Assistant	Nil cost	Model development and supervision of Computer Science PhD Student; review meetings.	Nil	Informal support to work (ad-hoc)
Rebecca Stoner	QMUL PhD Candidate Clinical Fellow	100%	Model development & refinement	Nil	COVID 19 prevented earlier start
Jared Wohlgemut	QMUL PhD Candidate Clinical Fellow	100%	Clinical Decision Support system Interface development, assessment and workflow integration	Nil	COVID 19 prevented earlier start
Erhan Pisirir	QMUL PhD Candidate	100%	Modelling, Validation, Data Analysis		
Stavros Fotiadis	Operations manager Centre for Trauma Sciences QMUL	20%	Liaison with funder, QMUL Research Office, financial oversight and contractual management	QMUL employee supported by multiple C4TS awards	
Prof Karim Brohi	CI	2.67%	Critical analysis, leadership, authority for Centre 4 trauma science data resources.	Full time QMUL academic	
Rachel Sumner	QMUL Comp Science RA	60% (has left project – funds to be redirected in to replacement hire)	Software and System Developer	Full time QMUL academic	

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nil significant to report

What other organizations were involved as partners?

- *Other.*

Nil significant to report

7. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: *N/A*

QUAD CHARTS: *N/A*

8. APPENDICES: *N/A*

ANNUAL REPORT TECHNICAL NARRATIVE W81XWH-19-2-0047 – YR 3

EXPLOITATION OF BAYESIAN NETWORKS FOR CLINICAL DECISION SUPPORT ON THE BATTLEFIELD

AUTHOR COL NIGEL TAI (PRINCIPLE INVESTIGATOR)

REPORT DATE 13 SEPT 2022 (SUBMITTED 12 OCT 2022)

DETAILED PROGRESS REPORT Year 4.

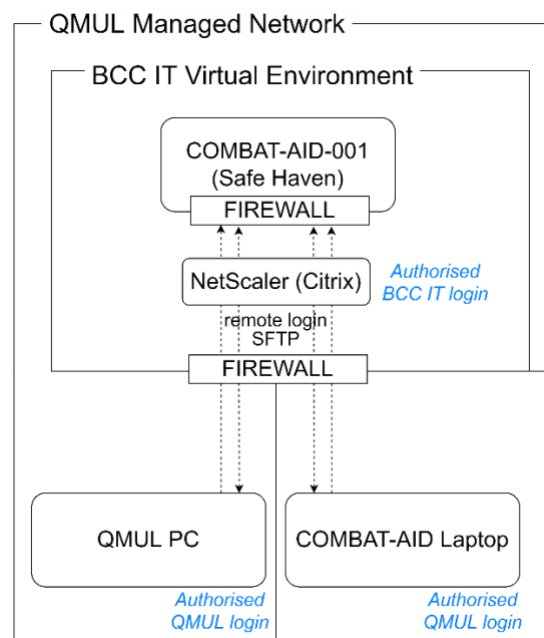
Specific Aim 1, Major Task 1: TIC-MIL development and internal validation

COMPLETED, as per previous reports

During this year, a manuscript was re-submitted (6/7/23) to Journal of Biomedical Informatics (Kyrimi et al 2023), entitled, “Updating and recalibrating causal probabilistic models on a new target population”. This article outlines the process of updating and recalibrating the TIC model originally designed in a civilian population, for the TIC-MIL model designed using a military population.

Specific Aim 1, Major Task 2: TIC-MIL external validation

The data safe haven named COMBAT-AID-001 is set up. COMBAT-AID-001 is a virtual machine behind a firewall, managed by Barts Cancer Centre IT, a safe environment disconnected from the world wide web for storing and processing of external validation data. The safe haven architecture is shown in Figure below.



The external validation data is an anonymised dataset for injury profile, physiology measurements, and lab test results for military patients injured in field operations, by USA Department of Defense Joint Trauma Registry (JTS DoDTR). The agreements for obtaining the de-identified data have been completed with the Defence Health Agency (DSA 23-3041) and the data has been received to store in COMBAT-AID-001 safe haven (August 2023). An

application has (Log # E00655.2a) been made to the US Office of Human and Animal Research Oversight (OHARO) (September 2023) for exploitation. Processing, cleaning, and preparation of the data and the external validation of TIC-MIL model is planned for the six month no-cost extension duration.

Specific Aim 2, Major Task 3: LIMB-MIL development and validation

COMPLETED, as per previous reports

Specific Aim 2, Major Task 4: LIMB-MIL external validation

COMPLETED, as per previous reports

Specific Aim 3: Prototype CDSS development

Software development: The BNDS architecture, the current version of the live server (based on QMUL servers <http://newton.eecs.qmul.ac.uk/>), and the prototype development for the usability study are described in the previous reports.

The back-end, front-end and manager API test suites are completed using relevant test packages for python and JavaScript languages. The server development is ongoing with the following features planned to be implemented: handling continuous variables in the prediction models (front-end adjustments for data input and back-end adjustments for model inference), explanation (integrating the explanation generation capabilities to the BNDS server), and uncertainty (dealing with uncertainty fragments in our developed models with the correct data input behaviour as tested in the prototype usability study), and incorporating feedback from the prototype usability study to improve UI/UX.

CDSS with Explanation Capabilities: The development of the BN-explanation environment is ongoing. The current (not yet completed) developments focus on the following categories:

Non-causal intermediate variables: The prediction explanation generation uses the concept of intermediate variables: these are not observed or measured as input to the model, however, they are the connecting links between inputs and the model prediction. Originally, all intermediate variables were assumed to play the same role in the generated explanation text. We discovered that there are intermediate variables which do not increase the probability of the outcome with their existence, but instead “explain away” the observations, thus reducing the probability of the outcome. This year’s work focuses on incorporating the capability to correctly generate causal and logical explanations with “explaining away intermediate variables” to the explanation generation algorithm.

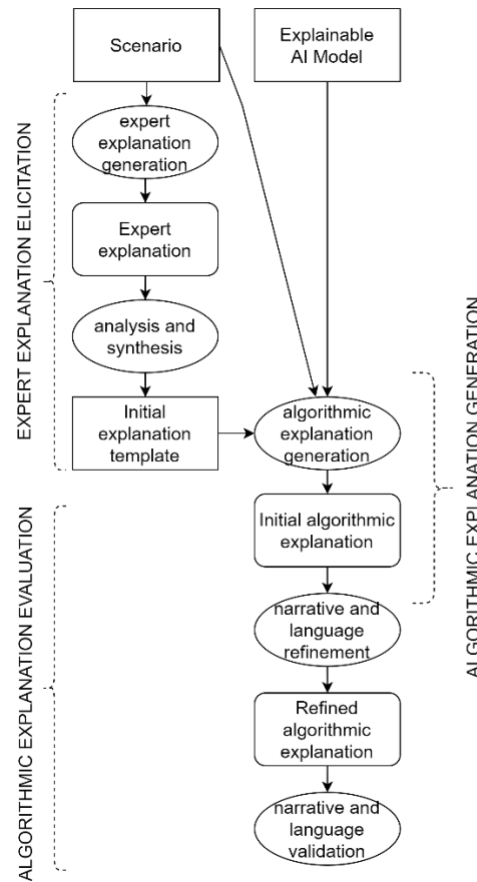
Continuous variables: The first version of the explanation generation environment required continuous variables in the prediction models to be discretised due to the limitations of BN inference libraries we use. This year’s work aims to incorporate a likelihood based inference method which can deal with non-discretised continuous variables, so that our explanation generation environment can take BNs including continuous variables without the need to discretise these nodes and potentially lose information.

A method for evaluating explanation narratives: In addition to the environment development, we developed and proposed an explanation template and narrative generation and evaluation process and used TIC-MIL model as a case example. See SA4 for details.

Specific Aim 4: Clinical usability evaluation of developed CDSS

The systematic literature review study is accepted and published, titled “Methods used to evaluate usability of mobile clinical decision support systems for healthcare emergencies: a systematic review and qualitative synthesis” on JAMIA Open 6(3) (Wohlgemut, et al 2023). CDSS adoption framework scoping review titled “A Scoping Review, Novel Taxonomy and Catalogue of Implementation Frameworks for Clinical Decision Support Systems” is submitted to BMC Medical Informatics and Decision Making (Wohlgemut et al, 2023). The method for evaluating the prediction explanation narrative study, titled “A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models” is presented

at IEEE 11th International Conference of Healthcare Informatics, Texas, USA, and accepted for publication in IEEE proceedings (Pisirir et al, 2023). This work proposes the evaluation and validation process shown together with the explanation generation algorithm in the Figure below.



The explanation generation process starts with a set of scenarios related to a BN prediction model. With these scenarios, experts are asked what they think the “outcome” for this case would be and to explain why they think that is the outcome for each case. These explanations are then analysed to derive a template for the algorithmic explanation. Below our TIC-MIL based case study is summarised to illustrate the process.

Expert Elicitations: For the expert explanation elicitation sessions, we prepared 10 scenarios for patients with a traumatic injury and potential risk to develop TIC. The observations for the patients include clinical assessments, pre-hospital test results, measures that can be taken en route to the hospital, and lastly emergency department measures and results which can be taken upon arrival. Three trauma clinicians are invited to the explanation elicitation sessions and presented with the 10 scenarios. The experts are asked to assign a risk level of developing TIC to the patients and explain how and why they came to their prediction for each patient. One of the scenarios presented to the experts is given as an example below.

Box 1. Example scenario used in expert elicitation sessions

Pre-Hospital Information:

Pedestrian vs Car

Young Male with Head Injury, Low GCS

Handover:

Age: 35-year-old male

Mechanism of injury: Pedestrian vs Car

Energy of injury: High

Observation on Scene:

Systolic BP Un-recordable, Pulse 110, GCS 8

On arrival of HEMS: No Radial pulse

Patient intubated and ventilated.

Resuscitation Fluid: 1100mls (750mls crystalloid, 350mls Hetastarch), 1g TXA

Emergency Room - Resuscitation:

Observations:

Systolic BP 130, Pulse 108

Intubated and ventilated.

Chest X-Ray: Clear

Long Bone Fracture: None

Pelvis X-Ray: Overlapping bone of the pubic symphysis and impaction fracture at the posterior aspect of the left sacroiliac joint.

FAST: Not done

ABG: pH 7.428, Lactate 4, Base Excess -4, pCO₂ 3.88, pO₂ 42

When presented with this scenario, three clinicians gave the coagulopathy predictions and their reasoning behind these predictions (example below).

Box 2. Reasoning explanation of expert 1

I think this person's got a **moderate risk of developing TIC**. Because initially **he has some compromised perfusion** which evidence for that is **not being able to feel a radio pulse, not being able to record a blood pressure, and the tachycardia, and the fact he was also given resuscitation fluids to try and treat it**. In addition to the perfusion problems is **the moderate amount of tissue injury** which the evidence for that it's a pedestrian vs a car, so **high mechanism of injury, and he's broken his pelvis on x-ray which takes a lot of energy to break, and no other major injuries were seen**. In addition to that, the resuscitation fluids he's been given, so the **one litre of fluids could also have a very mild impact on his coagulation function**. So yeah, in summary only I think **moderate because of small degree of shock and some moderate injury**.

Expert Explanation Narrative Analysis: Expert explanations for all the scenarios are analysed to detect the selection and flow of information, the narrative, and the chains of reasoning which connect the observations to the outcome prediction. Example below shows the codified version of the explanation given above.

Box 5. Codified expert 1 explanation

```
(subject) + [target risk level] + (target)

{for i is an intermediate & affecting target:
  "Because" + [most probable state of i] + (i)
+ "for which the evidence"

  {for e is an evidence & related to i:
    [observed value of e] + (e)
  }
}
```

Expert Explanation Narrative Synthesis: After analysing the themes and language of the expert explanations from three clinicians on all the given scenarios, the flow of information and narrative codes can be synthesised to generate the initial template for the AI explanations. See example below:

Box 8. Template for the algorithmic AI explanation

```
(subject) + [target risk level] + (target)
[target risk level] + "because" +
{for i is an intermediate & affecting target:
  (i) + [most probable state of i]
}
(target) + [target risk level] + "because" +
{for e is an evidence & consistent with target
risk:
  (e) + [observed value of e]
}
"although" +
{for e is an evidence & conflicting with target
risk:
  (e) + [observed value of e]
}
{for i is an intermediate & affecting target:
  (i) + [most probable state of i] + "because"
  {for j is an evidence & consistent with risk
of i:
    (e) + (observed value of e)
  }
  "although" +
  {for j is an evidence & conflicting with
risk of i:
    (e) + (observed value of e)
  }
}
```

Narrative Generation for Algorithmic Explanations: This is the explanation generation step, using our mathematical explanation content generation algorithm, and the created template above to convert the content into a narrative. The technical details of the environment and the development are described in the previous annual report. Example below is the initial explanation narrative based on the template above.

Box 9. Initial algorithmic explanation of reasoning for TIC-BN

This patient has a high risk of trauma-induced coagulopathy. The high risk is because tissue injury is most likely to be severe and the hypoperfusion is most likely to be no. Trauma-induced coagulopathy is high risk because energy of injury is high, unstable pelvic fracture is yes, prehospital fluids given is greater than 500ml, and base excess is -4; although no long bone fracture is no, pH is 7.428, and systolic blood pressure is 130. Tissue injury is most likely to be severe because energy of injury is high, mechanism of injury is blunt, unstable pelvic fracture is yes, and GCS is 8; although long bone fracture is no and haemothorax is no. Hypoperfusion is no because systolic blood pressure is 130, heart rate is 108, and pH is 7.428; although lactate is 4 and base excess is -4.

Refinement of Algorithmic Explanations: The last stage of our explanation generation and validation process is the evaluation study focusing on the narrative of the algorithmic explanation. For TIC-BN, within our prototype usability evaluation study, we also conducted an evaluation study of the algorithmic explanations. Based on the initial participant feedback, the questionnaire and interview answers, and usability scores, the narrative given above is refined to reflect the user needs. Example given below.

Box 10. Refined algorithmic explanation of reasoning for TIC-BN

Brief explanation:

This patient has high risk of trauma-induced coagulopathy. High risk is because tissue injury burden is most likely to be severe, although the patient is unlikely to be shocked.

Full explanation:

The risk of trauma-induced coagulopathy is high because the patient has:

- a high energy injury
- an unstable pelvic fracture
- a low base excess
- more than 500ml of prehospital fluids given

The risk is high despite the patient having:

- no long bone fracture
- a normal pH value
- a normal blood pressure

Tissue injury burden is most likely to be severe because the patient has a high energy blunt injury, an unstable pelvic fracture, and a low level of consciousness – despite the patient having no long bone fracture and haemothorax.

The patient is unlikely to be shocked because the patient has a normal blood pressure, a normal heart rate, a normal pH value – despite the patient having low base excess and high lactate.

This version of the explanation narrative is used in the main stage of the prototype CDSS evaluation study and the results from the study show that the clinicians find this explanation narrative easy to understand, clear, concise, and natural, and they believe the presence of the explanation helps with making the CDSS and its predictions more transparent and useful.

Specific Aim 5, Major Task 2: Transfusion model development

A description of some of the challenges of prediction of blood transfusion requirement have been described and submitted as a research Letter to Military Medical Research (submitted 2023), entitled, “Enhancing the Clinical Relevance of Haemorrhage Prediction Models in Trauma” (Tandle et al, 2023)

Motivation: We aim to develop a clinically useful blood transfusion prediction model which uses reasonably available pre-hospital information to give time-based blood transfusion predictions.

Review: A review of existing blood transfusion prediction models in the literature is completed. 11 blood transfusion prediction models are reviewed: 8 of these models only aim to predict whether the patient will need a massive blood transfusion or not (more than 10 units of pRBCs in first 24 hours). 2 other model aims to predict whether the patient will

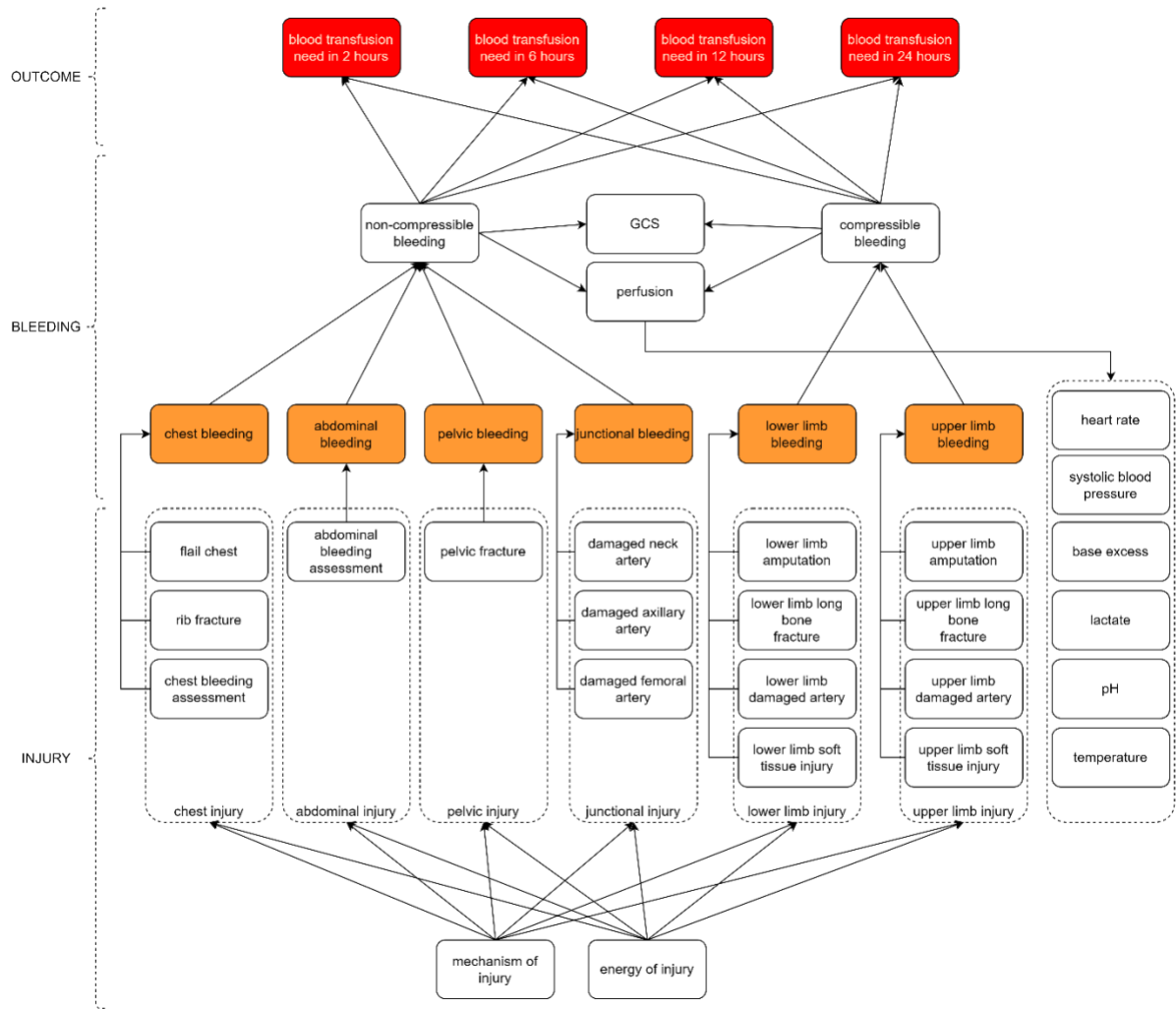
require blood transfusion at all or not, and 1 model aims to predict the amount of blood transfusion (with the accuracy of 0.764).

Data: The timestamped blood transfusion data from Camp Bastion Field Hospital, between January 2006 to December 2014 is obtained, processed and prepared for the model development and testing. This dataset includes patients who have received at least one unit of blood. Combined with the military patients from the rest of JTTR dataset (used for TIC-MIL development) who were injured but did not receive any blood transfusion, the training/test dataset for the blood transfusion prediction model is finalised. The Camp Bastion dataset provides timestamps for injury, arrival, and every unit of blood product transfused. The JTTR data provides the injury characteristics, injury severity, and physiology measurements. After removing fatalities from the dataset, the final training/test dataset includes 995 patients (547 of which have received at least one unit of blood transfusion within 24 hours from arrival). The reason for removing fatalities is to avoid survival bias.

Structure development: A knowledge- and data-driven structure development is completed. After iterations of model development, the final structure is given in the Figure below. The model structure features several new developments and novel fragments not used in the previously developed trauma models in the project.

Bleeding: The new structure which was not used in the previously developed models is the separation of compressible and non-compressible bleeding severity. Whether the bleeding is compressible or not is found to be a factor on the patient's transfusion needs. The injured body parts are categorised under these two bleeding severity categories: non-compressible bleeding (chest, abdomen, pelvis, junctional bleeding), and compressible bleeding (lower and upper limbs). The junctional bleeding is a new injury/bleeding structure which was not used in the previous models, explained in the next paragraph.

Injury: Injury characteristics and severity structure mostly follow the previously developed TIC and TAKI injury structure. The new development in this part of the model (which was not required in the previous models) is the junctional bleeding and separation of the arterial damage in upper and lower limbs based on the type. For the model structure to be fully credible and clinically accurate, the non-compressible vascular damage in lower and upper limbs should be separated from the compressible ones. For the upper limbs axillary (armpit) vascular damage and for the lower limbs femoral (groins) vascular damage is non-compressible. These two, together with neck vascular damage (which already exists in the model), make up the junctional bleeding. The model structure is updated to distinguish axillary and femoral arteries from the rest of the upper and lower limb ones and re-categorised the relevant injuries in the training and test data. The JTTR data registry contains AIS (abbreviated injury scale) scores for injuries. Originally, for lower limb arterial damage, the following injury codes were used: femoral artery (820204, 820206, 820208, 820299) and popliteal artery (820604, 820606, 820608, 820699). The upper limb arterial damage consisted of axillary (720206, 720208, 720299) and brachial (720604, 720606, 720608, 720699). For the revised model, patient injuries in the dataset are re-classified and femoral and axillary damage has been moved to non-compressible bleeding path.



Outcome: Another novel feature is that the model includes multiple outcome variables for different timeframes: 0-2 hours after arrival, 2-6 hours after arrival, 6-12 hours after arrival, and 12-24 hours after arrival. The decision to model different timeframes incrementally (as opposed to cumulatively: 0-2 hrs, 0-6 hrs, 0-12 hrs, and 0-24 hrs) came from the need to reduce the node table sizes and model complexity. The output nodes are discretised categorical variables based on true numerical values. After several iterations of model development based on data analysis and expert knowledge, the following discrete categories for the outcome nodes are defined:

Blood transfusion in 0-2 hours: [0-2], [3-6], [7-10], [11-20], [21+]
 Blood transfusion in 2-6 hours: None, [1-3], [4-10], [11-20], [21+]
 Blood transfusion in 6-12 hours: None, [1-3], [4-10], [11+]
 Blood transfusion in 12-24 hours: None, [1-3], [4-10], [11+]

This approach aims to discretise the continuous numerical true outcomes to make the modelling and predictions possible but not lose too much information. With our outcome structure which includes different time-based predictions and multiple categories of the blood amount, we aim to advance the existing blood transfusion prediction literature which is mostly based on binary massive transfusion or not predictions.

Parameter learning: The data-driven parameter learning is completed using EM algorithm. For the combinations of injuries where there was not enough data (extreme cases), we utilised the capabilities of Bayesian networks which allow combining knowledge and data,

and did expert knowledge based parameter elicitation to improve model performance for extreme cases.

Validation: Due to the complex outcome structure, the traditional ML prediction performance metrics such as AUROC, accuracy, calibration etc. are not applicable to evaluate the model performance. We developed two validation approaches, one based on categorised discrete outcome and one based on the numerical continuous true outcome and expected values of prediction. The results here are based on a 10-fold cross validation approach.

Probabilistic confusion matrix: The probabilistic confusion matrix shows the true outcomes and the model predictions using the discretised states of the model outcome. As far as true outcomes go, each patient falls into one category, however, the model prediction has a distribution over the possible states. Each patient can contribute to all the states of the prediction with a certain probability.

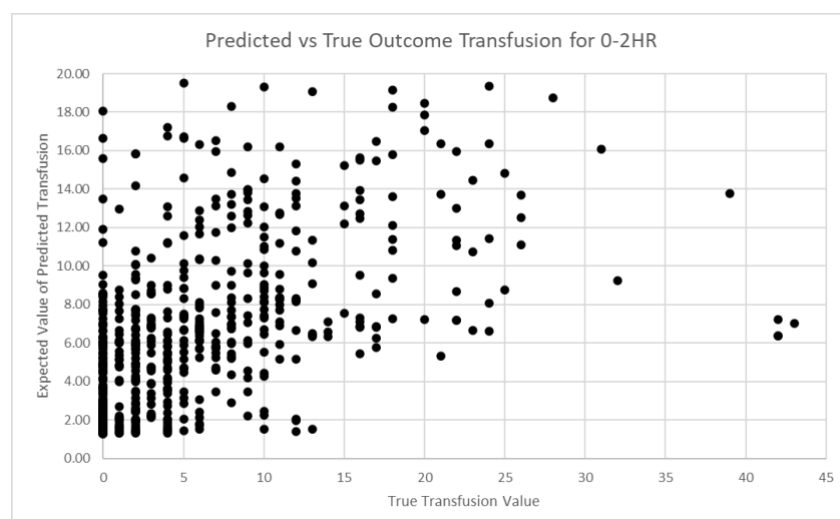
In the probabilistic confusion matrix, we categorised the prediction performance based on the states. A prediction is considered to have an “excellent” performance if the true outcome and true outcome and the prediction falls into the same discretised state. If the true outcome and prediction are in one adjacent state, then the prediction performance is “good”. If the true outcome and prediction fall into two adjacent states, the prediction performance is “mediocre”. if they are three states apart from each other, the performance is “poor”, and if they are four states apart from each other, the performance is “bad”. In the Table below, the confusion matrix (coloured based on the categorisation) is shown.

		True outcome				
		[0, 2]	[3, 6]	[7, 10]	[11, 20]	[21,)
Prediction	[0, 2]	526.21	67.24	31.23	24.00	7.79
	[3, 6]	57.03	27.49	19.22	17.71	8.31
	[7, 10]	28.88	19.25	15.62	16.28	6.42
	[11, 20]	23.39	17.06	15.48	17.55	10.53
	[21,)	7.50	5.95	6.44	10.45	5.94

Performance of the prediction based on the probabilistic confusion matrix and where the true outcome lies:

Prediction classification		Prediction classification		Prediction classification	
Excellent	59.70 %	Excellent		Excellent	81.40 %
Good	21.70 %	Good	92.25 %	Good	
Moderate	10.85 %	Moderate		Moderate	
Poor	6.21 %	Poor	7.75 %	Poor	18.60 %
Bad	1.54 %	Bad		Bad	

Expected value of prediction: Even though the model outcome is categorical, the categories are discretised intervals for the true numerical values of blood need, so we can calculate the expected value of the distribution for a model prediction. The value of each state is taken as the median of the state borders, and for the last state (>20 units), the maximum value observed in the dataset is used as the state value so that the maximum observed value is possible to expect from the distribution (in the hypothetical case that the model prediction has zero probability for every other state and a probability of one for the “>20 units” state). The true outcome values and the expected values of the prediction are distributed as shown in Figure below.



The prediction performance is categorised as excellent, good, moderate, poor, or bad based on the distance between the true outcome and the expected value of the prediction. The thresholds are defined as following: ± 2 from true outcome: excellent, $\pm 2-4$ from true outcome: good, $\pm 4-7$ from true outcome: moderate, $\pm 7-10$ from true outcome: poor, $> \pm 10$ from true outcome: bad.

Performance of the prediction based on the expected value of the prediction and the true continuous outcome:

Prediction classification	
Excellent	56.70 %
Good	19.44 %
Moderate	12.49 %
Poor	5.64 %
Bad	5.74 %

Prediction classification	
Excellent	
Good	88.62 %
Moderate	
Poor	
Bad	11.38 %

Prediction classification	
Excellent	
Good	76.13 %
Moderate	
Poor	23.87 %
Bad	

As both validation methods show, the model shows good performance (around 0.90%) when we look at the ratio of excellent, good and moderate predictions over poor and bad predictions (mispredictions).

Specific Aim 5, Major Task 3: Transfusion model external validation

See SA1, MT2. The safe haven environment is set up and the external validation data is obtained.

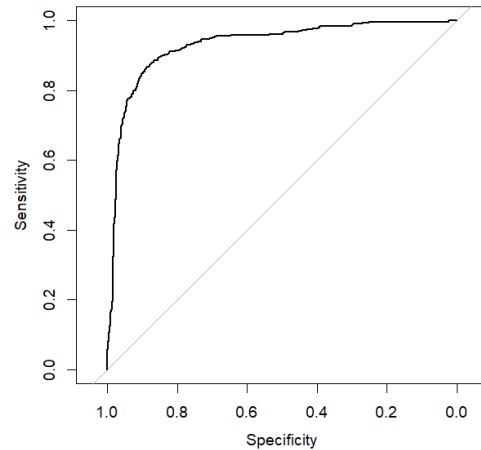
Processing, cleaning, and preparation of the data and the external validation of TIC-MIL model is planned for the six month no-cost extension duration.

Specific Aim 6, Major Task 4: SURVIVAL-MIL

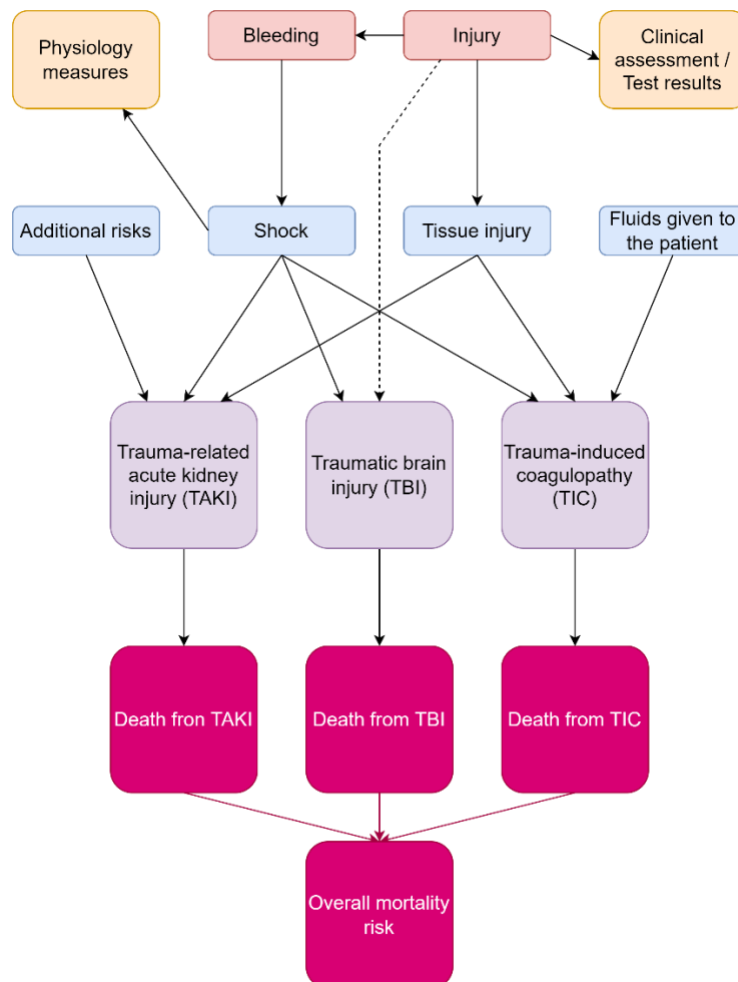
Previous annual report summarised the work on TIC and TAKI model as well as TBI model fragment development.

Previously developed TBI model is validated with the ACIT data and the initial validation results are given below.

- AUROC = 0.93 (0.92 - 0.95)
- Specificity at 0.90 sensitivity = 0.84
- Specificity at 0.80 sensitivity = 0.92
- Brier Score (BS) = 0.09



The survival model structure consists of these three trauma related risk prediction models, combined with “death from the condition” fragment. The overall model structure is given in Figure below.



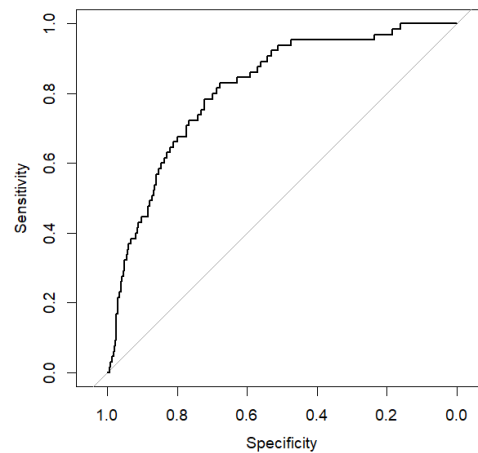
Here the first layer of variables (orange and red) represents the observable input variables: injury characteristics, injury severity, physiology measurements, and test results. The second layer of variables (blue) is the intermediate variables (tissue injury burden and shock) and other direct factors of the conditions (additional risks, prehospital fluids given to the patient). This layer leads to risk predictions of three trauma-related conditions: TIC,

TAKI, and TBI. The risks of three conditions then lead to the risk of death from these conditions, and eventually overall risk of trauma related mortality.

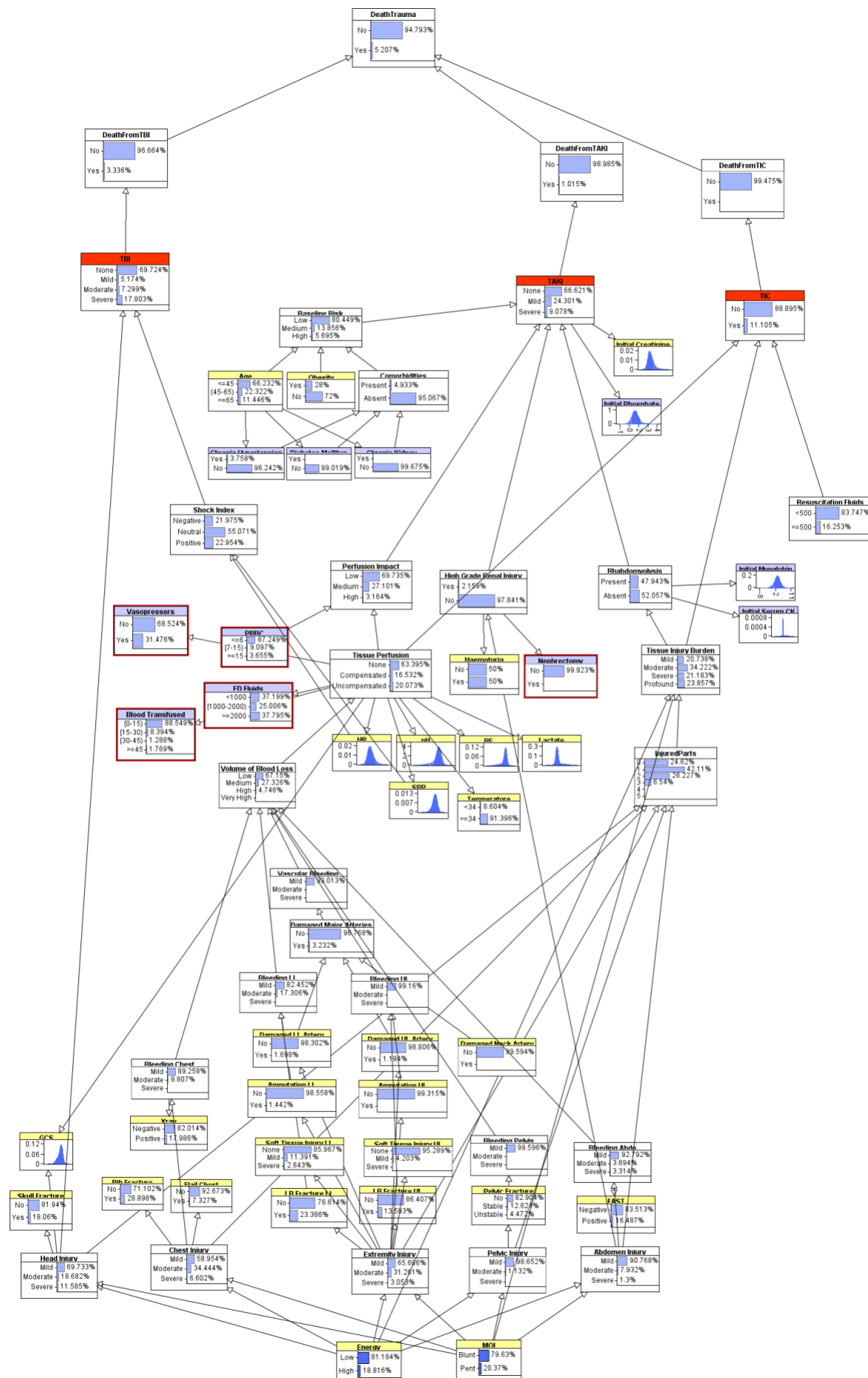
The model with all the parameters in the Bayesian network software agena.ai is shown in the Figure below.

The overall mortality model is trained using the ACIT data obtained from Royal London Hospital. The full data size is 1234 patients (mortality rate 5%). The true outcomes for TIC, TAKI, and TBI are obtained from the data, with a mix of data-driven learning and expert labelling, and the “death from these conditions” variable is labelled based on these true outcomes and patient mortality.

- AUROC = 0.81 (0.77 - 0.85)
- Specificity at 0.90 sensitivity = 0.54
- Specificity at 0.80 sensitivity = 0.70
- Brier Score (BS) = 0.05
- Skill Score = 0.05



Below, initial validation results are given. The next step is the adjustments/fine tuning to the model structure and parameters based on data and knowledge, 10-fold internal cross validation and generate relevant performance metrics such as AUROC, accuracy, and calibration.



Specific Aim 6, Major Task 5: SURVIVAL-MIL external validation

See SA1, MT2. The safe haven environment is set up and the external validation data is obtained.

Processing, cleaning, and preparation of the data and the external validation of TIC-MIL model is planned for the six month no-cost extension duration.

Specific Aim 7, Major Task 1: OVERFLIGHT-MIL development

Development of OVERFLIGHT model is planned for the six month no-cost extension duration. The development is planned to include the overall trauma-related mortality model and the concept of time-based outcomes, proven to perform well in the blood transfusion prediction work (See SA5, MT2).

Specific Aim 7, Major Task 2: OVERFLIGHT-MIL external validation

External validation will follow model development (SA 7, MT 1), using the safe haven environment COMBAT-AID-001 and the US DoDTR data (See SA1, MT2).

Specific Aim 8: CDSS development for SURVIVAL and OVERFLIGHT models

CDSS development will follow the current work mentioned in SA 3. The developed BNDS environment (with the explanation capabilities and the improved interface) will host SURVIVAL and OVERFLIGHT models once the models are fully developed and validated.

Specific Aim 9: Clinical usability evaluation of developed CDSS

SA 4 information applies here too, adjustments specifically related to SURVIVAL and OVERFLIGHT models might be added if necessary.

KNOWLEDGE PRODUCTS

Abstracts

International **Oral** Presentations:

- June 26-29th 2023 **A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models.** Pisirir E, Wohlgemut JM, Kyrimi E, Stoner RS, Perkins ZB, Tai NRM, Marsh DWR. The 11th IEEE International Conference on Healthcare Informatics, Houston, TX, USA
- May 9th 2023 **Usability of an Artificial Intelligence Clinical Decision Support System for Trauma: A Pilot Study.** Wohlgemut JM, Pisirir E, Kyrimi E, Stoner RS, Marsh W, Perkins ZB, Tai NRM) The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia
- May 8th 2023 **Diagnostic accuracy of Pre-hospital Clinical Assessment of Major Traumatic Haemorrhage.** Wohlgemut JM, Pisirir E, Stoner RS, Kyrimi E, Grier G, Christian M, Marsh W, Perkins ZB, Tai NRM. The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia
- Dec 8th 2022 **Critical decision analysis of pre-hospital major haemorrhage protocol activation.** Wohlgemut JM, McMullen H, Marsden M, Pisirir E, Stoner RS, Kyrimi E, Marsh W, Perkins ZB, Tai NRM. The London Trauma Conference 2022, London, UK
- Sep 21-24th 2022 **Developing an AI prediction model for Trauma-induced Acute Kidney Injury.**
R Stoner, E Kyrimi, E Pisirir, J Wohlgemut, W Marsh, Z Perkins, N Tai
Association of Surgery for Trauma, 81st Annual Meeting, Chicago, Illinois, USA

International **Poster** Presentations:

- May 17-19th 2023 **Critical decision analysis of pre-hospital major haemorrhage protocol activation.** Wohlgemut JM, McMullen H, Marsden M, Pisirir E, Stoner RS, Kyrimi E, Marsh W, Perkins ZB, Tai NRM. Association of Surgeons of Great Britain & Ireland (ASGBI) 2023 Congress, Harrogate, UK
- Sep 21-24th 2022 **Machine Learning to Aid Decision Making in Patients who may Benefit from Prehospital Tranexamic Acid.** Wohlgemut JM, Almuwallad AO, Kyrimi E, Rossetto A, Stoner RS, Cole E, Marsh W, Perkins ZB, Tai NRM, Davenport R. American Association of Surgery for Trauma, 81st Annual Meeting, Chicago, Illinois, USA
- Sep 12-15th 2022 **Incorporating Clinician Accuracy and Uncertainty of Prehospital Injury Diagnosis Improves Clinical Validity of a Decision Support System Without Compromising Performance of a Bayesian Network to Predict Trauma-Induced Coagulopathy.** Wohlgemut JM, Kyrimi E, Rossetto A, Stoner R, Pisirir E, Marsh W, Perkins Z, Tai N. Military Health Services Research Symposium 2022, Kissimmee, Florida, USA

Papers

Accepted by Peer-Reviewed Journals:

Durrands T, Murphy M, Wohlgemut JM, De'Ath HD, Perkins ZB. Diagnostic accuracy for identification of life-threatening torso injuries: A Meta-analysis. BJS 2023 (accepted)
Marsden MER, Kellett S, Bagga R, Wohlgemut JM, Lyon R, Perkins ZB, Gillies K, Tai N. Understanding pre-hospital blood transfusion decision-making for injured patients: an interview study. Emergency Medical Journal (accepted 01/08/2023)

Wohlgemut JM, Pisirir E, Kyrimi E, Stoner RS, Marsh W, Perkins NRM, Tai NRM. Methods Used to Evaluate Usability of Mobile Clinical Decision Support Systems for Healthcare Emergencies: A Systematic Review and Qualitative Synthesis. JAMIA Open 2023 (accepted 26/6/2023)

Pisirir E, Wohlgemut JM, Kyrimi E, Stoner RS, Perkins ZB, Tai NRM, Marsh DWR. A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models. IEEE Xplore 2023. (accepted March 2023)

Wohlgemut JM, Marsden MER, Stoner RS, Pisirir E, Kyrimi E, Grier G, Christian M, Hurst T, Marsh W, Tai NRM, Perkins ZB. Diagnostic Accuracy of Clinical Examination to Identify Life- and Limb-Threatening Injuries in Trauma Patients. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 2023;31:18.

Submitted to Peer-Reviewed Journals:

6. Kyrimi et al, "A Scoping Review, Novel Taxonomy and Catalogue of Implementation Frameworks for Clinical Decision Support Systems" BMC Medical Informatics and Decision Making, submitted 2023.
7. Wohlgemut et al, "Identification of Major Hemorrhage in Trauma Patients in the Pre-Hospital Setting: Diagnostic Accuracy and Impact on Outcome" Trauma Surgery Acute Care Open, submitted 14/7/23
8. Kyrimi et al, "Updating and recalibrating causal probabilistic models on a new target population" Journal of Biomedical Informatics, resubmitted 6/7/23
9. Kyrimi et al, "Counterfactual reasoning using Bayesian networks as a healthcare governance tool" Artificial Intelligence in Medicine, submitted 5/6/23
10. Tandle et al, "Enhancing the Clinical Relevance of Haemorrhage Prediction Models in Trauma" Military Medical Research, submitted 2023.

Prizes

May 2023 **TSACO Prize-winning oral abstract**
22nd European Congress of Trauma and Emergency Surgery (ECTES) 2023
Resident paper competition, winning cost of article processing charge for TSACO (Trauma Surgery and Acute Care Open), ECTES 2023, Ljubljana, Slovenia, 7-9 May 2023

- Feb 2023 **Post-Graduate Research Fund, Queen Mary University of London**
Awarded £1000 to present research at the 22nd European Congress of Trauma and Emergency Surgery (ECTES) 2023, Ljubljana, Slovenia, 7-9 May 2023
- Sept 2022 **Poster Award: Honorable Mention**
2022 Military Health Services Research Symposium
Gaylord Palms Resort and Convention Center, Kissimmee, Florida, USA, 12-15 Sept 2022
<https://mhsrs.amedd.army.mil/SitePages/Awards-Winners.aspx>
<https://health.mil/News/Articles/2022/09/26/Young-Investigators-and-Poster-Awardees-Recognized-at-Symposium>

Conferences

- June 26-29th 2023: The 11th IEEE International Conference on Healthcare Informatics, Houston, TX, USA. *J Wohlgemut (on behalf of E Pisirir)*
- May 17-19th, 2023: Association of Surgeons of Great Britain & Ireland (ASGBI) 2023 Congress, Harrogate, UK. *J Wohlgemut*
- May 7-9th 2023: The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia. *J Wohlgemut*
- March 1st, 2023: Closer, Better, Longer: MDS-DSTL Blood Symposium, Victory Service Club London, UK. *Z Perkins and W Marsh Presenting, on behalf of N Tai.*
- Dec 6-9th, 2022: The London Trauma Conference 2022, London, UK. *J Wohlgemut*
- Sep 21-24th, 2022: American Association of Surgery for Trauma, 81st Annual Meeting, Chicago, Illinois, USA. *RS Stoner, J Wohlgemut*
- Sept 12-15th, 2022: Military Health Services Research Symposium 2022, Kissimmee, Florida, USA. *J Wohlgemut*

Abstract 1

Oral Presentation to the 11th IEEE International Conference on Healthcare Informatics, Houston, TX, USA, June 26-29th 2023.

A Process for Evaluating Explanations for Transparent and Trustworthy AI Prediction Models

Erhan Pisirir, Jared M Wohlgemut, Evangelia Kyrimi, Rebecca S Stoner, Zane B Perkins, Nigel RM Tai, D William R Marsh

Department of Electronic Engineering and Computer Science, Queen Mary University of London, UK

Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, UK

Abstract— This study proposes a process to generate and validate algorithmic explanations for the reasoning of an AI prediction model, implemented using a Bayesian network (BN). The intention of the generated explanations is to increase the transparency and trustworthiness of a decision-support system that uses a BN prediction model. To achieve this, explanations should be presented in an easy-to-understand, clear, and concise natural language narrative. We have developed an algorithm for explaining the reasoning of a prediction made using a BN. For the narrative part of the explanation, we use a template which presents the ‘content’ part of the explanation; this content is a word-less information structure that applies to all BNs. The template, on the other hand, needs to be designed specifically for each BN model. In this paper, we use a BN for the risk of trauma-induced coagulopathy, a critical bleeding problem. We outline a process for using experts’ explanations as the basis for designing the explanation template. We do not believe that an algorithmic explanation needs to be indistinguishable from expert explanations; instead we aim to imitate the narrative structure of explanations given by experts, although we find that there is considerable variation in these. We then consider how the generated explanations can be evaluated, since a direct comparison (in the style of a Turing test) would likely fail. We describe a study using questionnaires and interviews to evaluate the effect of an algorithmic explanation on the transparency and also on the trustworthiness of the predictions made by the system. The preliminary results of our study suggest that the presence of an explanation makes the AI model more transparent but not necessarily more trustworthy.

Abstract 2

Oral Presentation to The 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia, 7-9th May 2023.

Usability of an Artificial Intelligence Clinical Decision Support System for Trauma: A Pilot Study

Wohlgemut JM, Pisirir E, Kyrimi E, Stoner RS, Marsh W, Perkins ZB, Tai NRM

Department of Electronic Engineering and Computer Science, Queen Mary University of London, UK

Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, UK

Introduction

Trauma-induced coagulopathy (TIC) affects 25% of severely injured patients, and mortality approaches 50%. An Artificial Intelligence (AI) model was developed and validated to predict TIC, and integrated into a clinical decision support system (CDSS). This pilot study aimed to evaluate our usability study design, determining the usability, usefulness, and potential barriers to adoption of our AI CDSS.

Material & Methods

This was a prospective, mixed-methods, pilot usability study. Participants (n=5) tested the CDSS by inputting information into its interface, generating a risk prediction output (Fig 1). Usability issues were captured using "think aloud" methodology, categorised using Nielsen's Heuristics (NH), and graded for severity (1=cosmetic, 2=minor, 3=major, 4=catastrophic). A validated questionnaire was administered based on the System Usability Scale (SUS), Technology Acceptance Model (TAM), and nonadoption, abandonment, scale-up, sustainability and spread (NASSS) adoption framework.

Results

Participants identified 18 study design problems: 13 user testing, 5 questionnaire-related. Usability issues were identified in 9/10 NH categories (Fig 2): 40 input- and 11 output-related (average severity 2.1/4 and 1.8/4, respectively). The most common issue was "consistency and standards" (n=14). The CDSS achieved an SUS score of 60 (standard is ≥68). Participants said the CDSS would be useful in their clinical practice (60%), improve their clinical performance (60%), and enhance their clinical effectiveness (40%). Potential barriers were related to the adopters, organisation and external environment.

Conclusions

This pilot study identified issues with study design, and usability of the AI CDSS. After iterative improvements, a prospective study will evaluate the usability, clinician trust, and impact on decision-making of an updated AI CDSS in trauma.

Figure 1. CDSS user interface

A) Input

B) Output

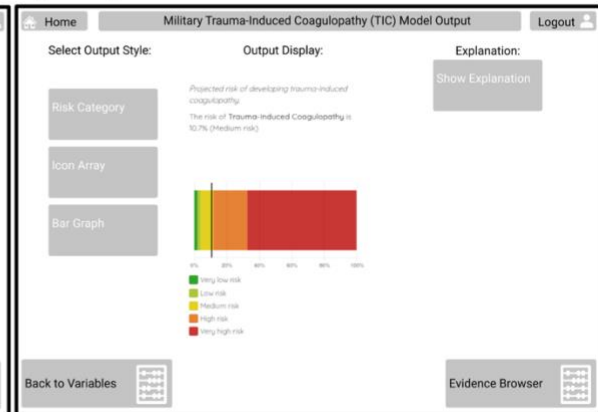
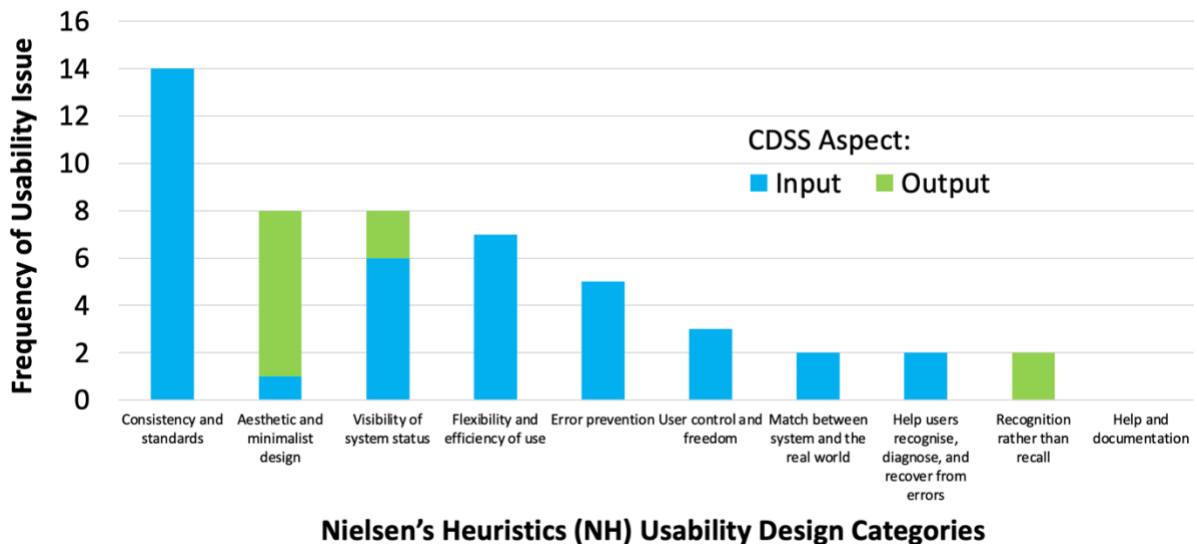


Figure 2. NH usability issues



Abstract 3

Oral Presentation to the 22nd European Congress of Trauma and Emergency Surgery, Ljubljana, Slovenia, 7-9th May 2023.

Diagnostic accuracy of Pre-hospital Clinical Assessment of Major Traumatic Haemorrhage

Wohlgemut JM, Pisirir E, Stoner RS, Kyrimi E, Grier G, Christian M, Marsh W, Perkins ZB, Tai NRM

Department of Electronic Engineering and Computer Science, Queen Mary University of London, UK

Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, UK

Introduction

Traumatic haemorrhage is the largest cause of potentially preventable death after injury. Pre-hospital diagnosis of major haemorrhage (MH) can be difficult, even for expert clinicians. The aim was to determine the accuracy of pre-hospital MH diagnosis, and analyse factors associated with missed MH diagnosis and mortality.

Material & Methods

Retrospective evaluation of consecutive adult (≥ 16 years) patients injured in 2019-2020, assessed by expert clinicians in a mature pre-hospital trauma system, and admitted to a Major Trauma Centre. Diagnostic performance to identify MH pre-hospital was compared to in-hospital diagnosis. Pre-hospital MH protocol activation was used as a surrogate for MH diagnosis. MH was defined using the critical admission threshold (CAT+): ≥ 3 units of red blood cells given within any 60min period in the first 24hr of injury. Multivariate logistical regression analyses were conducted, with different dependent variables: missed MH diagnosis and mortality.

Results

There were 947 patients included in the study. Median age was 31 years (IQR 23-47), 821 (87%) were male, 569 (60%) had blunt injury, median injury severity score was 10 (IQR 4-22), and 79 (8%) died. Pre-hospital clinicians correctly diagnosed MH in 97/138 CAT+ patients (sensitivity 70%; 95% CI 62-77%), and correctly ruled out MH in 764/809 CAT- patients (specificity 94%; 95% CI 93-96%, Tab 1). Factors associated with missed MH diagnosis were penetrating mechanism (OR 2.4; 95% CI 1.2-4.7) and major abdominal bleeding injury (OR 4.0; 95% CI 1.7-8.7, Tab 2). The odds of mortality increased by 3.3 (95% CI 1.1-9.7) when MH was missed.

Conclusions

Clinical assessment by expert clinicians had only moderate ability to identify MH in the pre-hospital setting. Understanding the limitations of clinical assessment, and developing solutions to aid identification of MH are warranted.

Tab 1. Contingency table of MH vs CAT

	CAT+ (≥ 3 U RBC/60min in 24hr)	CAT- (< 3 U RBC/60min in 24hr)
MH ("Code Red")	97	45

No MH (No “Code Red”)	41	764
-----------------------	----	-----

Tab 2. Multivariate logistic regression of missed MH

			Dependent variable: missed MH diagnosis, n=41			
Factor	Variable	Referent	Univariate OR (95% CI)	p	Multivariate* OR (95% CI)	p
Patient	Age	-	0.989 (0.967-1.01)	0.27		
	Female Sex	Male Sex	1.62 (0.682- 3.42)	0.24		
	Penetrating MOI	Blunt MOI	2.45 (1.30- 4.74)	0.006	2.37 (1.24- 4.65)	0.010
	GCS	-	0.999 (0.929-1.08)	0.98		
	SBP	-	0.993 (0.985-1.00)	0.09	0.997 (0.988- 1.01)	0.56
	HR	-	1.00 (0.989- 1.01)	0.98		
	Polytrauma	-	1.11 (0.890- 1.36)	0.33		
Bleeding Injuries	Chest	No chest bleed	3.58 (1.55- 7.53)	0.001	2.36 (0.963- 5.33)	0.05
	Abdomen	No abdominal bleed	4.90 (2.18- 10.2)	<0.001	4.04 (1.74- 8.70)	<0.001
	Unstable pelvis	Stable pelvis	1.40 (0.222- 4.86)	0.65		
	Peripheral	No peripheral bleed	2.16 (0.338- 7.73)	0.31		
	Major vascular	No major vascular injury	5.25 (2.74- 9.98)	<0.001**	Collinear	
Clinician Experience	1-5 years	<1 year	0.644 (0.190-1.66)	0.41		
	5-10 years	<1 year	1.95 (0.452- 5.80)	0.29		
	>10 years	<1 year	1.63 (0.452- 5.80)	0.52		
Clinician Base	Anaesthetics	EM	1.14 (0.575- 2.20)	0.70		
Specialty	ICM	EM	0.904 (0.211-2.68)	0.87		
Environment	Nightshift	Dayshift	1.09 (0.584- 2.08)	0.78		
MH= major haemorrhage; OR= odds ratio; CI= confidence interval; MOI= mechanism of injury; GCS= Glasgow Coma Scale; SBP= systolic blood pressure; HR= heart rate; Polytrauma= number of abbreviated injury score (AIS) body regions injured; EM= emergency medicine; ICM= intensive care medicine *Multivariate variables were limited to four due to event rate of 41 patients.						

***'Major vascular bleeding injury' excluded from multivariate due to collinearity with other bleeding injuries (chest, abdomen, unstable pelvis, peripheral), as it combined these injuries in one category.

Variables excluded from the multivariate included age, sex, GCS, HR, polytrauma, unstable pelvis, peripheral bleeding injury, major vascular bleeding injury, clinician experience, clinician base specialty, and shift.

Abstract 4

Oral Presentation to London Trauma Conference, London, UK, 6-9th Dec 2022.

Poster Presentation to the ASGBI 2023 Congress, Harrogate, UK, 17-19th May 2023.

Critical decision analysis of pre-hospital major haemorrhage protocol activation

Jared M Wohlgemut^{1*}, Heather McMullen², Max Marsden¹, Erhan Pisirir³, Rebecca S Stoner¹, Evangelia Kyrimi³, William Marsh³, Zane B Perkins⁴, Nigel RM Tai⁵

¹Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, London, UK

²Global Public Health Unit, Centre for Public Health and Policy, Wolfson Institute of Population Health, Queen Mary University of London, London, UK

³Department of Electrical Engineering and Computer Science, Queen Mary University of London, London, UK

⁴London's Air Ambulance, 17th Floor, Royal London Hospital, London, UK

⁵Trauma Service, Royal London Hospital, Barts Health NHS Trust, London, UK

*j.wohlgemut@qmul.ac.uk

Background

Major haemorrhage protocol activation (MHPA) is designed to treat major bleeding by facilitating immediate blood product availability and institutional capacity readiness. In the London trauma system, two thirds of MHPA decisions occur pre-hospital. Understanding what influences the pre-hospital MHPA decision is a prerequisite to improving clinician decision-making processes. Our aim was to identify the critical determinants of the pre-hospital MHPA decision.

Methods

This prospective qualitative study was approved by the Health Research Authority (REC 22/PR/0089). Physicians from one helicopter emergency medical service (HEMS) were purposively sampled, and written informed consent was obtained. Cases where the MHPA decision was difficult within the last six months were included. Semi-structured interviews were conducted on the Microsoft Teams platform, according to the critical decision analysis method.[1] Video- and audio-recordings were transcribed and analysed according to the 6-step inductive thematic analysis method using NVivo software.[2]

Results

Twelve interviews were conducted with nine participants: median age 37, median post-graduate experience 12 years (inter-quartile range (IQR) 10.5-14.8), and median HEMS experience 1 year (IQR 0.5-4.9). Cues for the MHPA decision included: hard evidence of bleeding, initial information fit with their mental model of a bleeding patient, patient deterioration, acknowledgement that injuries and bleeding can be missed, and when alternative diagnoses were considered less likely. Cues against the MHPA decision included: initial information did not fit with their mental model of a bleeding patient, responsive to resuscitation, uncertainty of clinical findings, bleeding judged to be non-severe, and when alternative diagnosis were considered more likely. When considering the MHPA decision, clinicians were influenced by previous experience, colleagues, time pressure, overwhelming task volume, and the desire to prevent delay while ensuring diagnostic accuracy and professional reputation.

Conclusion

There are multiple determinants of the pre-hospital MHPA decision. Acknowledging these cognitive processes will inform efforts to improve decision-making in trauma care.

References

1. Klein GC, Macgregor D. Critical Decision Method for Eliciting Knowledge. IEEE Transactions on Systems, Man and Cybernetics. 1989; 19(3): 462-72.
2. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006; 3(2):77-101.

Abstract 5

Poster Presentation to the American Association of Surgery for Trauma, 81st Annual Meeting, Chicago, Illinois, USA, 21-24th September 2022.

Machine Learning to Aid Decision Making in Patients who may Benefit from Prehospital Tranexamic Acid

Wohlgemut JM, Almuwallad AO, Kyrimi E, Rossetto A, Stoner RS, Cole E, Marsh W, Perkins ZB, Tai NRM, Davenport R

Background: Trauma-induced coagulopathy (TIC) is characterised in part by hyperfibrinolysis. Tranexamic acid (TXA) confers maximum benefit in patients at risk of TIC if given within one hour of injury. Early TXA treatment relies on clinicians identifying patients at risk of TIC, before any viscoelastic assessments can aid decision making. This can be challenging, thus TXA is often given late or not given. A validated Bayesian Network (BN) model that accurately predicts TIC (TIC-BN) using prehospital (PH) variables may help identify patients in whom PH TXA is indicated. The aim was to determine the proportion of patients in each TIC-BN risk strata who received TXA ≤1hr from injury. Secondary aims were to compare TXA treatment within TIC-BN risk strata among outcome-related subgroups.

Methods: Retrospective study of the UK trauma audit research network database (TARN) between 2015-2019, of patients aged ≥16 years, who had hemorrhage control intervention or blood transfusion, excluding pregnancies and burns. Patients were risk-stratified using TIC-BN: very low (<2%), low (2-4%), moderate (4-12%), high (12-33%), and very high (>33%).

Results: A total of 27,272 patients were included. Median age was 51 years (interquartile range (IQR) 32-71), 68% were male, 86% suffered blunt injury, median injury severity score was 13 (IQR 9-25). Overall, 5571 (20%) patients received TXA ≤1hr, including 368/3403 (11%) with very low risk of TIC, 494/4764 (10%) low risk, 1805/9408 (19%) medium risk, 2003/8173 (25%) high risk, and 901/1524 (59%) very high risk (χ^2 for trend TXA ≤1hr vs rest: $p < 0.001$). The proportion of patients who had TXA ≤1hr increased as TIC risk increased among outcome-related subgroups: major transfusion (≥4U blood products), massive transfusion (≥10U), early (≤24hr) and late (≤28dy) mortality (all $p < 0.05$). However, of those with medium or higher TIC risk, 53% of all patients, 37% of massive transfusion patients, and 46% of patients who died early, did not receive TXA ≤1hr.

Conclusions: Eighty percent of the patients for whom TXA is indicated do not receive it within one hour of injury, including over half of those at medium TIC risk or greater. Earlier identification of TXA-eligible patients predicated on accurate machine learning risk prediction using prehospital variables, may allow more tailored treatment for patients at risk of TIC.

Abstract 6

Poster Presentation to the Military Health Services Research Symposium 2022, Kissimmee, Florida, USA, 12-15th September 2022.

Incorporating Clinician Accuracy and Uncertainty of Prehospital Injury Diagnosis Improves Clinical Validity of a Decision Support System Without Compromising Performance of a Bayesian Network to Predict Trauma-Induced Coagulopathy

Wohlgemut JM, Kyrimi E, Rossetto A, Stoner R, Pisirir E, Marsh W, Perkins Z, Tai N

¹Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, UK

²Royal London Hospital, Barts NHS Health Trust, London, UK

³Department of Electrical Engineering and Computer Science, Queen Mary University of London, UK

⁴Academic Department of Military Surgery and Trauma, Royal Centre of Defence Medicine, Birmingham, UK

Learning Objectives:

1. Describe the accuracy of prehospital clinical assessment in identifying major injuries (hemothorax, abdominal bleeding, unstable pelvis, and long bone fracture) and investigate the impact of clinician uncertainty on diagnostic accuracy
2. Describe how incorporating clinician accuracy and uncertainty into a Bayesian Network model for trauma-induced coagulopathy impacts its predictive performance
3. Discuss how end user credibility of a clinical decision support system could be strengthened by incorporating clinician accuracy and uncertainty

Introduction

Traumatic hemorrhage is the leading preventable cause of death after injury. Trauma-induced coagulopathy (TIC) is a complication of severe traumatic hemorrhage which is associated with increased mortality. Early treatment of TIC can be life-saving, but identification of at-risk patients can be difficult, especially in prehospital and austere settings. The likelihood of TIC calculated by a clinical decision support system (CDSS) could aid decisions about treatment, transport, and prioritization. The use of such CDSS may result in improved patient outcomes from more timely and informed interventions.

A Bayesian Network (BN) has been developed to predict trauma-induced coagulopathy (TIC-BN). BNs are graphical probabilistic models which can be designed using expert knowledge and clinical data. The TIC-BN model incorporates patient demographic, physiological, laboratory, and injury variables. However, the accuracy of prehospital clinical assessment to diagnose injuries may be imperfect, and clinician uncertainty may further reduce diagnostic accuracy. We hypothesized that incorporating clinician accuracy and uncertainty of injury diagnoses in the TIC-BN structure would improve clinical validity by ensuring the CDSS can handle real-world information from users in forward and en route combat casualty care environments, without reducing predictive accuracy.

Our aims were to 1) determine the diagnostic accuracy of prehospital clinical assessment by expert trauma clinicians in identifying major injuries; 2) determine the frequency of clinician uncertainty of major injuries, and the impact on diagnostic accuracy; and 3) determine how incorporating diagnostic accuracy and uncertainty in an updated TIC-BN model structure impacts predictive performance.

Materials and Methods

Study Design

Retrospective validation study incorporating clinician diagnostic accuracy and uncertainty nodes into the structure of an existing TIC-BN model. The study has been reported according to STARD and TRIPOD guidelines.

Study Population

Consecutively injured patients assessed by expert prehospital trauma clinicians in an urban helicopter emergency medical service and admitted to the regional Level 1 Trauma Center between 01/01/2019 and 12/31/2020 were included. Patients were excluded if the model outcome was unknown (TIC defined as international normalized ratio (INR) >1.2), age <16 years, and non-trauma or thermal injury.

Data Sources

Patient demographics, physiology, injuries, and clinician uncertainty of injury diagnoses were collected from prehospital documentation. Laboratory data and discharge diagnoses were collected from the Trauma Center clinical registry. Classification of true outcome (TIC) was facilitated by comparing an expectation-maximization algorithm to the INR results. Discrepancies were resolved by expert consensus.

Injury Accuracy and Uncertainty

For each injury, prehospital diagnosis was compared to discharge diagnosis. 2x2 contingency tables were constructed, and sensitivity, and specificity were reported. Contingency tables were repeated for each injury for which clinicians documented diagnostic uncertainty.

Updating the TIC-BN structure

BN is a directed acyclic graph with qualitative and quantitative parts. The qualitative BN *structure* comprises nodes representing random variables and directed arcs signifying causal or influential relationships. The qualitative BN *parameters* comprise a set of conditional probability functions. Variables of TIC-BN relate to injury (energy, mechanism, Glasgow Coma Scale, hemothorax, abdominal bleeding, unstable pelvis fracture, and long bone fracture), perfusion (lactate, base excess, pH, prehospital heart rate, systolic blood pressure, and temperature), and coagulopathy (prehospital fluid administration). The previously developed TIC-BN had excellent predictive performance at external validation: area under the receiver operator curve (AUC) 0.95, calibration slope 0.96 (ideally 1), Brier Score (BS) 0.05 (ideally 0), and Brier Skill Score (BSS) 0.46 (ideally 1). The TIC-BN structure was revised to incorporate prehospital diagnostic accuracy and clinician uncertainty for hemothorax, abdominal bleeding, and unstable pelvis; and diagnostic accuracy only for long bone fracture, as uncertainty was infrequent.

TIC-BN Model Performance

The predictive performance of the updated TIC-BN model was assessed by its discrimination, accuracy and calibration, in three iterations. The 'prehospital' iteration included prehospital clinical assessment in injury nodes that represent in-hospital diagnostic tests. The 'prehospital + accuracy' iteration added nodes for prehospital clinical assessment of injuries. The 'prehospital + accuracy + uncertainty' iteration added nodes for clinician uncertainty.

Statistical Analysis

Diagnostic accuracy analysis was conducted using Prism v9.0.2. TIC-BN model reasoning was conducted using AgenaRisk and the model's performance was evaluated using R v4.1.2.

Results

Study Population

Our validation cohort comprised 811 patients who met eligibility criteria. Their median age was 32 years (interquartile range (IQR) 23-48), 703 (86.7%) were male, 497 (61.3%) suffered blunt mechanism, median injury severity score (ISS) was 13 (IQR 5-25), and mortality was 69/811 (8.5%). The rate of TIC in our cohort was 82/811 (10.1%).

Injury Accuracy and Uncertainty

The sensitivity and specificity of prehospital clinical assessment were respectively 32.4% and 92.5% for hemothorax; 42.2% and 89.2% for abdominal bleeding; 19.2% and 99.0% for unstable pelvis; and 65.5% and 95.5% for long bone fracture.

Clinicians expressed uncertainty in 21/79 of hemothorax (26.6%), in 66/108 of abdominal bleeding (61.1%), in 3/13 of unstable pelvis (23.1%), and in 8/109 of long bone fracture diagnoses (7.3%).

When clinicians were uncertain, the sensitivity and specificity of prehospital clinical assessment were respectively 100.0% and 0.0% for hemothorax; 93.3% and 3.7% for abdominal bleeding; 0% and 89.3% for unstable pelvis; and 100.0% and 33.3% for long bone fracture.

TIC-BN Model Performance

Prehospital

Discrimination was moderate with AUC 0.77 (95% confidence interval (CI) 0.73–0.82). Accuracy was 0.81, the BS was 0.08, and the BSS was 0.14. Calibration slope was 0.925, and calibration intercept was 0.028.

Prehospital + Accuracy

Discrimination was similar with AUC 0.78 (95% CI 0.74–0.82), accuracy 0.79, BS 0.08, BSS 0.16, calibration slope 0.922, and calibration intercept 0.018.

Prehospital + Accuracy + Uncertainty

Discrimination was similar with AUC 0.78 (95% CI 0.74–0.82), accuracy 0.78, BS 0.08, BSS 0.15, calibration slope 0.905, and calibration intercept 0.018.

Conclusions

Prehospital clinical assessment by expert trauma clinicians to diagnose major injuries is imperfect. Clinician uncertainty worsens accuracy by reducing specificity, due to more false positive diagnoses. In the prehospital context - compared to the in-hospital context in which the original model which was developed and validated - discrimination reduced, though accuracy and calibration remained excellent. By incorporating prehospital accuracy and uncertainty to the TIC-BN structure, predictive performance was not reduced between iterations. The rationale for incorporating this information in the CDSS was to allow users to express their uncertainty, mirroring the cognitive reality of prehospital and austere clinical practice, and take into account known diagnostic inaccuracies, improving its clinical validity.

Limitations to our study included the surrogate definitions of clinician accuracy and uncertainty. Future work includes replicating this methodology on our military TIC-BN model to evaluate impact on predictive performance in the context of military casualties; and prospective user testing of the updated TIC-BN which incorporates clinician diagnostic accuracy and uncertainty, to evaluating model credibility and trust.