

A comparison of methods for assessing penetrating trauma on retrospective multi-center data

Bilal A. Ahmed^a, Michael E. Matheny^{b,c,e}, Phillip L. Rice^d, John R. Clarke^f, Omolola I. Ogunyemi^{g,*}

^a University of Toronto, Faculty of Medicine, Toronto, Ont., Canada

^b Division of General Internal Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

^c Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

^d Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

^e Center for Health Services Research, Tennessee Valley Health System, Veteran's Administration, Nashville, TN, USA

^f Department of Surgery, Drexel University, Philadelphia, PA, USA

^g Center for Biomedical Informatics, Charles Drew University of Medicine and Science, Los Angeles, CA, USA

ARTICLE INFO

Article history:

Received 24 March 2008

Available online 1 October 2008

Keywords:

Artificial intelligence

Computer-assisted diagnosis

Stab wounds

Gunshot wounds

Penetrating trauma

ABSTRACT

Objective: TraumaSCAN-Web (TSW) is a computerized decision support system for assessing chest and abdominal penetrating trauma which utilizes 3D geometric reasoning and a Bayesian network with subjective probabilities obtained from an expert. The goal of the present study is to determine whether a trauma risk prediction approach using a Bayesian network with a predefined structure and probabilities learned from penetrating trauma data is comparable in diagnostic accuracy to TSW.

Methods: Parameters for two Bayesian networks with expert-defined structures were learned from 637 gunshot and stab wound cases from three hospitals, and diagnostic accuracy was assessed using 10-fold cross-validation. The first network included information on external wound locations, while the second network did not. Diagnostic accuracy of learned networks was compared to that of TSW on 194 previously evaluated cases.

Results: For 23 of the 24 conditions modeled by TraumaSCAN-Web, 16 conditions had Areas Under the ROC Curve (AUCs) greater than 0.90 while 21 conditions had AUCs greater than 0.75 for the first network. For the second network, 16 and 20 conditions had AUCs greater than 0.90 and 0.75, respectively. AUC results for learned networks on 194 previously evaluated cases were better than or equal to AUC results for TSW for all diagnoses evaluated except diaphragm and heart injuries.

Conclusions: For 23 of the 24 penetrating trauma conditions studied, a trauma diagnosis approach using Bayesian networks with predefined structure and probabilities learned from penetrating trauma data was better than or equal in diagnostic accuracy to TSW. In many cases, information on wound location in the first network did not significantly add to predictive accuracy. The study suggests that a decision support approach that uses parameter-learned Bayesian networks may be sufficient for assessing some penetrating trauma conditions.

© 2008 Elsevier Inc. All rights reserved.

1. Background

Probabilistic methods for computerized diagnostic decision support have been demonstrated to be useful in addressing medical problems such as pneumonia diagnosis, mammographic diagnosis of breast cancer, lymph-node disease, abdominal pain, and hypercalcemia [1–5]. In the penetrating trauma domain, approaches to providing computerized decision support have included rule-based systems [6], methods that incorporate 3D

models of anatomy [7–10], and methods that utilize machine learning with neural networks [11].

The assessment of injury to vital organs as a result of penetrating trauma is a complex clinical problem that requires fusing knowledge of human anatomy, physiology, and patient signs and symptoms. Complicated spatial and physiologic relationships need to be assessed rapidly in the setting of incomplete clinical information. Simultaneous injury to multiple organ-systems resulting in several conditions adds to the complexity of this task. Computerized decision support systems that encode the anatomic and physiological relationships necessary for penetrating trauma assessment may help to increase diagnostic accuracy during the early part of trauma resuscitation, before a patient is stabilized and definitive imaging studies can be performed. Well-established

* Corresponding author. Fax: +1 310 631 1495.

E-mail addresses: bilal.ahmed@utoronto.ca (B.A. Ahmed), lolaogunyemi@cdrewu.edu (O.I. Ogunyemi).

databases that provide detailed information linking external wound locations, organ injuries, detailed patient findings and injury outcomes for the penetrating trauma domain are not widely available. While the National Trauma Data Bank and similar trauma registries collate demographic information [12], Glasgow Coma scores, and diagnostic codes, they lack detailed information on patient findings (such as external wound locations, breath sounds, heart sounds, jugular vein distension, bullet locations from xray, hemoptysis, stridor, tenderness, guarding, rebound tenderness, ileus, hematuria, arm pulse strength, leg pulse strength, and guaic status) that facilitate deeper reasoning about trauma consequences. As a result, most of the existing computerized models for penetrating trauma decision support are not derived directly from such data registries.

Several studies have shown the utility of computerized decision support in predicting outcomes in the penetrating trauma domain. Hirshberg et al. studied neural networks as a means of predicting the need for damage control in abdominal gunshot injuries [11]. Their study shows that bullet trajectory information and blood pressure findings are important predictors of outcomes for patients with abdominal gunshot wounds. To reason about penetrating trauma for the Virtual Soldier project, Rubin et al. at Stanford utilize patient-specific 3D models of anatomy from CT-data [9,10]. They link these geometric models to the foundational model of anatomy [13], a comprehensive anatomical ontology developed at the University of Seattle, Washington, and also develop a reference ontology of regional perfusion for the heart in order to deduce injuries to anatomic structures that may not be visible in the segmented CT images. A related effort by Pouchard and Dickson also creates ontology-enriched visualization of human anatomy using the foundational model [14]. The Stanford project is an impressive effort to utilize canonical knowledge to reason from first-principles about the effects of penetrating trauma. As they state, their approach is a deterministic one that does not account for uncertainty associated with injury.

In other related works, Walczak utilized a back-propagation artificial neural network model to predict the blood transfusion requirements of trauma patients in an emergency room [15]. Shoemaker et al. applied a probabilistic decision support system to predict outcomes and the effects of therapy in 396 patients with severe thoracic and thoraco-abdominal injuries [16].

Between 2002 and 2006, we created a detailed thoraco-abdominal penetrating trauma dataset containing gunshot and stab injury related information on 637 patients from three hospitals: Brigham & Women's Hospital, Massachusetts General Hospital, and MCP-Hahnemann Hospital. In this paper, we compare Bayesian network machine learning approaches for penetrating trauma diagnosis derived from this dataset to TraumaSCAN-Web, which combines 3D geometric reasoning and expert-derived Bayesian networks (network structure and probabilities are specified by experts).

TraumaSCAN-Web (TSW) [7] is a computerized decision support system for assessing chest and abdominal penetrating trauma. It computes the probability of injury to anatomic structures and the probability of consequent conditions, such as pneumothoraces and hemothoraces, using geometric reasoning about anatomic structure involvement and probabilistic reasoning with Bayesian networks. Given the absence of detailed datasets for reasoning about penetrating trauma consequences at the time it was being developed, TSW's Bayesian network was derived using subjective probabilities from a trauma expert (JRC).

The goal of the present study is to determine whether Bayesian networks with predefined structure and probabilities learned from penetrating trauma data are comparable in diagnostic accuracy to TSW, which utilizes 3D geometric reasoning and subjective probabilities obtained from an expert traumatologist. It is well known that accurate structure learning is difficult, especially in the presence of sparse/missing data [17]. In Bayesian networks, the network structure is represented visually by a directed acyclic graph that encodes the causal and associational relationships among domain variables. This visual structure can be readily critiqued by experts, especially in domains where relationships are well defined and documented in the literature. Often, in such domains, learned structures are assessed as flawed by experts: containing spurious connections that have no real-world analogue or missing connections between nodes that have a known association. Penetrating trauma is a domain in which the consequences of disruption to the normal functioning of organs and organ-systems by projectiles/stabbing are well understood and documented in the literature. As demonstrated by the Stanford study previously referenced [9,10], using this knowledge, models that reason from first-principles about such disruptions have been developed. Since we

Table 1
Variables used to derive Bayesian Network models

Organ Injuries	Conditions	Patient findings	External Wound Locations*
Heart	Pericardial tamponade	Muffled heart sounds Hemodynamic shock	Right/left chest wound RUQ/RLQ/LUQ/LLQ wound
Right/left lung	Right/left hemothorax Right/left simple pneumothorax	Decreased breath sounds	RLLQ/LLLQ wound
Esophagus	Right/left tension pneumothorax	Tracheal deviation	Right/left flank wound
Descending aorta	Non specific intra-abdominal injury		Right/left back wound
Tracheobronchial tree		Distended neck veins	
Diaphragm	Right/left renal injury	Odynophagia	
Liver		Blood in nasogastric tube	
Stomach		Hemoptysis	
Intestines		Stridor	
Spleen		Weak arm pulses	
Pancreas		Weak leg pulses	
Right/left kidney		Positive X-ray for laceration of diaphragm Abdominal tenderness, guarding, rebound tenderness, ileus Abdominal distension Blood lavage Free air Positive renal IVP Hematuria	

* Not included in Bayesian network 2.

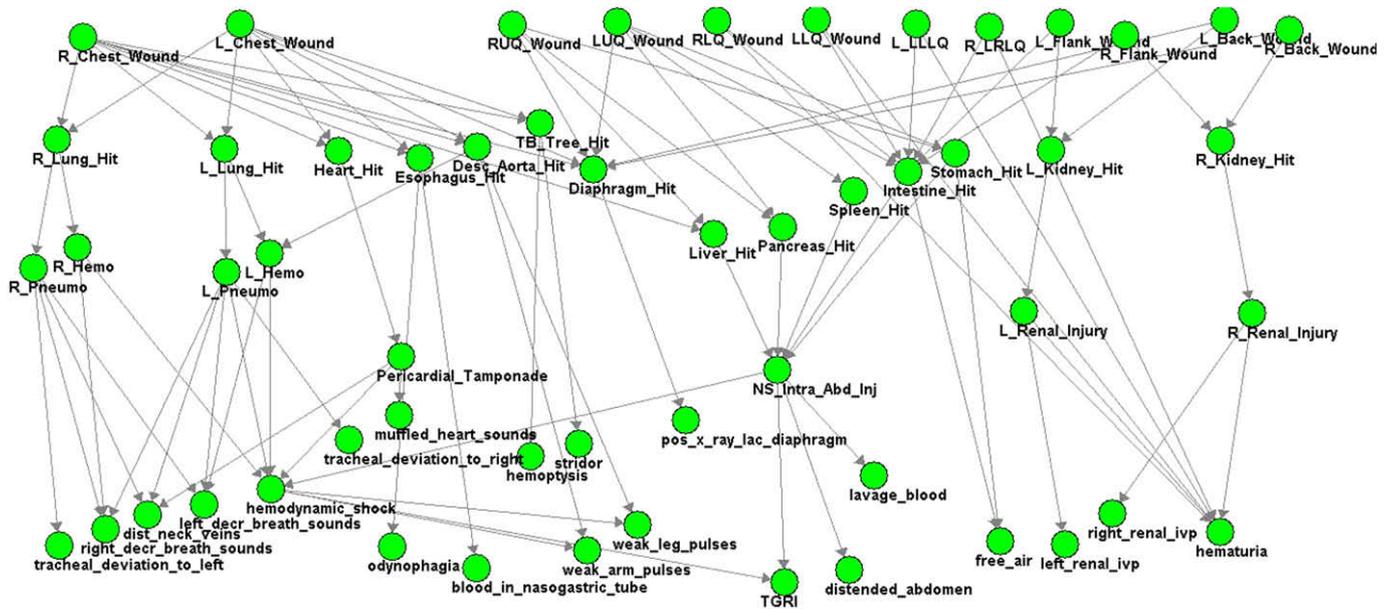


Fig. 2. Bayesian Network 1: Updated Bayesian Network for machine learning approach including external wound location nodes.

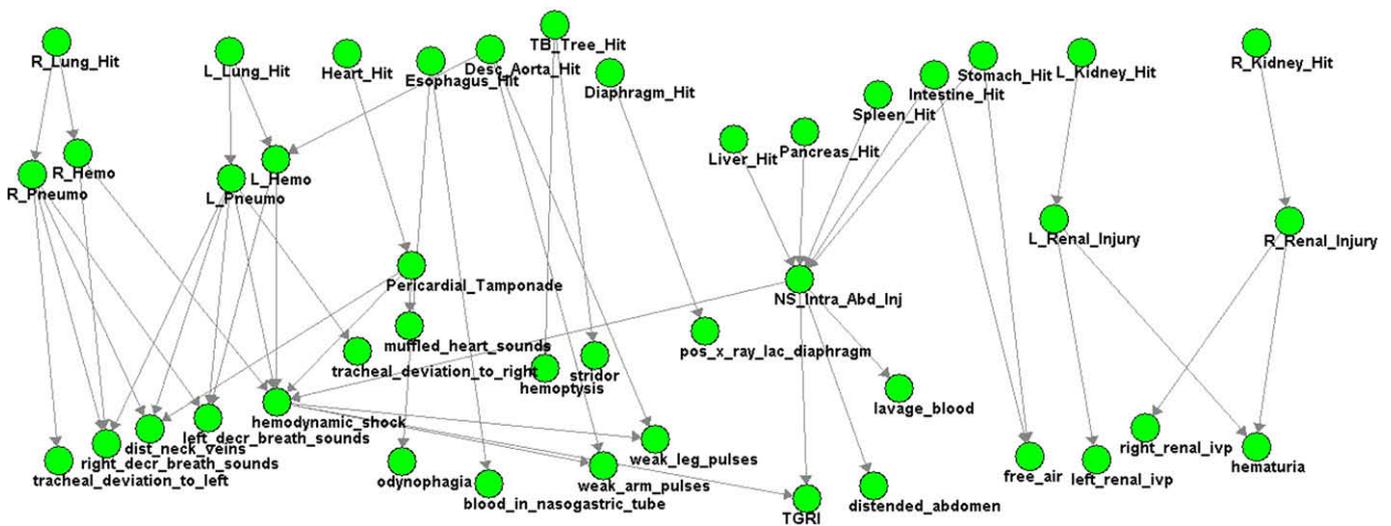


Fig. 3. Bayesian Network 2: Updated Bayesian Network for machine learning approach excluding external wound location nodes.

were unknown or were simply not recorded) and were marked accordingly.

Of these data, a total of 81 cases inappropriate for the study were excluded for a variety of reasons. These include:

- (a) Cases with no external wound locations in the thoracoabdomen and pelvis regions when the external wound location was peripheral to the thoracoabdomen and pelvis regions (neck, shoulder, and leg).
- (b) Cases with an inadequate amount of recorded information. This included acutely fatal cases where many less important signs, symptoms, and diagnoses were not recorded due to the nature of the overarching condition (i.e., internal anatomic injuries were not verified clinically in some patients who were dead on arrival or died shortly after arrival in the ED).
- (c) Cases with a missing or unclear case wound diagram or an irresolvable inconsistency between the diagram and the accompanying note.

- (d) Shotgun cases and cases with extensive bullet fragmenting, as they were never intended to be evaluated by TSW.

2.3. Training and test datasets and evaluation measures

The modified Bayesian network structures were stored in the Bayesian Interchange Format for use with Weka, an open-source machine learning toolkit [19]. Two distinct learning studies were performed. The first study involved parameter learning and assessing the performance of the modified Bayesian networks on penetrating trauma data from three hospitals. In the second study, we performed parameter learning by partitioning the data into training sets using data from three hospitals and test sets using data from two hospitals, in order to facilitate direct comparison with results from a previously published study.

In the first study, we combined data collected from a new hospital (MGH) with existing data from two hospitals (BWH, MCP). Using the two modified Bayesian network structures, we learned

Table 2
Summary of AUC results from 10-fold cross-validation

Diagnosis	Events	AUC (95% CI) Network 1	AUC (95% CI) network 2
Right lung injury	91	0.994 (0.983–1.000)	0.994 (0.985–1.000)
Left lung injury	126	0.996 (0.992–1.000)	0.982 (0.963–1.000)
Heart injury	31	0.794 (0.728–0.860)	0.543 (0.426–0.660)
Esophagus injury	0	–	–
Tracheal/TB tree injury	1	0.767 (0.500–1.000)	0.767 (0.500–1.000)
Descending aorta injury	8	0.582 (0.311–0.853)	0.525 (0.262–0.788)
Diaphragm injury	60	0.550 (0.469–0.630)	0.503 (0.426–0.581)
Liver injury	82	0.920 (0.897–0.942)	0.902 (0.877–0.927)
Intestinal injury	111	0.928 (0.908–0.947)	0.903 (0.879–0.926)
Stomach injury	44	0.831 (0.791–0.870)	0.829 (0.789–0.868)
Spleen injury	28	0.805 (0.734–0.876)	0.817 (0.749–0.886)
Pancreas injury	16	0.772 (0.683–0.861)	0.780 (0.703–0.856)
Left Kidney injury	21	1.000 (1.000–1.000)	1.000 (1.000–1.000)
Right kidney injury	12	1.000 (1.000–1.000)	1.000 (1.000–1.000)
Right hemothorax	47	0.964 (0.950–0.978)	0.964 (0.950–0.978)
Left hemothorax	50	0.945 (0.925–0.964)	0.945 (0.925–0.964)
Right simple pneumothorax	71	0.953 (0.915–0.990)	0.953 (0.915–0.990)
Left simple pneumothorax	95	0.971 (0.954–0.987)	0.971 (0.954–0.987)
Right tension pneumothorax	1	0.966 (0.952–0.980)	0.966 (0.952–0.980)
Left tension pneumothorax	1	0.969 (0.955–0.982)	0.969 (0.955–0.982)
Pericardial tamponade	6	0.989 (0.981–0.997)	0.989 (0.981–0.997)
NS IA injury	255	0.917 (0.891–0.944)	0.917 (0.891–0.944)
Right renal injury	12	1.000 (1.000–1.000)	1.000 (1.000–1.000)
Left renal injury	21	1.000 (1.000–1.000)	1.000 (1.000–1.000)

prior or conditional probabilities for each network node. The resulting networks were used to evaluate each of the 24 diagnoses listed in Table 2 using 10-fold cross-validation.

For the second study, we performed parameter learning and diagnostic accuracy estimation with a focus on evaluating gunshot and stab cases that had been previously assessed using TSW. Since TSW utilizes subjective probabilities supplied by an expert within its Bayesian network, this allows for a comparison of results derived in part from subjective probabilities with results from Bayesian networks that utilize parameter learning. To achieve this, we split the data into training and test sets: for gunshot injuries, data was split into 557 training cases and 80 test cases, the same 80 as were published in Matheny et al. [18], and for stab injuries, data were split into 523 training cases and 114 test cases, the same 114 as published previously. Weka was used to learn parameters and the AUC for each diagnosis was calculated. Finally, pair-wise comparisons were made between the empiric models and TraumaSCAN-Web using the Hanley–McNeil test [20]. A Bonferroni correction was applied to account for the multiple comparisons, giving $p < 0.0031$ and $p < 0.0028$ to indicate statistical significance for stab and gunshot cases, respectively.

3. Results

3.1. 10-fold cross-validation study

The results from the 10-fold cross-validation were used to calculate AUCs for the 24 possible diagnoses modeled by the two Bayesian networks (see Table 2). Esophageal injury could not be evaluated due to a lack of occurrences of that injury in the new dataset. For each condition, AUCs with the 95% confidence intervals (95% CI) are reported in Table 2. Discrimination estimates for thoracic injuries ranged from 0.953 to 0.994 for lung injuries, 0.767 for tracheobronchial tree injuries, and 0.543 to 0.989 for heart-related injuries. Injuries to the diaphragm and descending aorta ranged from 0.503 to 0.550 and 0.525 to 0.582, respectively. Non-specific intra-abdominal injuries were 0.917 while other intra-abdominal solid organ injury estimates ranged from 0.772 to 0.928. There were no significant differences in AUCs with the exception of heart

injury, for which Network 1, which includes external wound locations performed considerably better than Network 2, which does not.

3.2. Comparison study to TSW

Pairwise comparisons of the area under the ROC curve for stab injury data show that diagnostic accuracies of the learned models were greater than TSW for 3 out of 16 outcomes ($p < 0.0031$ indicating statistical significance, based on the Bonferroni correction), namely right and left lung injuries and right hemothorax. TSW's prediction for diaphragm and heart injuries were superior to the learned networks' estimates; however, these were not statistically significant. The learned models performed similarly to or better than TSW for all other estimates, without statistically significant differences. Six conditions that did not occur in the previously collected dataset were not compared.

For gunshot injury data, the learned networks performed better than or equal to TSW for the majority of outcomes, while TSW had superior performance for diaphragm, heart and descending aorta injuries. However, none of these differences were statistically significant differences ($p < 0.0028$ based on the Bonferroni correction). For conditions that did not occur in the previously collected dataset were not compared.

4. Discussion

In this study, we demonstrate that a trauma diagnosis approach using Bayesian networks with predefined structures and probabilities learned from penetrating trauma data is comparable to, and in some instances better than TraumaSCAN-Web in diagnostic accuracy based on areas under the ROC curve. The Bayesian networks performed very well in discriminating between the various thoraco-abdominal injuries evaluated using 10-fold cross-validation. For the 23 conditions evaluated, 16 conditions had AUCs over 0.90 while 21 conditions had AUCs over 0.75 for Network 1. For Network 2, 16 conditions had AUCs over 0.90 and 20 conditions had AUCs over 0.75. Lung and renal injuries were among the best estimations while diaphragm and descending aorta injury estimations

Table 3a

Comparison of AUC results for Bayesian Network 1 with AUC results for TraumaSCAN-Web on 114 stab wound cases

Diagnosis	Number of events	AUC (95%CI) TSW/expert	AUC learned Network 1	Pairwise <i>p</i> Network 1
Right lung injury	21	0.887 (0.823–0.950)	1.000 (1.000–1.000)	0.0005
Left lung injury	18	0.910 (0.857–0.962)	0.990 (0.972–1.000)	0.0018
Heart Injury	8	0.975 (0.937–1.000)	0.851 (0.741–0.961)	0.0305
Esophagus Injury	0	–	–	–
Tracheal/TB tree injury	1	1.000 (1.000–1.000)	1.000 (1.000–1.000)	1.000
Descending aorta injury	0	–	–	–
Diaphragm Injury	9	0.830 (0.721–0.938)	0.601 (0.438–0.764)	1.000
Liver injury	12	0.833 (0.721–0.946)	0.893 (0.772–1.000)	0.4643
Intestinal injury	11	0.701 (0.541–0.860)	0.942 (0.899–0.984)	0.0038
Stomach injury	4	0.814 (0.724–0.904)	0.863 (0.751–0.974)	1.0000
Left kidney injury	0	–	–	–
Right kidney injury	4	0.805 (0.482–1.000)	1.000 (1.000–1.000)	0.2353
Right hemothorax	12	0.810 (0.720–0.900)	0.953 (0.916–0.990)	0.0010
Left hemothorax	5	0.899 (0.827–0.972)	0.877 (0.783–0.971)	0.5785
Right simple pneumothorax	16	0.884 (0.811–0.957)	0.978 (0.956–1.000)	0.0103
Left simple pneumothorax	18	0.934 (0.889–0.979)	0.991 (0.978–1.000)	0.0105
Right tension pneumothorax	0	–	–	–
Left tension pneumothorax	0	–	–	–
Pericardial tamponade	2	0.978 (0.946–1.000)	0.964 (0.929–1.000)	1.0000
NS IA injury	36	0.755 (0.658–0.852)	0.894 (0.821–0.967)	0.0095
Right renal injury	4	0.809 (0.489–1.000)	1.000 (1.000–1.000)	0.2424
Left renal injury	0	–	–	–

were the poorest. Thoracic and retroperitoneal diagnoses generally outperformed abdominal diagnoses.

The comparison of the learned Bayesian networks to TSW shows that the learned models significantly outperformed TSW consistently for lung injuries, including hemothoraces. Other thoracic and abdominal diagnoses were also more accurately discriminated by the learned models, but the differences were not statistically significant.

The predictive accuracy of Network 1 (the graph with nodes representing external wound locations) is generally very similar to Network 2 (the graph without external wound location nodes), suggesting that for many diagnoses, information about patient signs, symptoms, and test results is sufficient for reasoning about penetrating trauma consequences. Diagnoses for which there appears to be a significant advantage to having external wound location information are heart injury, and to a lesser extent, descending aorta injury (see Table 2). No broad conclusions on diagnostic accuracy can be drawn from results in Tables 3a and b that show a

perfect AUC—for many of these cases, there was just one instance of a particular type of injury in the test set (Tables 4a and b).

Heart and diaphragm injuries are two diagnoses that were discriminated better by TraumaSCAN-Web than the learned Bayesian network models for stab or gunshot wounds. Diaphragm injury was significantly better discriminated by TSW for gunshot injury cases, although the improvement in discrimination was not statistically significant for stab injury cases. Accurately diagnosing diaphragmatic injury is generally considered to be difficult given that patients may remain asymptomatic with this type of injury [21–25]. Furthermore, imaging studies (CT scans, in particular) often miss diaphragmatic rupture [26]. While magnetic resonance imaging is shown to be more accurate [27], this imaging modality has less utility in acute trauma settings. It is therefore not surprising that a machine learning approach that relies only on clinical features and some imaging investigations (specifically X-rays for diaphragmatic rupture) and did not utilize any anatomical model would perform worse for this type of injury.

Table 3b

Comparison of AUC results for Bayesian network 2 with AUC results for TraumaSCAN-Web on 114 stab wound cases

Diagnosis	Number of events	AUC (95%CI) TSW/expert	AUC learned network 2	Pairwise <i>p</i> network 2
Right lung injury	21	0.887 (0.823–0.950)	1.000 (1.000–1.000)	0.0005
Left lung injury	18	0.910 (0.857–0.962)	0.990 (0.972–1.000)	0.0019
Heart injury	8	0.975 (0.937–1.000)	0.625 (0.391–0.859)	0.0036
Esophagus injury	0	–	–	–
Tracheal/TB tree injury	1	1.000 (1.000–1.000)	1.000 (1.000–1.000)	1.0000
Descending aorta injury	0	–	–	–
Diaphragm injury	9	0.830 (0.721–0.938)	0.505 (0.309–0.701)	1.0000
Liver injury	12	0.833 (0.721–0.946)	0.883 (0.778–0.988)	0.5082
Intestinal injury	11	0.701 (0.541–0.860)	0.918 (0.866–0.971)	1.0000
Stomach injury	4	0.814 (0.724–0.904)	0.876 (0.779–0.973)	1.0000
Left kidney injury	0	–	–	–
Right kidney injury	4	0.805 (0.482–1.000)	1.000 (1.000–1.000)	1.0000
Right hemothorax	12	0.810 (0.720–0.900)	0.953 (0.916–0.990)	0.0010
Left hemothorax	5	0.899 (0.827–0.972)	0.877 (0.783–0.971)	0.5785
Right simple pneumothorax	16	0.884 (0.811–0.957)	0.978 (0.956–1.000)	0.0103
Left simple pneumothorax	18	0.934 (0.889–0.979)	0.991 (0.978–1.000)	0.0105
Right tension pneumothorax	0	–	–	–
Left tension pneumothorax	0	–	–	–
Pericardial tamponade	2	0.978 (0.946–1.000)	0.964 (0.929–1.000)	1.0000
NS IA injury	36	0.755 (0.658–0.852)	0.894 (0.821–0.967)	0.0095
Right renal injury	4	0.809 (0.489–1.000)	1.000 (1.000–1.000)	0.2424
Left renal injury	0	–	–	–

Table 4a
Comparison of AUC results for Bayesian Network 1 with AUC results for TraumaSCAN-Web on 80 gunshot wound cases

Diagnosis	Number of events	AUC (95%CI) TSW/expert	AUC learned Network 1	Pairwise <i>p</i> Network 1
Right lung injury	11	0.910 (0.819–1.000)	0.993 (0.978–1.000)	1.0000
Left lung injury	19	0.915 (0.849–0.982)	1.000 (1.000–1.000)	0.0126
Heart injury	4	0.720 (0.459–0.982)	0.673 (0.348–0.997)	1.0000
Esophagus injury	0	–	–	–
Tracheal/TB tree injury	0	–	–	–
Descending aorta injury	2	0.814 (0.645–0.983)	0.734 (0.632–0.836)	0.9930
Diaphragm injury	12	0.857 (0.726–0.988)	0.599 (0.416–0.782)	0.0171
Liver injury	10	0.739 (0.593–0.884)	0.895 (0.822–0.968)	0.7390
Intestinal injury	18	0.811 (0.704–0.918)	0.944 (0.897–0.991)	0.0161
Stomach injury	13	0.898 (0.821–0.975)	0.871 (0.794–0.948)	0.5875
Left kidney injury	4	0.944 (0.887–1.000)	1.000 (1.000–1.000)	1.0000
Right kidney injury	1	0.975 (0.940–1.000)	1.000 (1.000–1.000)	1.0000
Right hemothorax	8	0.870 (0.790–0.950)	0.982 (0.954–1.000)	0.0057
Left hemothorax	13	0.893 (0.791–0.996)	0.974 (0.945–1.000)	0.0732
Right simple pneumothorax	6	0.881 (0.793–0.968)	0.982 (0.955–1.000)	0.0134
Left simple pneumothorax	13	0.845 (0.737–0.953)	0.966 (0.930–1.000)	0.0219
Right tension pneumothorax	0	–	–	–
Left tension pneumothorax	0	–	–	–
Pericardial tamponade	2	0.519 (0.111–0.927)	0.683 (0.206–1.000)	0.8330
NS IA injury	34	0.875 (0.793–0.957)	0.927 (0.852–1.000)	0.3118
Right renal injury	1	0.975 (0.940–1.000)	1.000 (1.000–1.000)	1.0000
Left renal injury	4	0.947 (0.891–1.000)	1.000 (1.000–1.000)	0.0695

There were a number of limitations of the data collected that resulted in some missing data and may therefore have reduced the predictive accuracy of the learned network. The data used for machine learning was collected from three different hospitals in two states, recorded by a large number of health care professionals, and collected by three different individuals. This potentially introduces sources of error and inconsistency. First, there is an inherent variability in how different physicians record medical information on discharge summaries and case wound diagrams. Some physicians may not explicitly mention negative signs and symptoms, even if they are pertinent negatives and the patient was examined for them, assuming that the absence of a record can safely imply that the sign or symptom is negative.

Cases of extensive trauma, especially those that resulted in the patient's death, were another source of missing data. For example, in cases where the heart was hit, there is often little explicit information on less pertinent matters like weakened peripheral pulses. Some of these cases had such extensive missing data (almost nothing

recorded beyond the fact that the patient expired) that they were inappropriate for the study and had to be excluded. Thus the learned models could not benefit from some of these interesting cases. This experience shows that there are instances in which expert physicians can significantly contribute to the knowledge base based on the knowledge they have acquired from handling such cases.

Several of the trauma cases involved injury to other parts of the body in addition to the thoracoabdomen. While this did not affect most of the signs and symptoms, some cases inevitably introduced "noise" in the data as a result of this. For example, a case that involved a gunshot wound to the back and the wrist resulted in a weak right radial pulse, causing an incorrect association between a back wound and radial pulses in the absence of the contextual knowledge of a gunshot wound to the wrist.

A distinct advantage of the new Bayesian network models over TraumaSCAN-Web is that they can handle cases when the internal locations of any bullets are not yet known. This is important be-

Table 4b
Comparison of AUC results for Bayesian network 2 with AUC results for TraumaSCAN-Web on 80 gunshot wound cases

Diagnosis	Number of events	AUC (95%CI) TSW/expert	AUC learned network 2	Pairwise <i>p</i> network 2
Right lung injury	11	0.910 (0.819–1.000)	0.997 (0.988–1.000)	1.0000
Left lung injury	19	0.915 (0.849–0.982)	1.000 (1.000–1.000)	1.0000
Heart injury	4	0.720 (0.459–0.982)	0.618 (0.291–0.945)	1.0000
Esophagus injury	0	–	–	–
Tracheal/TB tree injury	0	–	–	–
Descending aorta injury	2	0.814 (0.645–0.983)	0.526 (0.000–1.000)	1.0000
Diaphragm injury	12	0.857 (0.726–0.988)	0.542 (0.355–0.729)	1.0000
Liver injury	10	0.767 (0.601–0.933)	0.909 (0.844–0.974)	0.1138
Intestinal injury	18	0.811 (0.704–0.918)	0.901 (0.835–0.967)	0.1149
Stomach injury	13	0.898 (0.821–0.975)	0.869 (0.792–0.946)	0.5999
Left kidney injury	4	0.944 (0.887–1.000)	1.000 (1.000–1.000)	1.0000
Right kidney injury	1	0.975 (0.940–1.000)	1.000 (1.000–1.000)	1.0000
Right hemothorax	8	0.870 (0.790–0.950)	0.982 (0.954–1.000)	0.0057
Left hemothorax	13	0.893 (0.791–0.996)	0.974 (0.945–1.000)	0.0732
Right simple pneumothorax	6	0.881 (0.793–0.968)	0.982 (0.955–1.000)	0.0134
Left simple pneumothorax	13	0.845 (0.737–0.953)	0.966 (0.930–1.000)	0.0219
Right tension pneumothorax	0	–	–	–
Left tension pneumothorax	0	–	–	–
Pericardial tamponade	2	0.519 (0.111–0.927)	0.683 (0.206–1.000)	1.0000
NS IA injury	34	0.875 (0.793–0.957)	0.927 (0.852–1.000)	0.3118
Right renal injury	1	0.975 (0.940–1.000)	1.000 (1.000–1.000)	1.0000
Left renal injury	4	0.947 (0.891–1.000)	1.000 (1.000–1.000)	0.0695

Appendix A

Mapping of Bayesian Network abbreviations to clinical descriptions used in Results tables

Bayesian network node label	Label used in table	Description*
R_Lung_Hit	Right lung injury	Potential traumatic injury to the right lung
L_Lung_Hit	Left lung injury	Potential traumatic injury to the left lung
Heart_Hit	Heart injury	Potential traumatic injury to the heart
Esophagus_Hit	Esophagus injury	Potential traumatic injury to the esophagus
Desc_Aorta_Hit	Descending aorta injury	Potential traumatic injury to the descending aorta
TB_Tree_Hit	Trachea/tracheobronchial tree injury	Potential traumatic injury to trachea/tracheobronchial tree
Diaphragm_Hit	Diaphragm injury	Potential traumatic injury to the diaphragm
Spleen_Hit	Spleen injury	Potential traumatic injury to the spleen
Liver_Hit	Liver injury	Potential traumatic injury to the liver
Pancreas_Hit	Pancreas injury	Potential traumatic injury to the pancreas
Intestine_Hit	Intestine injury	Potential traumatic injury to the intestine
Stomach_Hit	Stomach injury	Potential traumatic injury to the stomach
L_Kidney_Hit	Left kidney injury	Potential traumatic injury to the left kidney
R_Kidney_Hit	Right kidney injury	Potential traumatic injury to the right kidney
R_Pneumo	Right pneumothorax	Potential consequence of right lung injury; mutually exclusive states in Bayesian network node are simple, tension, or no pneumothorax
R_Hemo	Right hemothorax	Potential consequence of right lung injury
L_Pneumo	Left pneumothorax	Potential consequence of left lung injury; mutually exclusive states in Bayesian network node are simple, tension, or no pneumothorax
L_Hemo	Left hemothorax	Potential consequence of left lung injury
Pericardial_Tamponade	Pericardial tamponade	Potential consequence of heart injury
NS_Intra_Abd_Inj	Non-specific intra-abdominal injury	Potential consequence of injury to one or more abdominal area organs
L_Renal_Injury	Left renal injury	Potential consequence of left kidney injury
R_Renal_Injury	Right renal injury	Potential consequence of right kidney injury

*Nodes have yes/no or true/false states unless otherwise noted (e.g., pneumothorax).

cause the 3D modeling approach used by TSW requires pairing up entry and exit wounds or entry wounds with bullets (the sum of external wounds and bullets must therefore be even), excluding cases where external wounds are known but the internal locations of bullets are not. The ability to make relatively accurate penetrating trauma predictions using just those patient findings that can be obtained without in-hospital tests (e.g., decreased breath sounds, muffled heart sounds, distended neck veins, etc.) is a key step for creating a decision support tool that could be used by Emergency Medicine Technicians (EMTs) in a pre-hospital setting. Of course, the accuracy of a Bayesian network-only solution depends a great deal on the amount of patient findings provided, while TSW's does not. However, the results obtained using Network 2 suggest that predictive power may still be decent for certain traumatic conditions in the absence of information on external wound location. If these results are replicated in further studies, it increases the number of viable strategies that can be employed for developing a pre-hospital decision support solution.

An advantage of TSW is that it provides greater control in specifying wound locations and can more accurately specify the 3D space traversed by the wound. This not only avoids the lack of spatial specificity encountered in the machine learning approach, but also overcomes the difficulty of describing whether a wound is superficial or deep (and how this affects the probability of organ injury). Furthermore, by providing a wound diagram for the user to mark, TSW avoids the fuzzy boundaries of the anatomical structures that were encountered during data collection.

In this study, we have shown that for many trauma diagnoses, Bayesian Networks with expert-defined structures and parameter learning performed as well as or better than an approach that combines 3D geometric reasoning with a Bayesian network derived from subjective expert probabilities. To the extent that this helps to address the problem of requiring wound-wound or bullet-wound pairing in TSW, these results demonstrate a potential solution for penetrating trauma decision support in the pre-hospital setting,

which we will investigate further. A parameter learning approach to Bayesian networks is promising in the domain of penetrating trauma injury and future studies are warranted to explore the combination of expert knowledge with knowledge learned from data in other domains of medical decision support.

Acknowledgments

This study was supported by the National Library of Medicine under Grants 1R01LM07167 and 1T15LM07092, and the CREMS Summer Program Award at the University of Toronto.

References

- [1] Aronsky D, Haug PJ. Diagnosing community-acquired pneumonia with a Bayesian network. *Proc AMIA Symp* 1998;632–6.
- [2] Burnside E, Rubin D, Shachter R. A Bayesian network for mammography. *Proc AMIA Symp* 2000;106–10.
- [3] Cooper GF. A Computer-based medical diagnostic aid that integrates causal and probabilistic knowledge. PhD thesis, Stanford, CA: Stanford University; 1984.
- [4] de Dombal FT. Computer-assisted diagnosis of abdominal pain. In: *Advances in medical computing: proceedings of the third international symposium on computers in medicine*; 1975. p. 10–9.
- [5] Heckerman DE, Horvitz EJ, Nathwani BN. Toward normative expert systems: the pathfinder project. Technical Report. Stanford, CA: Knowledge Systems Laboratory, Stanford University; 1991 [Report No. KSL-91-44].
- [6] Webber B, Rymon R, Clarke JR. Flexible support for trauma management through goal directed reasoning and planning. *Artif Intell Med* 1992;4(2):145–63.
- [7] Ogunyemi O. Methods for reasoning from geometry about anatomic structures injured by penetrating trauma. *J Biomed Inform* 2006;39(4):389–400.
- [8] Ogunyemi OI, Clarke JR, Ash N, Webber BL. Combining geometric and probabilistic reasoning for computer-based penetrating-trauma assessment. *J Am Med Inform Assoc* 2002;9(3):273–82.
- [9] Rubin DL, Bashir Y, Grossman D, Dev P, Musen MA. Using an ontology of human anatomy to inform reasoning with geometric models. *Stud Health Technol Inform* 2005;111:429–35.
- [10] Rubin DL, Dameron O, Bashir Y, Grossman D, Dev P, Musen MA. Using ontologies linked with geometric models to reason about penetrating injuries. *Artif Intell Med* 2006;37(3):167–76.

- [11] Hirshberg A, Wall Jr MJ, Mattox KL. Bullet trajectory predicts the need for damage control: an artificial neural network model. *J Trauma* 2002;52(5):852–58.
- [12] The National Trauma Data Bank. Available from: <http://www.facs.org/trauma/ntdb.html> [accessed 28.01.07].
- [13] Rosse C, Mejino Jr JL. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003;36(6):478–500.
- [14] Pouchard LC, Dickson SP. Ontology-enriched Visualization of Human Anatomy: Computer Science and Mathematics Division, Oak Ridge National Laboratory; 2005 [Report No. ORNL/TM-2005/113].
- [15] Walczak S. Artificial neural network medical decision support tool: predicting transfusion requirements of ER patients. *IEEE Trans Inf Technol Biomed* 2005;9(3):468–74.
- [16] Shoemaker WC, Bayard DS, Wo CC, Chan LS, Chien LC, Lu K, et al. Outcome prediction in chest injury by a mathematical search and display program. *Chest* 2005;128(4):2739–48.
- [17] Neapolitan RE. Learning Bayesian Networks. In: Russell S, Norvig P, editors. Upper saddle river, NJ: Pearson-Prentice Hall; 2004.
- [18] Matheny ME, Ogunyemi OI, Rice PL, Clarke JR. Evaluating the discriminatory power of a computer-based system for assessing penetrating trauma on retrospective multi-center data. *AMIA Annu Symp Proc* 2005:500–4.
- [19] Witten I, Frank E. Data mining: practical machine learning tools and techniques. 2nd ed. Morgan Kaufmann; 2005.
- [20] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148(3):839–43.
- [21] Murray JA, Demetriades D, Asensio JA, Cornwell 3rd EE, Velmahos GC, Belzberg H, et al. Occult injuries to the diaphragm: prospective evaluation of laparoscopy in penetrating injuries to the left lower chest. *J Am Coll Surg* 1998;187(6):626–30.
- [22] Nel JH, Warren BL. Thoracoscopic evaluation of the diaphragm in patients with knife wounds of the left lower chest. *Br J Surg* 1994;81(5):713–4.
- [23] Robin AP, Andrews JR, Lange DA, Roberts RR, Moskal M, Barrett JA. Selective management of anterior abdominal stab wounds. *J Trauma* 1989;29(12):1684–9.
- [24] Leppaniemi A, Haapiainen R. Occult diaphragmatic injuries caused by stab wounds. *J Trauma* 2003;55(4):646–50.
- [25] Dissanaik S, Griswold JA, Frezza EE. Treatment of isolated penetrating flank trauma. *Am Surg* 2005;71(6):493–6.
- [26] Reber PU, Schmied B, Seiler CA, Baer HU, Patel AG, Buchler MW. Missed diaphragmatic injuries and their long-term sequelae. *J Trauma* 1998;44(1):183–8.
- [27] Shanmuganathan K, Mirvis SE, White CS, Pomerantz SM. MR imaging evaluation of hemidiaphragms in acute blunt trauma: experience with 16 patients. *AJR Am J Roentgenol* 1996;167(2):397–402.