

Original Article

Open Access



# Bayesian process mapping for intraoperative decision making in laparoscopic cholecystectomy

Isaac Tranter-Entwistle<sup>1</sup> , Bill Wilson<sup>2</sup> , Tim Eglinton<sup>1,3</sup>, Saxon Connor<sup>3</sup> 

<sup>1</sup>Department of Surgery and Critical Care, The University of Otago Christchurch, Christchurch 8140, New Zealand.

<sup>2</sup>Te Whatu Ora Waitaha, Christchurch, Canterbury 8140, New Zealand.

<sup>3</sup>Department of General Surgery Christchurch Hospital, Christchurch, Canterbury 4710, New Zealand.

**Correspondence to:** Dr. Isaac Tranter-Entwistle, Department of Surgery and Critical Care, The University of Otago Christchurch, 36 Cashel Street, Christchurch Central, Christchurch 8140, New Zealand. E-mail: tranter.isaac@gmail.com

**How to cite this article:** Tranter-Entwistle I, Wilson B, Eglinton T, Connor S. Bayesian process mapping for intraoperative decision making in laparoscopic cholecystectomy. *Art Int Surg* 2024;4:77-91. <https://dx.doi.org/10.20517/ais.2024.04>

**Received:** 17 Jan 2024 **First Decision:** 26 Mar 2024 **Revised:** 5 Jun 2024 **Accepted:** 5 Jun 2024 **Published:** 21 Jun 2024

**Academic Editor:** Andrew A. Gumbs **Copy Editor:** Dong-Li Li **Production Editor:** Dong-Li Li

## Abstract

**Aim:** Surgeon's intraoperative decisions significantly impact patient outcomes. In the reconciliation cycle, interoperative decisions are guided by probabilistic reasoning, which is informed by the evolving intraoperative features. This paper aims to compare the utility of a traditional logistic regression (LR) model for critical view of safety (CVS) achievement to Bayesian network (BN) maps using intraoperative features. It hypothesizes that BN mapping better integrates with surgeon heuristics.

**Methods:** Using prospectively gathered intraoperative data, BN maps were developed and tested to determine their ability to predict critical view of safety achievement. Performance was compared to traditional logistic regression models to consider their utility in practice.

**Results:** In total, 4,663 patients were identified. Of these patients, 2,837 (61%) presented acutely and 3,122 (67%) were female. CVS was achieved in 4,131 (92%) of patients. In total, four BN were developed. Optimal performance was seen in model 2 with an AUC of 0.879 (0.798-0.960) ( $P < 0.001$ ). Selecting a cut-off of 0.6 gave an optimized sensitivity of 99% and a specificity of 45% for CVS achievement. In comparison to this, for the combined acute LR model, ROC curve analysis gave an AUC of 0.829 (0.787-0.872) ( $P < 0.001$ ). A cut-off of 75% probability resulted in a sensitivity of 95% and a specificity of 38% for CVS achievement.

**Conclusion:** The present study illustrates how BN modeling can map to surgeon decision making to facilitate



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



reasoning in complex environments. Further work is needed to facilitate data capture and implementation. Despite this, they represent a promising avenue for intuitive decision support tools.

**Keywords:** Operative process, laparoscopic cholecystectomy, artificial intelligence, Bayesian networks

## INTRODUCTION

Surgeon's intraoperative decisions significantly impact patient outcomes. In the model described by Cristancho *et al.*, interoperative decisions are guided by probabilistic reasoning, which is informed by the evolving intraoperative features<sup>[1]</sup>. Traditional decision support models using logistic regression analysis to generate odds ratios fail to capture these probabilistic accumulations and inter-related probabilities<sup>[2]</sup>. In contrast to this, Bayesian network (BN) maps generate conditional probabilities among a set of connected variables<sup>[3]</sup>. This mirrors the intraoperative decision-making process described by Cristancho *et al.*<sup>[1]</sup>. BN maps, therefore, have the potential as intraoperative decision support tools<sup>[4,5]</sup>.

BNs are useful in probabilistic causal and risk modeling due to their transparency, as well as their ability to cope with missing observations and noisy data. The node probability tables required to inform BNs can be supplied via expert judgment, learned from training data, or a combination of expert judgment and empirical data. However, BNs are computationally intensive, and only recently have advances in computing power and the underlying algorithms made BNs more accessible to researchers in many domains, including medical research<sup>[6]</sup>. In a scoping review of BNs in healthcare, McLachlan *et al.* found BNs referencing twenty-one categories of medical condition. The four most common, in descending order, were (i) cardiac conditions, (ii) cancer, (iii) psychological and psychiatric disorders, and (iv) lung and breathing disorders. Of note, there has been little focus on intraoperative decision making<sup>[7]</sup>. A crucial decision in laparoscopic cholecystectomy arises if hostile intraoperative findings necessitate a bailout strategy where a critical view of safety (CVS) cannot be safely achieved<sup>[8]</sup>. Current clinical guidelines lack clear indications for this decision<sup>[9-11]</sup>. BN mapping providing updated real-time probabilities using salient intraoperative features may inform this decision, providing clearer indications for when to proceed with a CVS or to perform a bailout approach. This paper aims to compare the utility of a traditional logistic regression (LR) model for CVS achievement to BN maps using intraoperative features. It hypothesizes that BN mapping better integrates with surgeon heuristics.

## METHODS

### Data capture

Data were prospectively captured from patients presenting to Christchurch Hospital for laparoscopic cholecystectomy between 2016 and 2022 using "Solutions Committed to Operative Procedure Excellence" (scOPe) (Scope Solutions Ltd), a perioperative workflow solution<sup>[12]</sup>. Data capture was routinely completed by the attending clinical team throughout the patient's operative journey. Specifically, intraoperative data were recorded by the primary surgeon at the time of the operation using synoptic operation reports. Data captured included demographics, time of booking, date of admission, and postoperative complications.

### Ethics

Ethical approval was deemed out of scope by the National Health and Disability Ethics Committee. Locality approval was sought and granted through the Canterbury District Health Board low-risk pathway (RO22146).

## Outcomes

The primary outcome was critical view of safety achievement. Achievement of CVS was determined by the operating surgeon at the time of the operation. Achievement of CVS was defined using the SAGES consensus guidelines for safe cholecystectomy<sup>[8]</sup>:

1. The hepato-cystic triangle was cleared of fat and fibrous tissue.
2. The lower one-third of the gallbladder was separated from the liver to expose the cystic plate.
3. Two and only two structures are seen entering the gallbladder.

## Univariate analysis and factor selection

A literature review in conjunction with expert review of the local data set was conducted to identify those factors impacting the likelihood of achieving the CVS<sup>[2,13-15]</sup>. In keeping with BN development best practice, the risk idiom was identified and used for network generation, as the research objective was to understand the relative influence of several risk factors, both individually and in temporal combinations. Factors shown to significantly impact the risk of not achieving a CVS were then included. Liver function tests were excluded due to variation in performance < 24 h preoperatively to ensure dataset completeness. Factors identified included:

- Admission: acute or elective
- Patient sex: male or female
- Patient age: quartiles
- White cell count: quartiles
- C-reactive protein (CRP): quartiles
- Liver appearances: normal or cirrhotic, fibrotic, fatty
- Rouviere's sulcus: visible or not visible
- Ease of gallbladder retraction: straightforward retraction or requiring dissection of adhesions
- Operative grade: Straight forward as defined by North Shore Difficulty Grade 1 and 2 or challenging as defined by North Shore Difficulty Grade 3 and 4

To assess for the significance of these factors in predicting CVS achievement, statistical analysis was conducted using IBM SPSS for Windows version 22 (IBM Corp., Armonk, N.Y., USA). Factors were dichotomized or divided into quartiles for the purposes of the BN model. Contingency tables stratifying the impact of these factors on CVS achievement were generated and Chi-square testing performed. Statistical significance was considered at  $P < 0.05$ .

For the development and assessment of the BN model, data were split, with 70% used for model training and 30% used for model testing. For the logistic regression, data were again split, allocating 70% for model training and 30% for model testing. Patients with missing data points were excluded.

## BN model development

BNs are a method to formally represent conditional probability and causal relationships among a set of variables. Using Bayes' theorem, BNs facilitate both causal and diagnostic (inverse) reasoning based on the probability distributions of the phenomena of interest<sup>[16]</sup>. A BN consists of a directed acyclic graph, where no node can be the parent of itself, with associated node probability tables (NPT) for each node within the graph. The nodes represent each variable and edges connect dependent variables. An edge from node A to node B represents a causal assumption that node A has a causal or influential function on node B. Node A is often described as a parent of node B. The NPT is the probability distribution of the node, given the parent nodes. For nodes without parents, the NPT is just the probability distribution of the node<sup>[6,16]</sup>.

The recommended best practice for constructing BNs is to build up small network fragments using appropriate reasoning idioms<sup>[17]</sup>. General reasoning idioms include notions such as cause-consequence, definition, and measurement<sup>[16]</sup>. Specific medical idioms have also been identified as helpful in building BN through medical reasoning<sup>[17]</sup>. Examples of medical reasoning idioms include manifestation, pathogenesis, risk, treatment, complications, and comorbidities. For the Critical View of Safety BN, the risk idiom was identified as the most appropriate, as the research objective was to understand the relative influence of several risk factors, both individually and in temporal combinations.

To construct the BN, expert judgment (SC and IT-E) and relevant literature were used to identify the risk factors<sup>[2,18]</sup>. The identified risk factors serve as the nodes of the BN. These nodes can be configured to suit category, Boolean, ordinal, continuous, or integer interval data. For this research, the use of Boolean nodes supported the underlying decision-making focus of the research. The information of interest is not a specific value, but whether that value has reached a threshold for a required action. Expert clinical input also recommended separating the data into acute and elective cohorts, given the differing presentation pathways. This decision was also influenced by the differences in the available data for these two cohorts.

The options for the structural relationship of the nodes began with the simplest hypothesis, with all the predictor variables treated as independently influencing the outcome predictor variable - the critical view of safety being achieved (CVSA). A second hypothesis proposed by the surgeons was the role of the gall bladder visible inflammation grade as a mediating variable between the patient history risk factors and the outcome. Identifying such a mediating variable would provide an early warning indicator that CVSA was at risk. White cell count and CRP were hypothesized as directly influencing the inflammation grade, without additional causal paths to the outcome variable. Age and Gender were considered to have this mediating path while also retaining a direct path to the CVSA outcome. To assess these two different causality hypotheses for two patient cohorts, four BN models were built using the AgenaRisk software application (<https://www.agena.ai/>). The four BN models tested are summarized in [Table 1](#), and the structural relationships between variables for each model are shown in [Figures 1-4](#).

Having constructed the BN graphs, the NPTs for each model were trained by importing the relevant training data set. The BN uses the empirical data to define the NPT parameters for each node, by determining a joint probability distribution that best explains the relationships between the predictor and outcome variables in the data. The data contained some null values. For data with missing values, the AgenaRisk software runs an expectation maximization algorithm to achieve maximum likelihood estimation<sup>[16]</sup>. Simply put, maximum likelihood estimates the parameters of an assumed probability distribution to give the highest likelihood of the observed data<sup>[16]</sup>.

The authors followed the recommendations of Kyrimi *et al.* (2021) for aiding the reproducibility and usefulness of the BN models<sup>[3]</sup>. Specifically: (i) the potential clinical benefits have been explained; (ii) the BN structure and variables have been provided, with the data sets and AgenaRisk models available on reasonable request; (iii) an assessment of accuracy has been provided, and (iv) consideration of the model use in practice.

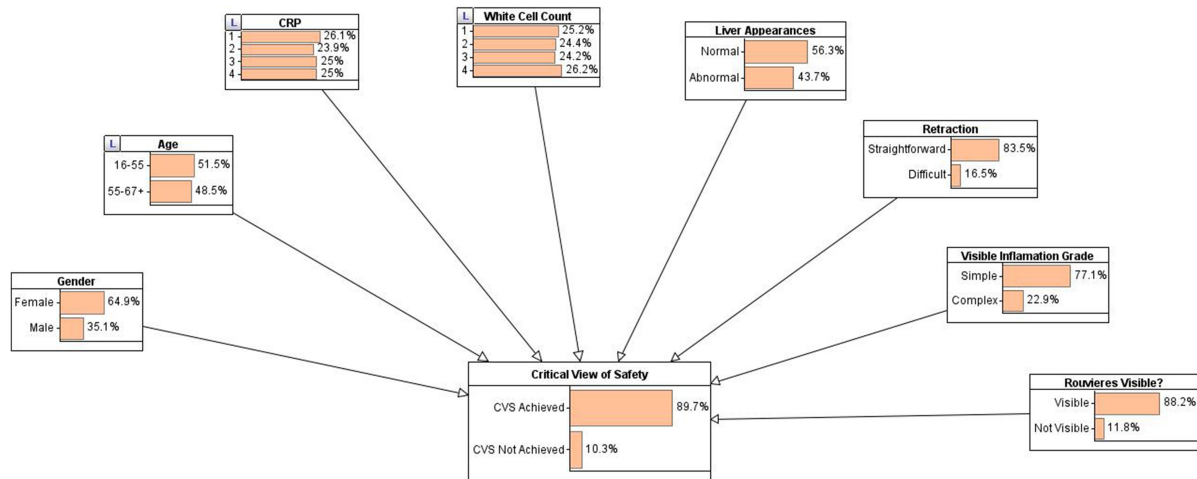
### **Logistic regression model**

Significant features ( $P < 0.05$ ) from the univariate analysis were included in the logistic regression model developed using IBM SPSS Data Modeller v18.4 (IBM Corp., Armonk, N.Y., USA). Models were considered for the combined, acute, and elective data sets.

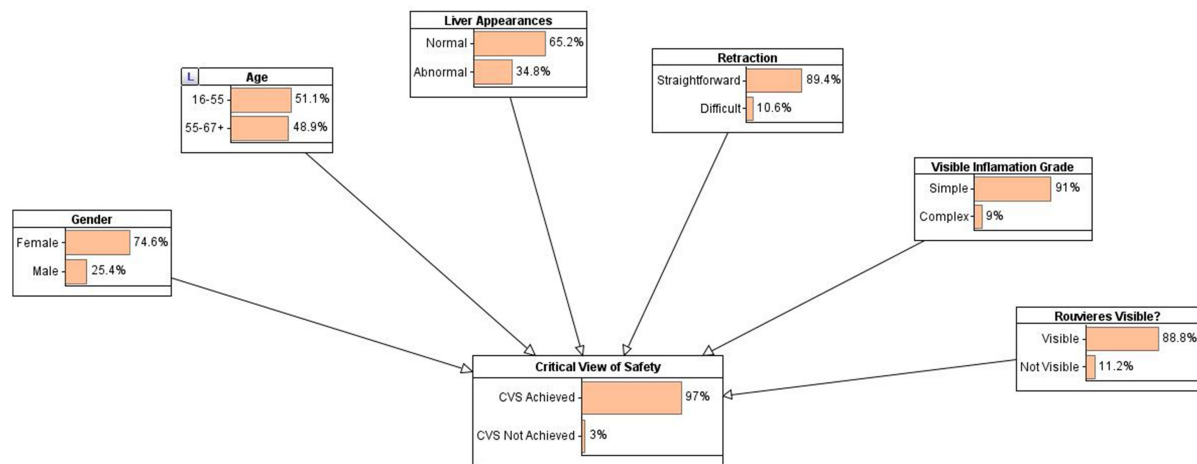
**Table 1. Differences between the four critical view of safety BN models developed**

BN model ID	Causality assumption	Population
Model 1	No parent-child dependency amongst predictor variables	Acute presentations
Model 2	No parent-child dependency amongst predictor variables	Elective presentations
Model 3	Causal chain for inflammation grade	Acute presentations
Model 4	Causal chain for inflammation grade	Elective presentations

BN: Bayesian network.



**Figure 1.** Critical view of safety: model 1 (acute presentations only).



**Figure 2.** Critical view of safety: model 2 (elective presentations only).

**Validation**

Internal validation for both the BN and LR models was performed on the partitioned test set. For the LR model, AUC testing performance was assessed by generating receiver operating characteristic (ROC) curves to determine the area under the curve (AUC) and optimal sensitivities and specificities<sup>[19]</sup>. This methodology is in keeping with literature comparing BN and LR models<sup>[4,5]</sup>.

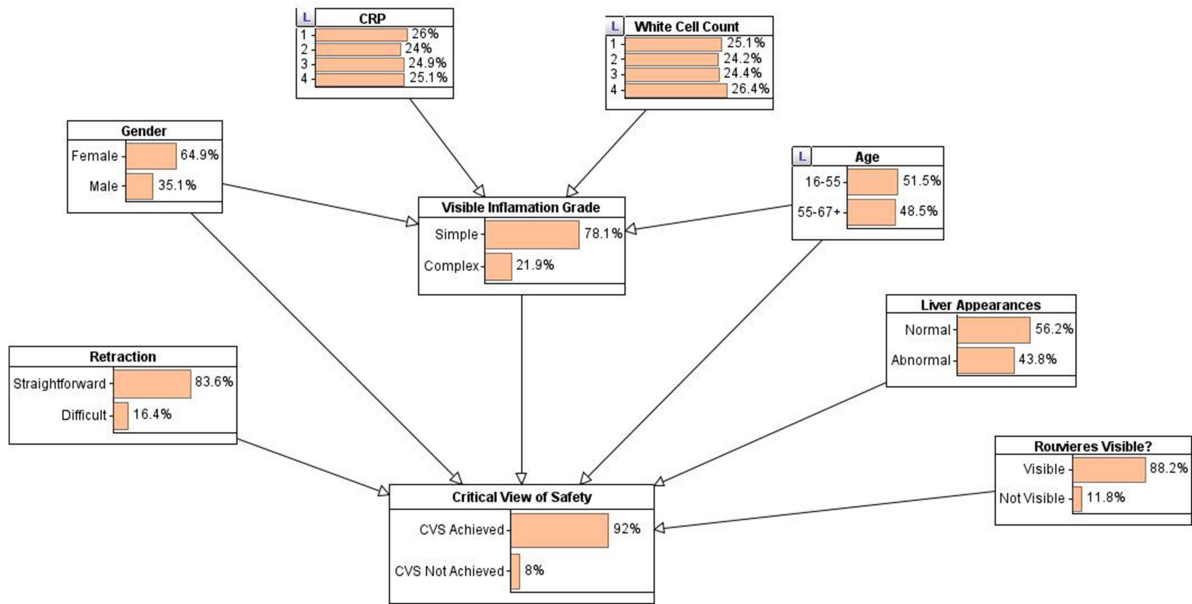


Figure 3. Critical view of safety: model 3 (acute presentations only).

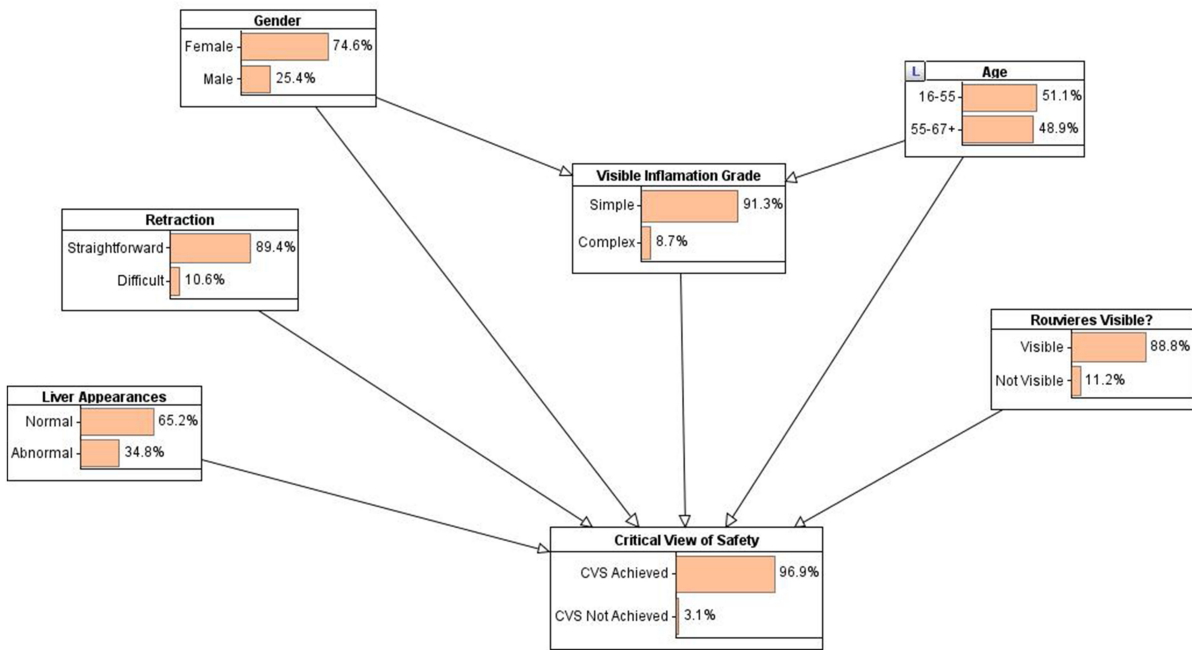


Figure 4. Critical view of safety model 4 (elective presentations only).

## RESULTS

In total, 4,663 patients were identified. Of these patients, 2,837 (61%) presented acutely and 3,122 (67%) were female. CVS was achieved in 4,131 (92%) patients. Data set characteristics following division of the data into training and test sets are provided in [Table 2](#) and the univariate analysis in [Table 3](#).

**Table 2. Demographic factors**

	Factors	Training set %	Test set %
Number of patients		3,072 (70)	1,551 (30)
Median quartiles	0-41	769 (25)	633 (41)
	42-55	791 (26)	324 (21)
	56-67	671 (22)	227 (15)
	67+	810 (26)	345 (22)
	Missing	31 (1)	
Gender	Male	959 (31)	573 (37)
	Female	2,113 (69)	978 (63)
Admission	Acute	1,835 (60)	978 (63)
	Elective	1,227 (40)	573 (37)
Grade	Easy	2,306 (82)	1,162 (75)
	Hard	493 (16)	242 (16)
	Missing	272 (9)	147 (10)
Liver appearances	Normal	1,691 (55)	949 (61)
	Fatty, cirrhotic fibrotic	1,136 (37)	472 (30)
	Missing	245 (8)	130 (8)
Retraction	Easy	2,559 (83)	1,228 (79)
	Difficult	420 (14)	217 (14)
	Missing	93 (3)	106 (7)
CRP	Quartile 1	336 (11)	175 (11)
	Quartile 2	309 (10)	158 (10)
	Quartile 3	319 (10)	150 (10)
	Quartile 4	319 (10)	181 (12)
	Missing	1,281 (42)	887 (57)
WCC	Quartile 1	389 (13)	211 (14)
	Quartile 2	392 (13)	200 (13)
	Quartile 3	377 (12)	179 (12)
	Quartile 4	382 (12)	205 (13)
	Missing	1,532 (50)	756 (49)
CVS	Yes	2,774 (90)	1,233 (85)
	No	195 (6)	115 (7)
	Missing	103 (3)	114 (7)

CRP: C-reactive protein; WCC: white cell count; CVS: critical view of safety.

## Logistic regression model training

### *Combined data set*

Dividing the data into a training set comprising 70% and a remaining 30% yielded a final training set of 2,636 patients, excluding those with missing variables. Logistic regression modeling produced a McFadden's pseudo R-squared of 0.206 ( $P < 0.001$ ) [Table 4]. The single most significant factor in predicting whether CVS was achieved was the operative grade with an Exp(B) of 4.2764 ( $P < 0.001$ ). Other significant factors included the youngest age quartile, easy liver retraction, and elective procedures [Table 4].

### *Acute data set*

When considering the acute data set separately, 1,102 patients were included in the training set once patients with missing data points were excluded. Logistic regression modeling produced a McFadden's pseudo R-squared of 0.185 ( $P < 0.001$ ). CRP was the most significant factor in predicting the likelihood of CVS achievement, with a CRP in the first quartile having an Exp(B) of 10.87 ( $P < 0.001$ ). Other significant

**Table 3. Univariate analysis**

		CVS achieved	CVS not achieved	P-value
Median quartiles	0-41	728 (98)	18 (2)	$P < 0.001$
	42-55	728 (95)	38 (5)	
	56-67	579 (90)	66 (10)	
	67+	709 (91)	72 (9)	
Gender	Male	828 (89)	98 (11)	$P < 0.001$
	Female	1,946 (95)	97 (5)	
Admission	Acute	1,631 (92)	148 (8)	$P < 0.001$
	Elective	1,143 (96)	47 (4)	
Grade	Easy	2,200 (96)	83(4)	$P < 0.001$
	Hard	380 (78)	108 (22)	
Liver appearances	Normal	1,593 (95)	82 (5)	$P < 0.001$
	Fatty, cirrhotic fibrotic	1,019 (91)	102 (9)	
Retraction	Easy	2,419 (96)	114 (4)	$P < 0.001$
	Difficult	337 (81)	80 (19)	
CRP	Quartile 1	331 (100)	1 (0)	$P < 0.001$
	Quartile 2	284 (95)	16 (5)	
	Quartile 3	277 (90)	32 (10)	
	Quartile 4	245 (80)	60 (20)	
WCC	Quartile 1	261 (95)	18 (5)	$P < 0.001$
	Quartile 2	366 (96)	16 (4)	
	Quartile 3	335 (91)	32 (9)	
	Quartile 4	318 (86)	51 (14)	

CVS: Critical view of safety; CRP: C-reactive protein; WCC: white cell count.

factors included operative grade, the visibility of Rouviere's sulcus, and ease of gallbladder retraction [Table 4].

#### *Elective data set*

When considering the elective data set separately, 1,005 patients were included in the training set once patients with missing data points were excluded. Logistic regression modeling produced a McFadden's pseudo R-squared of 0.314 ( $P < 0.001$ ). Age was the most significant factor in predicting the likelihood of CVS achievement, with an age in the second quartile having an Exp(B) of 12.09 ( $P < 0.001$ ). Other significant factors included operative grade, the visibility of Rouviere's sulcus, and ease of gallbladder retraction [Table 4].

#### **Logistic regression model testing**

For the combined regression model, the final 30% test split resulted in 1,207 patients in the test data set. ROC curve analysis gave an AUC of 0.829 (0.787-0.872) ( $P < 0.001$ ). A cut-off of 75% probability resulted in a sensitivity of 95% and a specificity of 38%. The significant class imbalance seen in the rate of CVS achievement precluded ROC curve analysis for the acute and elective subsets.

#### **Bayesian models**

To assess for the impact of different variables on network performance, four network models were developed and assessed [Figures 1-4]. The relative importance of each feature can be seen in Figure 5.



**Table 4. Logistic regression models**

	<b>Factor</b>	<b>Combined model Exp(B) (CI)</b>	<b>Significance</b>	<b>Acute model Exp(B) (CI)</b>	<b>Significance</b>	<b>Elective model Exp(B) (CI)</b>	<b>Significance</b>
Grade	1 and 2	4.274 (3.044-6.001)	$P < 0.001$	2.224 (1.344-3.680)	$P = 0.002$	8.421 (3.995-17.754)	$P < 0.001$
	3 and 4	Reference		Reference		Reference	
Age quartile	1	2.644 (1.577-4.433)	$P < 0.001$	1.830 (0.852-3.927)	$P = 0.121$	2.484 (0.833-7.4)	$P = 0.102$
	2	1.401 (.907-2.164)	$P = 0.128$	0.911 (0.502-1.654)	$P = 0.759$	12.095 (1.55-94.071)	$P = 0.017$
	3	0.913 (0.605-1.378)	$P = 0.666$	0.770 (0.436-1.359)	$P = 0.367$	0.739 (0.327-1.669)	$P = 0.466$
	4	Reference		Reference		Reference	
Liver appearances	Normal liver appearances	1.320 (0.956-1.823)	$P = 0.91$	1.162 (0.740-1.825)	$P = 0.515$	1.237 (0.601-2.544)	$P = 0.564$
	Cirrhotic, fibrotic or fatty liver	Reference		Reference		Reference	
Liver retraction	Easy	3.012 (2.139-4.234)	$P < 0.001$	2.091 (1.289-3.393)	$P = 0.003$	4.069 (1.943-8.520)	$P < 0.001$
	Difficult	Reference		Reference		Reference	
	Rouviere's visible	Reference		Reference		Reference	
	Rouviere's not visible	0.311 (0.216-0.448)	$P < 0.001$	0.414 (0.240-0.713)	$P = 0.002$	0.350 (0.154-0.797)	$P = 0.012$
Gender	Male	Reference		Reference		Reference	
	Female	1.334 (0.958-1.859)	$P = 0.088$	1.757 (1.113-2.773)	$P = 0.016$	1.011 (0.911-4.008)	$P = 0.086$
Admission	Acute	0.609 (0.417-0.890)	$P = 0.010$	NA	NA	NA	NA
	Elective	Reference					
CRP	Quartile 1	NA	NA	10.767 (3.105-37.339)	$P < 0.001$	NA	NA
	Quartile 2			2.518 (1.275-4.972)	$P = 0.008$		
	Quartile 3			1.431 (0.818-2.505)	$P = 0.209$		
	Quartile 4			Reference			
WCC	Quartile 1	NA	NA	0.866 (0.416-1.802)	$P = 0.700$	NA	NA
	Quartile 2			0.949 (0.462-1.948)	$P = 0.886$		
	Quartile 3			0.801 (0.447-1.434)	$P = 0.455$		
	Quartile 4			Reference			

CRP: C-reactive protein; WCC: white cell count; NA: not applicable.

### Model 1

Model 1 is seen in [Figure 1](#), with an AUC of 0.780 (0.726-0.833) ( $P < 0.001$ ) [[Figure 2](#)]. An optimal cut-off of 0.69 gave an optimized sensitivity of 96% and a specificity of 30%. Comparing model 1 to model 3 gave a Bayes factor of 1.026 [[Table 5](#)].

**Table 5. Model comparison\***

<b>Marginal probability of CVS achieved - acute</b>				
Test data set [Evidence]	Model 1 [Hypothesis 1]	Model 3 [Hypothesis 2]	$\frac{P(E H1)}{P(E H2)}$	Bayes factor
0.903	0.897	0.920	$\frac{(0.903 \times 0.897)}{(0.903 \times 0.920)}$	0.975 (H2 performs 1.026× better than H1)
<b>Marginal probability of CVS achieved - elective</b>				
Test data set [Evidence]	Model 2 [Hypothesis 1]	Model 4 [Hypothesis 2]	$\frac{P(E H1)}{P(E H2)}$	Bayes factor
0.950	0.970	0.969	$\frac{(0.950 \times 0.970)}{(0.950 \times 0.969)}$	1.001 (no difference between hypotheses)
<b>Risk scenario R12: probability CVS achieved (acute, male, 55-67+, straightforward retraction, complex grade)</b>				
Test data set [Evidence]	Model 1 [Hypothesis 1]	Model 3 [Hypothesis 2]	$\frac{P(E H1)}{P(E H2)}$	Bayes factor
0.774	0.712	0.810	$\frac{(0.774 \times 0.712)}{(0.774 \times 0.810)}$	0.879 (H2 performs 1.138× better than H1)

\*Comparing the predictive models using Bayes factors. The same assumptions apply to the prior probability distribution for each model (H1 and H2), so the prior probabilities are set to 1. The probability for the evidence (E) from the test data set is also the same for each model. In these circumstances, the Bayes factor is simply the likelihood ratio. The results in [Table 1](#) show Bayes factors ranging from 1.0 to 1.13 which is regarded as very weak evidence. There is no basis to favour one hypothesis (model) over the other. Risk scenario R12 is an example with multiple risk factors. Further Risk scenarios are included in the [Supplementary Material](#). CVS: Critical view of safety.

### Model 2

Model 2 is seen in [Figure 2](#), with an AUC of 0.879(0.798-0.960) ( $P < 0.001$ ). Selecting a cut-off of 0.6 gave an optimized sensitivity of 99% and a specificity of 45%. Comparing model 2 to model 4 gave a Bayes factor of 1.001 [[Table 5](#)].

### Model 3

Model 3 is seen in [Figure 3](#), with an AUC of 0.785 (0.734-0.836) ( $P < 0.001$ ). Selecting a cut-off of 0.77 gave an optimized sensitivity of 96% and a specificity of 30%.

### Model 4

Model 4 is seen in [Figure 4](#), with an AUC of 0.879 (0.798-0.960) ( $P < 0.001$ ). Selecting a cut-off of 0.6 gave an optimized sensitivity of 99% and a specificity of 45%.

## DISCUSSION

BN performance was comparable to the logistic regression models, with AUCs ranging from 0.780 to 0.879 ( $P < 0.001$ ). This is in keeping with prior literature showing comparable performance between regression and Bayesian models<sup>[3]</sup>. Performance across all BN models was comparable, as evidenced by the Bayes factors [[Table 5](#)]<sup>[20]</sup>. Across both acute and elective models, operative grade was the most consistent factor predicting CVS achievement [[Figure 5](#) and [Table 4](#)]. This is in keeping with earlier work reaffirming the importance of considering operative technical difficulty when assessing operative process<sup>[13,21,22]</sup>. The performance of elective models was most likely superior due to reduced operative process variance. To the authors' knowledge, this is the first study illustrating the utility of BN using intraoperative features to predict an intraoperative outcome. The comparable performance across both LR and BNs depicts the potential utility of BN mapping to support intraoperative decision making. For clinical utility, higher sensitivity and specificity are needed. Achieving this will require both a better understanding of operative processes and variable capture. If improved data performance is achieved, prospective model testing in an offline manner is required before multicentre validation.



Figure 5. Tornado graph for models 1-4.

The persistent rate of common bile duct injury in the era of laparoscopic cholecystectomy has led to concerted efforts to increase the rate of CVS achievement<sup>[8]</sup>. A significant amount of this work has focused on surgeon education and improved documentation as a means of increasing achievement rates<sup>[11,23]</sup>. This has been achieved at the authors' local institution through the implementation of a standard process, resulting in a high rate of CVS achievement and excellent outcomes<sup>[24]</sup>. Increasingly, it is recognized that intraoperative factors significantly impact the ability of surgeons to achieve CVS<sup>[2,18]</sup>. This is reflected in the present study, which shows the significant impact of operative grade, gallbladder retraction and ability to identify Rouviere's sulcus on the likelihood of CVS achievement [Table 4 and Figure 5]. When looking at the tornado graphs in figure 5 for patients undergoing acute laparoscopic cholecystectomy, the ability to retract the liver had the most impact on the likelihood of CVS achievement. For elective patients, the predominant driver was inflammation grade. These findings suggest a systematic difference in the reasons for not achieving a CVS between elective and acute operations. Consequently, different operative processes might be needed to optimize the rates of achievement. These findings evidence how specific operative features impact decision making and outcomes. Generalization of this approach across different operations may lead to broader insights into optimizing operative techniques.

The clinical utility of BN models in operative decision-making lies in their ability to mirror the operative thought process of surgeons. This is in contrast to traditional LR models that provide fixed probabilities that do not integrate intuitively with clinical workflow. Additionally, given that associations between predictor variables are accounted for by BN models, they are not subject to the same issues with multicollinearity encountered in regression models<sup>[25]</sup>. Furthermore, their ability to incorporate patients with missing data further facilitates their application in clinically uncertain environments. Cristancho *et al.*'s model breaks intraoperative decision making into the steps of assessing the information, reconciling this with known information, and implementing a plan of action<sup>[1]</sup>. In this model, the surgeon proceeds with the plan of action that has the highest probability of success. This is currently implicit, but by mirroring this process, BN can make these probabilities explicit, with probabilities at each step of the operation updated as more information becomes available, as illustrated in the [Supplementary Material](#). This stepwise process also illustrates the significance of each factor [Supplementary Material]. For senior surgeons, these probabilities would supplement the current decision-making process, while for surgical trainees, BNs could act as prompts to consider all available evidence and facilitate reasoning in an uncertain environment. This is likely to be highly beneficial to trainees who have not yet had the procedural experience to develop the rich process maps and high-level heuristics of surgical consultants<sup>[26,27]</sup>. Alternatively, post-procedure reflection could be informed using counterfactual probabilities generated from the BN. These could provide prompts for discussing the possible impact of different steps taken during the operation.

To achieve this, however, high-accuracy models using well-validated, broad data sets will be required. Currently, models have a high sensitivity and low specificity for CVS achievement. This is reflective of the significant class imbalance in the present data set, with a 92% chance of achieving a CVS. This contrasts with the broader literature which shows a variable rate of CVS achievement, especially when this is externally reviewed<sup>[10,11]</sup>. In the present data set, a subtotal cholecystectomy was frequently performed when a CVS could not be achieved. This, coupled with the low McFadden's pseudo R-squared values seen in the regression models, would suggest that a considerable number of factors influencing decision making have not been captured in the current models. This is more marked in the acute setting with an R-squared of 0.185 compared to 0.314 for the elective patients. It is likely that this is due to elective patients representing a more homogenous range of pathophysiological status and a smaller amount of variation in process compared to acute patients<sup>[22]</sup>. This is driven by operative technical difficulty with previous work identifying a higher proportion of grade 3 and 4 operative findings in acute patients<sup>[13,24]</sup>. Taken together, these findings

suggest that the significant class imbalance seen could be addressed through subgroup analysis of subtotal cholecystectomy. Focused video analysis of these patients may identify significant factors driving intraoperative decision making. Given the lack of clear indications for subtotal cholecystectomy included in current guidelines, this would contribute significantly to standardizing practice and improving patient care.

Despite a significant clinical interest in the study of BNs, there is no literature concerning their implementation into practice or gold standard for model development<sup>[3]</sup>. Early work has demonstrated their utility in personalized risk prediction for malignancy in patients with Barrett's oesophagus. This reflects both challenges in development, data set requirements, and the significant barriers around testing and implementation. In the present study, BN models were developed by reviewing existing literature and expert opinion of features predicting CVS achievement. These were then arranged hierarchically using an iterative improvement process. Initial model development concerned the impact of setting an observation on a causal path. By creating a causal pathway as soon as the inflammation grade value was defined in an observation, the parent nodes such as age and gender were blocked. To correct this, a revised model was developed, including a direct and indirect path between the parent nodes, and the outcome variable (CVSA) was generated. A more effective means of developing BN maps may be using cognitive task analysis. Work by Hashimoto *et al.* has illustrated that surgeon knowledge can be made explicit through process mapping<sup>[26]</sup>. Such an approach may increase model performance by enabling the capture of a wider variety of data points.

A weakness of the present study was the lack of granularity seen across the network maps. Initially ranked nodes were felt to offer more granular insight than Boolean nodes. However, even with a large data set, the number of possible combinations resulted in insufficient procedures among the training and test data. Converting all nodes to Boolean helped but did not completely eliminate this problem. For eight Boolean predictor variables, there are 256 possible combinations that subdivide the data. This precluded the inclusion of specific surgeons as factors in predicting the likelihood of CVS achievement. In the present study, over 40 surgeons undertook one of the documented operations. The inclusion of each surgeon in the BN would have resulted in a vast number of probability pathways. This highlights a problem in BNs in general and is representative of a broader issue in applying machine learning techniques to healthcare. Given the huge volume of data generated by healthcare, the bottleneck lies not in its generation but rather in its capture and organization. This is further reflected in the scant examples of ML-based decision support tools in clinical practice in contrast to the rapidly increasing literature touting their potential utility. Integration of these tools into practice will require a re-imagining of the clinical workflow to enable data to be captured as part of workflow. Doing so will allow for the development, application, and validation of these novel tools.

In keeping with AI model development, the current paper represents a proof of concept. Further clinical work, along with more refined data sets, is imperative to enable better prediction. Upon achieving this offline, prospective validation is needed to ensure the safety of application. Following confirmation, real-time risk prediction becomes feasible. However, model development should progress using multicentre data to ensure generalizability. Subsequent local testing is then necessary to validate the model's applicability within its specific environment.

The present study illustrates how BN modeling can map surgeon decision making to facilitate reasoning in complex environments. Additional efforts are required to facilitate data capture and implementation. Nevertheless, they present a promising avenue for intuitive decision support tools.

## DECLARATIONS

### Authors' contributions

Conceived, performed data gathering, drafted and reviewed the present study: Tranter-Entwistle I  
Conceived, drafted and reviewed the present study, participated in video annotation for an unrelated study, contributed to the development of a freely available educational application around laparoscopic cholecystectomy: Connor S

Conceived, drafted and reviewed the present study: Wilson B

Drafted and reviewed the present study: Eglinton T

### Availability of data and materials

Data and materials are available upon reasonable request.

### Financial support and sponsorship

None.

### Conflicts of interest

Tranter-Entwistle I has received funding from Medtronic to pursue a PhD. Connor S has provided pro bono consultancy for Digital Surgery LTD. Wilson B declares no conflicts of interest. Eglinton T has provided consultancy for Digital Surgery Ltd, which is separate from the present study.

### Ethical approval and consent to participate

Ethical approval was deemed out of scope by the National Health and Disability Ethics Committee. Locality approval was sought and granted through the Canterbury District Health Board low risk pathway (RO#22146). Patients were consented for their procedures and data prospectively captured.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2024.

## REFERENCES

1. Cristancho SM, Apramian T, Vanstone M, et al. Thinking like an expert: surgical decision making as a cyclical process of being aware. *Am J Surg* 2016;211:64-9. [DOI](#) [PubMed](#) [PMC](#)
2. Onoe S, Maeda A, Takayama Y, Fukami Y, Kaneoka Y. A preoperative predictive scoring system to predict the ability to achieve the critical view of safety during laparoscopic cholecystectomy for acute cholecystitis. *HPB* 2017;19:406-10. [DOI](#) [PubMed](#)
3. Kyrimi E, McLachlan S, Dube K, Neves MR, Fahmi A, Fenton N. A comprehensive scoping review of Bayesian networks in healthcare: past, present and future. *Artif Intell Med* 2021;117:102108. [DOI](#) [PubMed](#)
4. Cai ZQ, Guo P, Si SB, Geng ZM, Chen C, Cong LL. Analysis of prognostic factors for survival after surgery for gallbladder cancer based on a Bayesian network. *Sci Rep* 2017;7:293. [DOI](#) [PubMed](#) [PMC](#)
5. Bradley A, Sami S, Hwei NG, et al. A predictive Bayesian network that risk stratifies patients undergoing Barrett's surveillance for personalized risk of developing malignancy. *PLoS One* 2020;15:e0240620. [DOI](#) [PubMed](#) [PMC](#)
6. Ivens GP. Chemical process dynamics: By Rezső Mohilla and Béla Ferencz; published by Elsevier, Amsterdam, 1982; 300 pp.; price, Dfl. 150. *Chem Eng J* 1983;27:120-1. [DOI](#)
7. McLachlan S, Dube K, Hitman GA, Fenton NE, Kyrimi E. Bayesian networks in healthcare: distribution by medical condition. *Artif Intell Med* 2020;107:101912. [DOI](#) [PubMed](#)
8. Brunt LM, Deziel DJ, Telem DA, et al; Prevention of Bile Duct Injury Consensus Work Group. Safe cholecystectomy multi-society practice guideline and state-of-the-art consensus conference on prevention of bile duct injury during cholecystectomy. *Surg Endosc* 2020;34:2827-55. [DOI](#) [PubMed](#)
9. Wauben LSG, van Grevenstein WMU, Goossens RHM, van der Meulen FH, Lange JF. Operative notes do not reflect reality in laparoscopic cholecystectomy. *Br J Surg* 2011;98:1431-6. [DOI](#) [PubMed](#)
10. Sebastian M, Sebastian A, Rudnicki J. Recommendation for photographic documentation of safe laparoscopic cholecystectomy. *World*

- J Surg* 2021;45:81-7. DOI PubMed PMC
11. Nijssen MA, Schreinemakers JM, Meyer Z, van der Schelling GP, Crolla RM, Rijken AM. Complications after laparoscopic cholecystectomy: a video evaluation study of whether the critical view of safety was reached. *World J Surg* 2015;39:1798-803. DOI PubMed
  12. Sakowska MM, Thomas MV, Connor S, Roberts R. Hospital-wide implementation of an electronic-workflow solution aiming to make surgical practice improvement easy. *ANZ J Surg* 2017;87:143-8. DOI PubMed
  13. Nassar AHM, Hodson J, Ng HJ, et al; CholeS Study Group; West Midlands Research Collaborative. Predicting the difficult laparoscopic cholecystectomy: development and validation of a pre-operative risk score using an objective operative difficulty grading system. *Surg Endosc* 2020;34:4549-61. DOI PubMed
  14. Wennmacker SZ, Bhimani N, van Dijk AH, Hugh TJ, de Reuver PR. Predicting operative difficulty of laparoscopic cholecystectomy in patients with acute biliary presentations. *ANZ J Surg* 2019;89:1451-6. DOI PubMed PMC
  15. Madni TD, Nakonezny PA, Imran JB, et al. A comparison of cholecystitis grading scales. *J Trauma Acute Care Surg* 2019;86:471-8. DOI PubMed
  16. Fenton N, Neil M. Risk assessment and decision analysis with bayesian networks. 2nd ed. New York: Chapman and Hall/CRC. 2018. DOI
  17. Kyrimi E, Neves MR, McLachlan S, Neil M, Marsh W, Fenton N. Medical idioms for clinical Bayesian network development. *J Biomed Inform* 2020;108:103495. DOI PubMed
  18. Nassar AHM, Ng HJ, Wysocki AP, Khan KS, Gil IC. Achieving the critical view of safety in the difficult laparoscopic cholecystectomy: a prospective study of predictors of failure. *Surg Endosc* 2021;35:6039-47. DOI PubMed PMC
  19. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861-74. DOI
  20. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;90:773-95. DOI
  21. Tranter-Entwistle I, Eglinton T, Hugh TJ, Connor S. The artery first technique: re-examining the critical view of safety during laparoscopic cholecystectomy. *Surg Endosc* 2023;37:4458-65. DOI PubMed
  22. Tranter-Entwistle I, Eglinton T, Hugh TJ, Connor S. Use of prospective video analysis to understand the impact of technical difficulty on operative process during laparoscopic cholecystectomy. *HPB* 2022;24:2096-103. DOI PubMed
  23. Mascagni P, Rodríguez-Luna MR, Urade T, et al. Intraoperative time-out to promote the implementation of the critical view of safety in laparoscopic cholecystectomy: a video-based assessment of 343 procedures. *J Am Coll Surg* 2021;233:497-505. DOI PubMed
  24. Tranter-Entwistle I, Simcock C, Eglinton T, Connor S. A prospective cohort study of operative outcomes in laparoscopic cholecystectomy using operative difficulty grade-adjusted CUSUM analysis. *Br J Surg* 2023;10:1068-71. DOI PubMed PMC
  25. Cao Y, Raoof M, Szabo E, Ottosson J, Näslund I. Using bayesian networks to predict long-term health-related quality of life and comorbidity after bariatric surgery: a study based on the scandinavian obesity surgery registry. *J Clin Med* 2020;9:1895. DOI PubMed PMC
  26. Hashimoto DA, Axelsson CG, Jones CB, et al. Surgical procedural map scoring for decision-making in laparoscopic cholecystectomy. *Am J Surg* 2019;217:356-61. DOI PubMed PMC
  27. Korovin LN, Farrell TM, Hsu CH, White M, Ghaderi I. Surgeons' expertise during critical event in laparoscopic cholecystectomy: an expert-novice comparison using protocol analysis. *Am J Surg* 2020;219:340-5. DOI PubMed