Making Personalized Predictions of Poor Outcome Post Resection of Pancreatic Ductal Adenocarcinoma (PDAC): a prognostic Bayesian network with pre- and post-operative application

Alison Bradley[1,2], R. Van Der Meer[1], C. McKay[2]
(bradley_alison@live.co.uk)
[1] Department of Management Science, University of Strathclyde Business School, Glasgow, Scotland
[2] West of Scotland Pancreatic Cancer Unit, Glasgow Royal Infirmary, Glasgow, Scotland

Background:

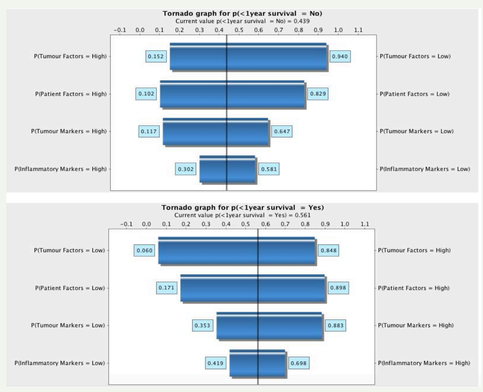
The narrative surrounding the management of potentially resectable pancreatic cancer is complex. Resection rates are low, the risk of operative morbidity and mortality are high, and survival outcomes remain poor. Surgical resection is the only potentially curative treatment but 5year survival rates for resected cases are between only 7% and 25%. The aim of this study was to create a prognostic Bayesian network that pre-operatively makes personalized predictions of post-resection survival time of 12months or less and also performs post-operative prognostic updating.

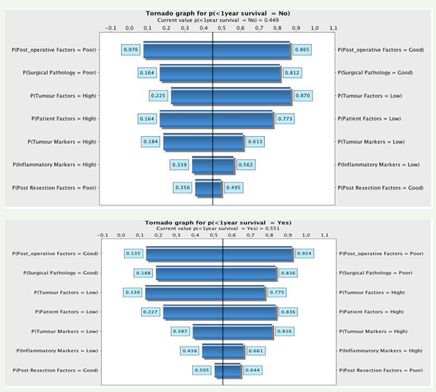
Methods 1: Bayesian Network

Based on probability theory, Bayesian networks (BN) model relationships between variables based on a graphical formalism of a joint or multivariate probability distribution over a set of variables. This is formalized as: BN= (*G,Pr)*. *G* is a graphical structure and *Pr* is the probability distribution. Within the graphical structure of a BN, *G*, variables are modeled as nodes (V(G)) with causal relationships between parent and child nodes represented by directed arcs (A(G)): *G* = V(G), A(G). Within a BN any number of nodes can be included therefore: V(G)= $\{V_1, V_2....V_n\}$ where *n*>1. Directed arcs, A(G), represent the probabilistic influence between parent (V_p)and child (V_c) nodes. The dependence and independence between nodes is defined by the joint probability distribution (*Pr*): $Pr(V_1, V_2....V_n) = \Pi^n_{i\ 1} Pr(V_p/\pi(V_c))$
where $\pi(V_c)$ represents the covariates of parent nodes to V_c. Each node therefore has a conditional probability table representing the probability of each value contained within that node given the condition of all its parent nodes. Through Bayes theorem the prior distribution and observed data are combined to update knowledge in the form of the posterior distribution. Missing data is handled through probabilistic inference with predictions made based on global averages of the patient population. In this way BN allow the modeling of the dynamic relationships between variables contained within the complex healthcare process, with predictions evolving and accuracy improving as more information becomes available.

Methods 3: Sensitivity Analysis

Pearl's inwards analysis and broadcasting analysis were used to perform sensitivity analysis. Hence sensitivity was defined as $S(X=x, T=t)$ and determined by setting values for all source variables, *X*, and assessing the impact on the target node, T, then changing only the target node, T, and assessing the changes on the source set, *X*, respectively with joint sensitivity of T to perturbations in source nodes defined as:$S(X=x,T=t) = p(T=t|e,X=x)p(T=t|e)$ where $p(T=t|e)$ is the current probability value for *T*, given evidence *e* and $p(T=t|e, X=x)$ is the new value of T for the set of source variable, *X*. Hence inwards analysis and broadcasting results were equivocal as: $p(T=t|e,X=x)p(T=t|e) = p(X=x|T=t,e)p(X=x|e)$. The results of BBN sensitivity analysis showed that for the pre-operative BBNs tumour factors had the greatest impact on outcomes, followed by patient factors. When post-operative data was incorporated into the BBN post-operative factors and surgical pathology had greatest impact on output followed by tumour factors and patient factors.

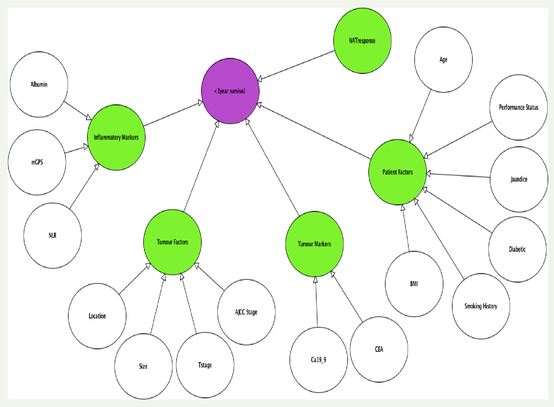




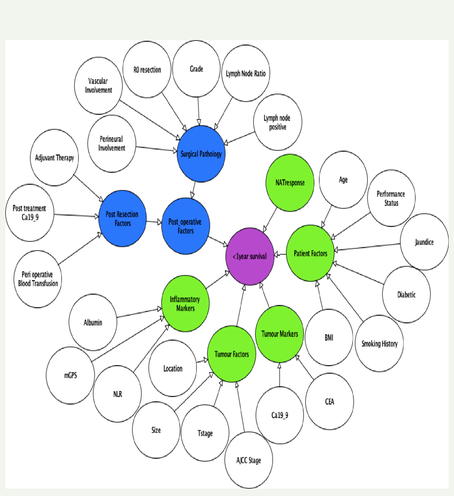
Methods 2:

A Bayesian network was created using AgenaRisk software by synthesizing data from 77 PubMed post-resection survival analysis studies (n=31,214) through a two-stage weighting process. Input variables included: inflammatory markers, tumour factors, tumour markers, patient factors and, if applicable, response to neoadjuvant treatment for pre-operative predictions.

New Risk Object	Scenario 1	Scenario 2	Scenario 3
Albumin	Low	Normal	Normal
mCPS	2	1	0
NLR	>5	>5	<5
Location	Body/Tail	HDP	HDP
Size	>2cm	>2cm	<2cm
T stage	T4	T3	T1
AJCC Stage	4	3	0
Ca19_9	>1000	50-999	<50
CEA	>5	<5	<5
Age	>70	>70	<70
Performance Status	Poor	Moderate	Good
Jaundice	Yes	No	No
Diabetic	Yes	Yes	No
Smoking History	Smoker	Non-smoker	Non-smoker
BMI	<18	Normal	Normal
NATresponse	Progressi...	No Answer	Responsi...



Prognostic updating was performed by inclusion of post-operative input variables including: pathology results and adjuvant therapy.



Lymph node positive	Yes
Lymph Node Ratio	>0.1
Grade	G3/G4
R0 resection	No
Vascular Involvement	Yes
Perineural Involvement	Yes
Adjuvant Therapy	No
Post treatment Ca19 9	>120
Peri operative Blood Transfusion	Yes

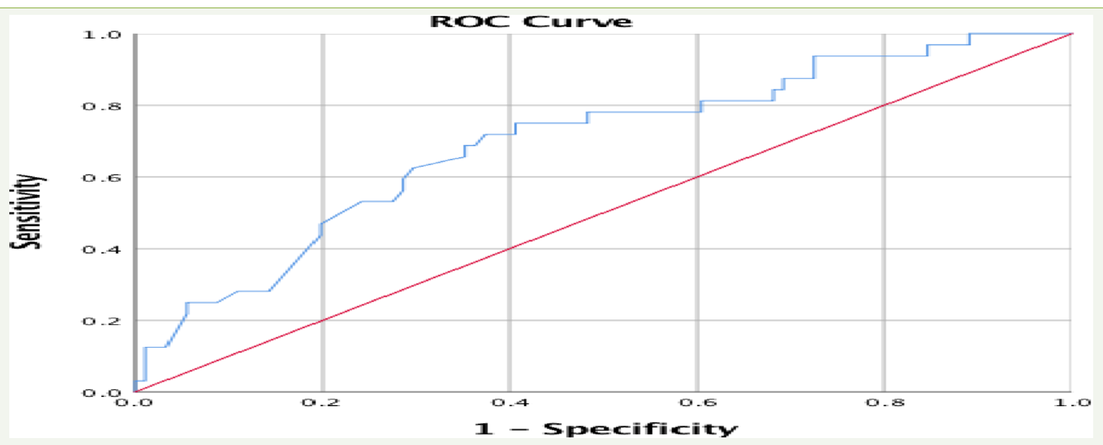
Lymph node positive	Yes
Lymph Node Ratio	<0.1
Grade	G3/G4
R0 resection	Yes
Vascular Involvement	Yes
Perineural Involvement	Yes
Adjuvant Therapy	Yes
Post treatment Ca19 9	<120
Peri operative Blood Transfusion	No

Results:

The performance of the model was validated against a 20year, prospectively maintained patient database from a tertiary referral centre. Individual patient data was entered into the BN and the personalized pre and post-operative predictions of poor prognosis were recorded and assessed against that individual's actual survival time therefore deeming predictions to be true or false.
This gave a pool of 387 and 251 patients against which the predictive performance of the pre and post-operative models were validated respectively.

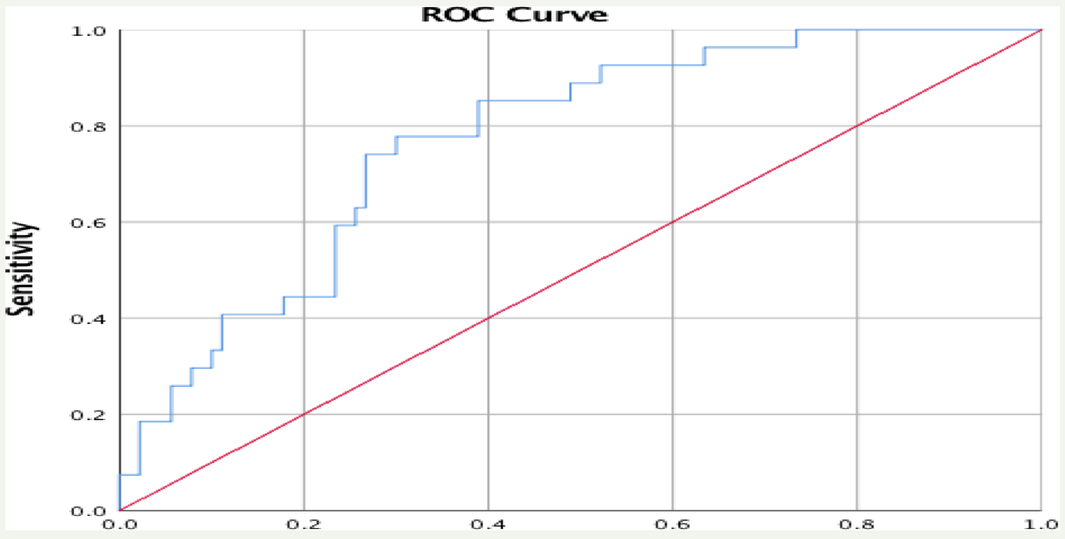
Pre-operative Results

	Sensitivity	Specificity	PPV	NPV	AUC
2 data points missing (n=123)	0.84	0.64	0.45	0.92	0. 70 (*P* value 0.001; 95% CI 0.589-0.801) Std. Error: 0.54
3 data point missing (n=139)	0.82	0.65	0.43	0.92	0. 70 (*P* value 0.001; 95% CI 0.578 -0.786) Std. Error: 0.53
4 data points missing (n=144)	0.83	0.65	0.44	0.92	0.70 (*P* value 0.001; 95% CI 0.591 - 0.791) Std. Error: 0.51
5 data points missing (n=176)	0.64	0.66	0.45	0.81	0. 65 (*P* value 0.009; 95% CI 0.537 -0.711) Std. Error: 0.44
6 data points missing (n=189)	0.66	0.63	0.43	0.82	0.64 (*P* value 0.024; 95% CI 0.518 - 0.690) Std. Error: 0.44
6+ data points missing (n=387)	0.64	0.54	0.46	0.72	0.60 (*P* value 0.559; 95% CI 0.502- 0.617) Std. Error: 0.29



Post-operative Results

	1 Missing Post-operative Data point	1-2 Missing Post-operative Data Point	1-3 Missing Post-operative Data Points	1-4 Missing Post-operative Data Points
2 Missing Pre-operative Data Points	Sensitivity: 0.97; Specificity: 0.62; PPV: 0.44; NPV: 0.98; AUC: 0.80; Standard Error:0.47; *P value*: 0.000; 95% CI: 0.678-0.862 (n=117)	Sensitivity: 0.93; Specificity: 0.63; PPV: 0.44; NPV:0.97; AUC: 0.80; Standard Error:0.51; *P value*: 0.000; 95% CI: 0.651-0.850 (n=120)		
2-3 Missing Pre-operative Data Point	Sensitivity: 0.94; Specificity: 0.63; PPV: 0.45; NPV: 0.97; AUC: 0.80; Standard Error:0.045; *P value*: 0.000; 95% CI: 0.685-0.862 (n=138)	Sensitivity: 0.94; Specificity: 0.63; PPV: 0.45; NPV: 0.97; AUC: 0.80; Standard Error:0.045; *P value*: 0.000; 95% CI: 0.685-0.862 (n=139)		
2-4 Missing Pre-operative Data Points	Sensitivity: 0.94; Specificity: 0.62; PPV: 0.45; NPV: 0.98; AUC: 0.80; Standard Error: 0.042; *P value*: 0.000; 95% CI: 0.708-0.872 (n=135)	Sensitivity: 0.94; Specificity: 0.62; PPV: 0.44; NPV: 0.97; AUC: 0.80; Standard Error: 0.045; *P value*: 0.000; 95% CI: 0.681-0.858 (n=140)		
2-5 Missing Pre-operative Data Points	Sensitivity: 0.97; Specificity: 0.62; PPV: 0.45; NPV: 0.98; AUC: 0.80; Standard Error: 0.041; *P value*: 0.000; 95% CI: 0.708-0.869 (n=137)	Sensitivity: 0.95; Specificity: 0.61; PPV: 0.45; NPV: 0.97; AUC: 0.80; Standard Error: 0.043; *P value*: 0.000; 95% CI: 0.681-0.849 (n=146)		
2-6Missing Pre-operative Data Points	Sensitivity: 0.97; Specificity: 0.61; PPV: 0.45; NPV: 0.98; AUC: 0.80; Standard Error:0.041; *P value*: 0.000; 95% CI: 0.707-0.869 (n=138)	Sensitivity: 0.95; Specificity: 0.59; PPV: 0.43; NPV: 0.97; AUC: 0.80; Standard Error:0.043; *P value*: 0.000; 95% CI: 0.663-0.832 (n=155)	Sensitivity: 0.95; Specificity: 0.59; PPV: 0.44; NPV: 0.97; AUC: 0.80; Standard Error:0.042; *P value*: 0.000; 95% CI: 0.672-0.835 (n=157)	
>6 Missing Pre-operative Data Points	Sensitivity: 0.97; Specificity: 0.61; PPV: 0.45; NPV: 0.98; AUC: 0.80; Standard Error:0.041; *P value*: 0.000; 95% CI: 0.710-0.870 (n=139)	Sensitivity: 0.94; Specificity: 0.55; PPV: 0.41; NPV: 0.96; AUC: 0.70; Standard Error:0.039; *P value*: 0.000; 95% CI: 0.667-0.818 (n=195)	Sensitivity: 0.94; Specificity: 0.54; PPV: 0.41; NPV: 0.96; AUC: 0.70; Standard Error: 0.037; *P value*: 0.000; 95% CI: 0.667-0.814 (n=205)	Sensitivity: 0.96; Specificity: 0.49; PPV: 0.44; NPV: 0.97; AUC: 0.70; Standard Error:0.033; *P value*: 0.000; 95% CI: 0.660-0.788 (n=251)



Discussion/ Conclusion: This Bayesian network is currently unique in the way it utilizes PubMed and patient level data to translate the existing empirical evidence surrounding potentially resectable pancreatic cancer to make personalized prognostic predictions. We believe such a tool is vital in facilitating better shared decision-making in clinical practice and could be further developed to offer a vehicle for delivering personalized precision medicine in the future.