Personalized Prognostic Bayesian Network for Pancreatic Cancer: Delivering personalized pancreatic cancer management throughout the patient journey

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:

Survival outcomes for pancreatic cancer remain poor. Surgical resection is the only potentially curative treatment but approximately 10% present with resectable disease with 5year survival for resected cases between only 7% and 25%. Adjuvant therapy after surgery is required to prolong survival but up to 50% of patients fail to receive adjuvant therapy. Neoadjuvant therapy has emerged as an alternative treatment strategy but carries the risk of loosing the window of resectability. Risks of failure therefore exist throughout both treatment pathways. Recent decision-analysis have suggested that the selection of optimal treatment pathways depends on individual patient and tumour factors. This paper presents a Bayesian Belief Network (BBN) model that evaluates the risk of failure and consequence factors across surgery-first and neoadjuvant pathways. BBN is capable of providing personalized predictions of good prognosis (3year survival) post resection of pancreatic ductal adenocarcinoma. This can support clinical decision action and patient counseling by identifying superior pathway selection at individualized level, and justify resource allocation by identifying patients at higher risk of failure who require additional actions to reduce risk.

Methods 1:

BBN, also referred to as acyclic directed graphs, is a graphical model based on probability theory that models relationships between variables. BBN consist of variables, know as nodes, with arcs depicting informational or causal relationships from parent to child nodes. Each node has a defined and exclusive set of states and the dependencies between nodes are quantified through a set of conditional probability tables (CPTs) whereby the conditional probability of a child node is defined by the state of each of its parent nodes. Nodes that do not have parent nodes are reduced to the unconditional probability (UP) structure. Where the UPs of a basic input parameter is not known a priori, equal weight are applied to states through the principle of insufficient reasoning.

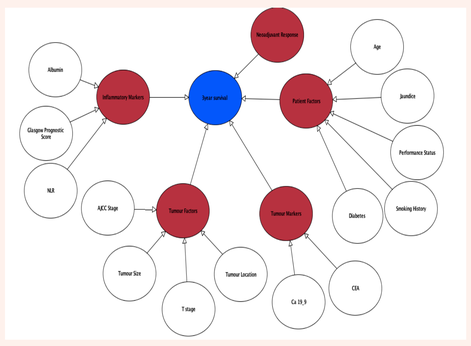
Through Baye's theorem, BBN can explicitly represent the conditional probability dependencies between variables, which has been proven to be an effective way of handling uncertainty. In a BBN, the updated probability for *n* number of mutually exclusive parameters X_i, where $_{(i= 1,2,3...n)}$, and given observed data, *Y*, can be computed as: $p(X_i|Y)= pYXix\ p(Xi)\sum jpYXjp(Xj$ where the posterior probability of *X* given the condition that *Y* occurs is represented by $p(X|Y)$, the posterior probability of *Y* given the condition that *X* occurs is represented by $p(Y|X)$, the prior occurrence probability of *X* is denoted by $p(X)$ and the marginal occurrence of *Y* is denoted as $p(Y)$. In this sense this is often viewed as the likelihood distribution.

This holds several advantages when modeling treatment pathways for potentially resectable PDAC. Through Bayes theorem the prior distribution and observed data are combined to update knowledge in the form of the posterior distribution. Therefore BBNs allow the modeling of relationships between variables at various stages of the healthcare process, with predictions of outcomes evolving throughout the process by utilizing all available patient data at that time. This means that the model could not only make predictions of outcome pre-operatively but also perform prognostic updating at the post-operative stage of the patient journey. Where patient information is limited probabilistic inference can still make predictions based on global averages of the patient population. As more information becomes available the predictions become more patient specific. Furthermore, BBN have the flexibility to perform bottom-up inference (inferring the state of the parent node from the observed state of the child node) and top-down inference (inferring the state of the child node given the observed state of the parent node). This is also known as diagnostic and decision analysis respectively and is called marginalization which is employed to compute the reliability of networks based on statistical data. For the pancreatic cancer model this process will also allow the testing of the model to perform scenario, or "what if" testing to anticipate cause and impact of failure events such as developing side effects from therapy or complications from surgery.

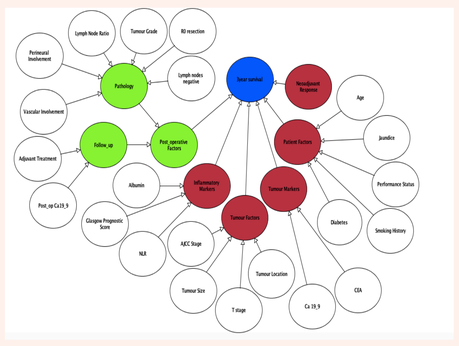
Methods 2:

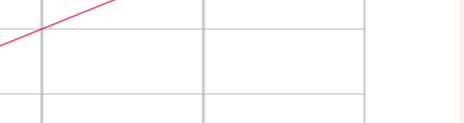
A weighted prognostic BBN was created using AgenaRisk software. The BBN was based on data from survival analysis studies gathered from a comprehensive search of PubMed database (n=48691) that was synthesized based on a two-stage weighting process. The model was validated against a prospectively maintained patient database from a tertiary referral centre (n=365). Variables know at the pre-operative stage of the patient journey were included in the pre-operative BBN and post-operative factors were added to the model to perform prognostic updating. 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.

Pre-Operative Predictive BBN



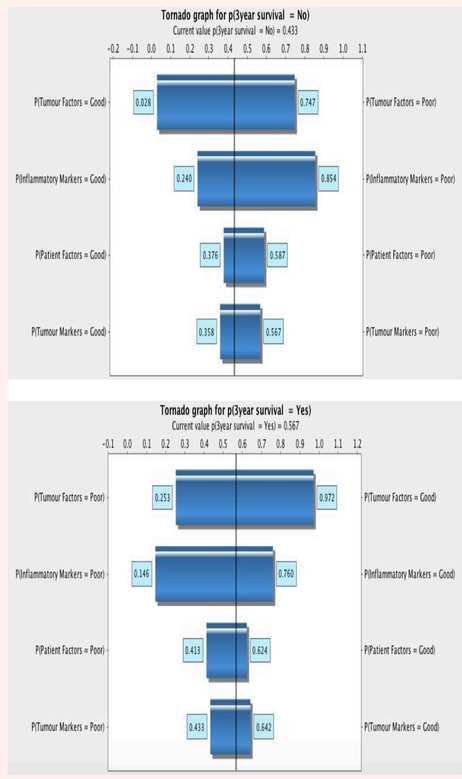
Post-Operative Prognostic Updating BBN

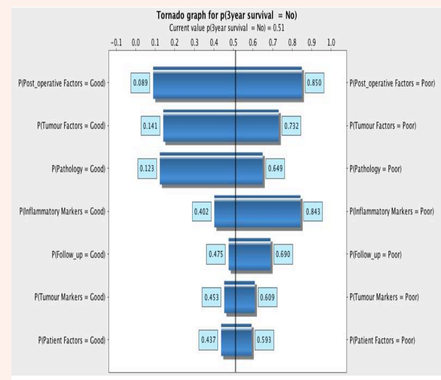


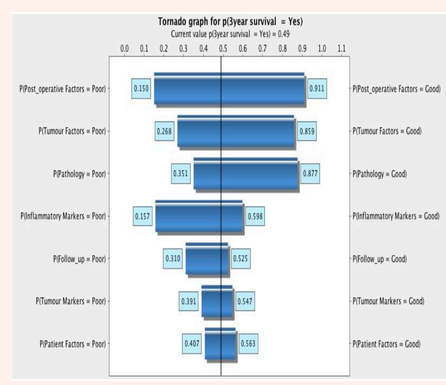


Methods 3: Sensitivity Analysis

Pearl's inwards analysis and broadcasting analysis were used to perform sensitivity analysis. This 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.



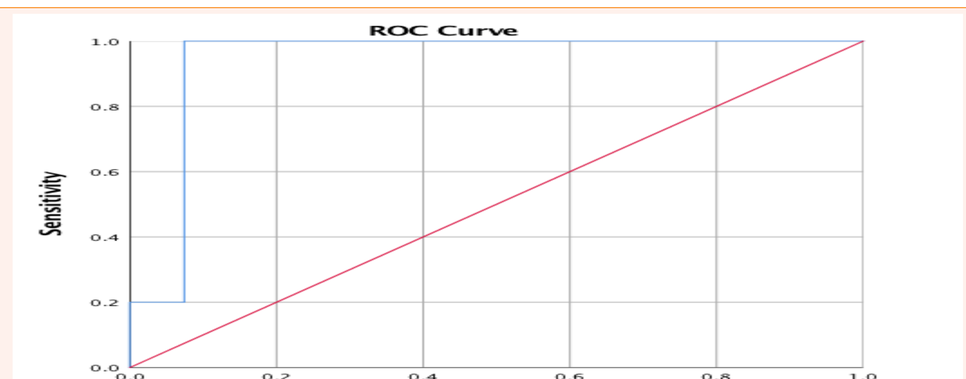




Results:

Pre-operative Results

	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Accuracy (ACC)	Area Under Curve (AUC)
Complete data set (n=32)	1	0.93	0.71	1	0.94	0.94 (*P* value 0.002; 95% CI 0.856-1.0) Std. Error 0.043
1 data point missing (n=119)	0.85	0.90	0.71	0.95	0.89	0.868 (*P* value 0.000; 95% CI 0.780-0.956) Std. Error 0.45
2 data points missing (n=132)	0.83	0.92	0.75	0.95	0.90	0.873 (*P* value 0.000; 95% CI 0.793-0.953) Std. Error 0.41
3 data points missing (n=135)	0.83	0.92	0.73	0.95	0.90	0.871 (*P* value 0.000; 95% CI 0.790-0.951) Std. Error 0.041
4 data points (n=175) missing	0.83	0.89	0.61	0.96	0.88	0.831 (*P* value 0.000; 95% CI 0.746-0.917) Std. Error 0.044
4+ data points missing (n=365)	0.62	0.86	0.42	0.93	0.82	0.735 (*P* value 0.000; 95% CI 0.660-0.809) Std. Error 0.038



Post-operative Results

	0 Missing Post-operative Data points	0-1 Missing Post-operative Data Point	0-2 Missing Post-operative Data Points	0-3 Missing Post-operative Data Points	0-4 Missing Post-operative Data Points
0 Missing Pre-operative Data Points	Sensitivity: 1 Specificity: 0.96 PPV: 0.86 NPV: 1 ACC: 0.97 AUC 0.97, P-value 0.000 (95% CI 0.908-1.000) (n=33)				
0-1 Missing Pre-operative Data Point	Sensitivity: 0.62 Specificity: 0.96 PPV: 0.83 NPV: 0.89 ACC: 0.83 AUC 0.81, P-value 0.000 (95% CI 0.705-0.910) (n=133)			Sensitivity: 0.62 Specificity: 0.96 PPV: 0.83 NPV: 0.89 ACC: 0.83 AUC 0.81, P-value 0.000 (95% CI 0.706-0.911) (n=114)	
0-2 Missing Pre-operative Data Points	Sensitivity: 0.62 Specificity: 0.90 PPV: 0.67 NPV: 0.88 ACC: 0.83 AUC 0.880, P-value 0.000 (95% CI 0.659-0.900) (n=121)	Sensitivity: 0.84 Specificity: 0.63 PPV: 0.56 NPV: 0.88 ACC: 0.69 AUC 0.880, P-value 0.000 (95% CI 0.702-0.911) (n=122)	Sensitivity: 0.62 Specificity: 0.91 PPV: 0.67 NPV: 0.89 ACC: 0.84 AUC 0.880, P-value 0.000 (95% CI 0.701-0.901) (n=126)		Sensitivity: 0.62 Specificity: 0.91 PPV: 0.67 NPV: 0.89 ACC: 0.84 AUC 0.80, P-value 0.000 (95% CI 0.726-0.902) (n=127)
0-3 Missing Pre-operative Data Points	Sensitivity: 0.62 Specificity: 0.91 PPV: 0.67 NPV: 0.89 ACC: 0.84 AUC 0.81, P-value 0.000 (95% CI 0.708-0.904) (n=130)			Sensitivity: 0.62 Specificity: 0.90 PPV: 0.64 NPV: 0.83 ACC: 0.80 AUC 0.81, P-value 0.000 (95% CI 0.705-0.902) (n=131)	
0-4 Missing Pre-operative Data Points	Sensitivity: 0.62 Specificity: 0.91 PPV: 0.67 NPV: 0.89 ACC: 0.84 AUC 0.81, P-value 0.000 (95% CI 0.711-0.905) (n=131)	Sensitivity: 0.62 Specificity: 0.91 PPV: 0.64 NPV: 0.89 ACC: 0.84 AUC 0.81, P-value 0.000 (95% CI 0.709-0.905) (n=138)			Sensitivity: 0.62 Specificity: 0.91 PPV: 0.64 NPV: 0.89 ACC: 0.84 AUC 0.80, P-value 0.000 (95% CI 0.705-0.903) (n=140)
>4 Missing Pre-operative Data Points	Sensitivity: 0.63 Specificity: 0.92 PPV: 0.70 NPV: 0.90 ACC: 0.86 AUC 0.880, P-value 0.000 (95% CI 0.721-0.969) (n=136)	Sensitivity: 0.66 Specificity: 0.92 PPV: 0.68 NPV: 0.91 ACC: 0.85 AUC 0.882, P-value 0.000 (95% CI 0.736-0.910) (n=172)	Sensitivity: 0.54 Specificity: 0.93 PPV: 0.68 NPV: 0.87 ACC: 0.83 AUC 0.880, P-value 0.000 (95% CI 0.655-0.838) (n=230)		

The BBN demonstrated a high level of predictive performance. Benefits over existing models include: pre-operative application, greater generalizability, and ability to encompass both upfront surgery and neoadjuvant treatment options. This marks an important step towards achieving the delivery of precision medicine, as the next step will be to incorporated genomic data into the model hence combining genetic, pathology and clinical data, creating a vehicle to deliver personalized precision medicine.