Title: Personalized Pancreatic Cancer Management: A Systematic Review of How Machine Learning is Supporting Decision-making

Running Head: Hope Is Replacing Hype But Challenges Remain

Authors: Alison Bradley, MRCSEd,*†

Robert van der Meer, PhD,*

Colin McKay, FRCS,†

*Department of Management Science, University of Strathclyde Business School, Glasgow, Scotland

†West of Scotland Pancreatic Cancer Unit, Glasgow Royal Infirmary, Glasgow, Scotland

Address Correspondence to: Alison Bradley, 34 Nasmyth Avenue, Bearsden, Glasgow, G61 4SQ, United Kingdom (e-mail: bradley_alison@live.co.uk). Tel: 07763191494 Address:

Abstract: This review critically analyzes how machine learning is being utilized to support clinical decision-making in the management of potentially resectable pancreatic cancer. Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, electronic searches of MEDLINE, Embase, PubMed and Cochrane Database were undertaken. Studies were assessed using the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modeling Studies (CHARMS) checklist. In total 89,959 citations were retrieved. Six studies met the inclusion criteria. Three studies were Markov decision-analysis models comparing neoadjuvant therapy versus upfront surgery. Three studies predicted survival time using Bayesian modeling ($n = 1$), Artificial Neural Network ($n = 1$), and one study explored machine learning algorithms including: Bayesian Network, decision trees, k-nearest neighbor, and Artificial Neural Networks. The main methodological issues identified were: limited data sources which limits generalizability and potentiates bias, lack of external validation, and the need for transparency in methods of internal validation, consecutive sampling, and selection of candidate predictors. The future direction of research relies on expanding our view of the multidisciplinary team to include professionals from computing and data science with algorithms developed in conjunction with clinicians and viewed as aids, not replacement, to traditional clinical decision-making.

Introduction

The management of pancreatic cancer is particularly challenging.[1,2] Surgical resection is the only potentially curative treatment yet despite advances in surgical technique and adjuvant therapies 10-year survival remains less than 1%.[3] Surgical risks remain high with any potential benefit often nullified by early disease reoccurrence.[4,5]

Several factors have aligned making decision-making in the management of pancreatic cancer more complex. The ageing population and obesity epidemic means patients are amassing a greater amount of clinical data to consider when making clinical decisions.[6] Treatment options are expanding with the emergence of neoadjuvant approach as an alternative to upfront surgery. While some are optimistic about the role of neoadjuvant therapy, others feel the current body of evidence is at best ambiguous with its role in the management of resectable pancreatic cancer being particularly controversial.[7-12] This is compounded by the current lack of randomized controlled trials comparing both upfront surgery and neoadjuvant treatment pathways.[12] Furthermore through the emergence of precision medicine databases will expand to reflect our understanding of disease at genomic level, creating a further 'data explosion'.[13] Patients therefore represent a big data challenge not only in amount of data amassed, but in being extremely complex data systems with multidimensional problems and interacting parameters with the rules governing behaviours within layers of these systems often unclear or simply unknown.[14]

Personalized predictive modeling has gained precedence as a means of supporting clinical decision-making.[15] However, existing predictive models, mainly based on non-liner regression techniques are limited in scope and volume regarding prognosis

as an isolated event at a pre-determined time.[15,16] Machine learning methods make predictions within complex systems against a background of competing risks and events.[14] Machine learning achieves this in one of three ways. Firstly supervised learning, where the computer utilizes partial labeling of data.[17,18] Alternatively unsupervised learning allows the computer to make predictions or explain data by utilizing structures detected within the data itself.[17,18] Thirdly reinforcement learning whereby, similar to operant conditioning,[19] the computer learns from its mistakes and successes to achieve a task.[17,20] Commonly employed methods of machine learning include, but are not limited to: Bayesian networks (BN), artificial neural networks (ANN) and Fuzzy-logic (FL) modeling.[14] Table 1 outlines the definition, strengths and limitations of these most commonly employed methods of machine learning.

In isolation the factors outlined as contributing to the complexity of decision-making may not be unique to pancreatic cancer. However, in the context of being one of the most challenging malignancies,[1,2] with comparatively lower resection rates compared to other gastrointestinal malignancies,[1-3] pancreatic cancer is the ideal vehicle to critically examine how successful machine learning is in dealing with complexity and uncertainty to support clinical decision-making.

The aim of this review is to use the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modeling Studies (CHARMS) checklist[21] to critically analyze the use of machine learning for decision-analysis, prognostic and predictive purposes to support clinical decision-making in the management of potentially resectable pancreatic cancer.

METHODS

The protocol for this review was published in the PROSPERO online database of systematic reviews (CRD42018108926). Electronic searches of MEDLINE, Embase, PubMed and Cochrane Database were undertaken with the entire databases included up to and including 14th September 2018, with no further date restrictions or limits applied. Full search strategies are detailed in Supplementary Digital Content 1. This review followed the PRISMA checklist.[22]

Study Selection

Manual screening was carried out after removal of duplicates, based on the title and abstract of articles identified in the database searches. Articles of probable or possible relevance were reviewed in full. This was decided based on the inclusion criteria of machine learning methods applied to pancreatic cancer management decision support, published in full-text in English language. This included predictive and prognostic modeling and decision-analysis studies. Application of machine learning to diagnosis, interpretation of diagnostic imagery, and risk of developing pancreatic cancer were excluded. The focus was on application of machine learning to support clinical decision-making in the management of pancreatic cancer, following diagnosis, including prediction of survival, quality adjusted survival, and complications of treatments. The reason for this focus is that it is anticipated that personalized predictive medicine will be able to forecast individualized outcomes across competing treatment strategies to facilitate clinical decision-making. Given the afore mentioned ambiguities and controversies regarding the best management pathways for potentially resectable pancreatic cancer,[7-12] this is the ideal vehicle through which to assess the contribution of machine learning in supporting clinical decision-making under uncertainty across the ever evolving patient journey. Reference lists and

citations of all included papers were manually searched to identify any additional articles. This process was repeated until no new articles were identified.

Data Extraction, Quality Assessment, and Data Analysis

Search design and data extraction were performed by the lead reviewer and with second author performing independent quality assurance. Discrepancies were resolved by inter-reviewer discussion. For predictive models data was extracted according to the CHARMS checklist, which also includes assessment of risk of bias.[21] This checklist is designed for appraisal of all primary prediction modeling studies including ANN and vector machine learning.[21]

RESULTS

Search Results

Abstracts of 89,959 citations were retrieved. Six studies met the inclusion criteria of machine learning methods applied to predictive modeling and decision-analysis related to pancreatic cancer management (Fig. 1).

Three studies were Markov decision-analysis models comparing two competing treatment options: neoadjuvant therapy versus upfront surgery.[23-25] Three studies focused on predicting survival time.[26-28] One of these studies also predicted lymph node ratio.[26] One of these studies additionally explored prediction of Eastern Cooperative Oncology Group (ECOG) quality-of-life scores, surgical outcomes and tumour characteristics.[27] One study performed direct comparison between predictive accuracy of machine learning techniques and linear and logistic regression.[27]

Three studies used Marko decision tree models,[23-25] 1 study used Bayesian modeling,[26] 1 study used ANN,[28] and 1 study explored machine learning algorithms including: BN, decision trees, *k*-nearest neighbor, and ANN (Table 2).[27]

Decision-analysis Models

Three studies attempted to employ Markov decision analysis to compare upfront surgery and neoadjuvant approach.[23-25] Sharma et al[24] used data drawn from 21 prospective phase II and III trials. De Gus et al[23] also included data from retrospective studies compiled from a literature search from a single search engine. Both these studies, although reportedly analyzing strategies for resectable pancreatic cancer used studies that included borderline resectable and locally advanced pancreatic cancer in an intention-to-treat analysis.[23,24] All 3 studies used an intention-to-treat approach to analysis and, although they reported a slight benefit from neoadjuvant approach, neither strategy was conclusively superior.[23-25] All three existing studies were solely based on synthesized evidence from published trials therefore share the limitations of the existing body of evidence mainly: heterogeneity and small underpowered sample size.[7,8]

Prediction Models

A cohort design, commonly recommended for prognostic model development,[29] was used for all 3 predictive models.[26-28] Two studies used retrospective single center databases (ANN, $n = 219$[28]; comparison study, $n = 91$[27]), which can limit generalizability, and 1 study used cancer data registry (BN, $n = 6400$).[26] Prospective cohort designed is recommended as it enables optimal measurement of predictors and outcome.[30] Retrospective cohorts are thought to yield poorer quality data but do enable longer follow-up time.[29]

Participant recruitment with inclusion criteria and description of cohort characteristics were well reported, as were study dates in all 3 studies.[26-28] Length of follow-up time was clear in 2 studies.[26,28] Consecutive sampling was reported in 1 study[28] but whether all consecutive participants were included, or number of participants who

refused to participate, could not be evaluated in any of the 3 studies.[26-28] Non-consecutive sampling can introduce a risk of bias.[31-33]

In all 3 studies outcomes were clearly defined with the same outcome definition and method of measurement applied to all patients.[26-28] The interactive Bayesian model predicted 6 month, 1-, 3- and 5-year survival post resection and lymph node ratio.[26] The ANN predicted 7-month mortality after resection.[28] Hayward et al focused on data mining techniques but treated survival outcome as a time-dependent-event for resected and un-resected patients, with ECOG measured at 6 months post-resection.[27] Number of candidate variables ranged from 7 to 19. The definition, method and timing of measurement of candidate predictors were clear in all 3 studies.[26-28] How candidate predictors were selected were not made clear in 2 studies[26,28] which may be illustrative of the non-transparent 'black-box' analysis sometimes employed by forms of artificial intelligence (AI). One study extensively explored algorithms for data mining and categorization of the datasets.[27] The other two studies used backward elimination methods.[26,28] The ANN used single hidden layer back propagation to train the model,[28] and the Bayesian model employed backward step down selection process.[26] All 3 studies used complete case analysis.[26-28] This approach results in loss of statistical power and can introduce bias as missing data rarely occurs randomly and often pertains to participant or disease characteristics.[30]

None of the studies underwent external validation. The interactive Bayesian model[26] and ANN[28] employed random split technique between training and validation datasets. This points to a potential key weakness in the application of machine learning techniques as random split technique can result in over and under fitting of the model, particularly as details of cross validation were not given.[34] Techniques of data splitting are poorly described and can result in a high degree of variance of

model performance.[34] More sophisticated techniques of data splitting that exploit the structure of the data exist and provide more confident results, but at higher computational cost.[34] Only the interactive Bayesian model reported calibration with goodness-of-fit statistic ($P = 0.300$ for prediction of lymph-node-ratio; $P = 0.4847$ for survival prediction).[26] The ANN reported discrimination as area under curve (AUC) of the receiver operated curve (ROC) (AUC, 0.6576; sensitivity, 91.30%; specificity, 38.27%).[28] The interactive Bayesian model reported discrimination as c-statistic (0.65; 95% confidence interval [CI], 0.63–0.66).[26] Although commonly used, the c-statistic can be influenced by predictor value distribution and be insensitive to inclusion of additional predictors.[21] The study by Hayward *et al.* compared machine learning to log regression and found that for survival prediction Bayesian modeling out performed log regression (accuracy 0.60 versus 0.42).[27] Furthermore in predicting outcome for ECOG at 6 months post-resection log regression performance improved from r-squared value, 0.26 to 0.32 when modified with machine learning algorithm 'linear regression with bagging'.[27]

DISCUSSION

Principal Findings

Our review found 6 studies applying machine learning techniques to support clinical decision-making in the management of pancreatic cancer. Three studies used Markov decision tree models to perform decision analysis.[23-25] Three studies used machine learning methods for predictive modeling: 1 study used Bayesian modeling,[26] 1 study used ANN,[28] and 1 study explored machine learning algorithms including: BN, decision trees, k-nearest neighbor, and ANN.[27] The main issues identified with decision-analysis studies were reliance on data from a single database search and the quality of the existing studies pertaining to treatment of potentially resectable

pancreatic cancer being mainly small and underpowered with a high degree of heterogeneity.[7,8] The issues identified with the predictive models were overreliance on single institution retrospective database, which could affect generalizability. There was also a lack of clarity as to whether consecutive sampling was employed and how candidate predictors were selected. A major issue identified was lack of external validation across all 3 predictive models. Although 2 studies used random-split technique, details of cross-validation were not provided which potentiates issues of over or under fitting. Only one study reported calibration of their model.

Implications for Research and Practice

Machine learning, although in its infancy, holds great potential in its application to decision-making under complexity.[14,35] However the application of machine learning to predictive modeling pertaining to the management of pancreatic cancer is currently limited in number therefore no conclusion can yet be drawn as to superiority of either machine learning or traditional modeling approaches. Only one study directly compared machine-learning methods with traditional approach to modeling.[27] Accuracy of machine learning predictions, particularly Bayesian modeling, were found to be superior and predictions form log regression approach were improved when combined with machine learning techniques.[27] However, it is important to note that of existing predictive studies using machine learning limitations in methodological approach were identified using the CHARMS checklist.[21] These issues are similar to issues highlighted in traditional approaches to predictive modeling and include: use of single centre database limiting generalizability, sample size, lack of blinding, transparency in candidate predictor selection, and lack of external validation.[21, 29-32]

Whilst much optimism surrounds the growing use for AI in healthcare delivery, machine learning also carries limitations that must be addressed in future research. Machine learning usually requires large amounts of data,[36] which in the case of management of potentially resectable pancreatic cancer can be difficult to obtain as the majority of patients present with advanced, unresectable disease.[1-3] Whilst the creation of national shared databases may be one solution to increase the volume of data, this is not without issue including dimensionality, missing data and control of bias[37,38] with minority groups often under represented in such databases.[38] Furthermore simply increasing volume of data is not the solution as machine learning is not yet at a stage where it can distinguish correlation and causation.[36] Future research should focus on better integration of machine learning with expert knowledge to overcome this challenge.[36] This review found little evidence of machine learning being actively integrated into clinical practice. Whilst this is mainly due to such techniques being in their infancy, it must also be acknowledged that some machine learning techniques are not yet sufficiently transparent which breeds distrust and resistance to their clinical application.[36] Machine learning requires high levels of technical skill and can be difficult to engineer with experts from medicine, computing and data sciences often speaking in different technical language and coming to problems from different perspectives which can inhibit shared understanding and limit achievement of its full potential.[36] The solution therefore lies with clinicians expanding their view of the multidisciplinary team to include professionals from computing and data science backgrounds with algorithms developed in conjunction with clinicians and viewed as aids, not replacement, to traditional clinical decision-making.[6]

Despite these challenges the study by Hayward et al[27] does however corroborate other studies where application of machine-learning methods to: breast, prostate and bladder cancers have demonstrated superiority in terms of accuracy of predictions over traditional log regression.[39-42] Artificial Neural Networks have also been found to perform as well as or better than traditional log regression models and also improve the diagnosis and management of pancreatitis and diagnosis pancreatic cancer.[35] Machine learning methods have also been shown to out perform log regression in: providing individualized prediction of the need for neonatal resuscitation,[43] predicting early mortality risk in coronary artery bypass graft surgery[44] and predicting severely depressed left ventricular ejection fraction following admission to intensive care unit.[45]

Strengths and Limitations

Although limited to studies available in English language, this review is the first of its kind and goes beyond the hype surrounding use of machine learning and AI in supporting clinical decision-making to ascertain what the current reality of its application to clinical practice in management of potentially resectable pancreatic cancer actually is. Furthermore it highlights limitations that should be addressed in future research. The current number of existing studies is limited therefore conclusions about superiority of machine learning over traditional predictive modeling techniques are limited.

CONCLUSION

To conclude clinical decision-making is going to become increasingly complex as our understanding of disease and treatment response at genomic level grows, resulting in a further 'data explosion'.[6,13,14] Utilizing this expanse of data to facilitate decision-making in a meaningful way for individual patients is beyond the capabilities of the

human mind working in isolation.[6,14,35] It is in this context that machine learning

holds the greatest potential by being able to handle large amounts of data and

integrate large, complex and varied databases.[35] However machine learning also

carries limitations and, whilst initial studies are promising, its application has yet to

be widely tested.[36] The future direction of research therefore relies on expanding our

view of the multidisciplinary team to include professionals from computing and data

science backgrounds with algorithms developed in conjunction with clinicians and

viewed as aids, not replacement, to traditional clinical decision-making.[6]

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics 2015. *CA Cancer J Clin*. 2015;6:5-29.

2. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49:1374-1403.

3. Pancreatic Cancer UK. Pancreatic Cancer UK policy briefing: every life matters: the real cost of pancreatic cancer diagnosis via emergency admission. Available at https://www.pancreaticcancer.org.uk/media/86662/every-im_policybriefing-final.pdf. Accessed September 2, 2017.

4. Winter JM, Brennan MF, Tang LH, et al. Survival after Resection of Pancreatic Adenocarcinoma: Results from a Single Institution over Three Decades. *Ann Surg Oncol*. 2012;19:169.

5. Lewis RS, Vollmer CM. Risk scores and prognostic models in surgery: pancreas resection as a paradigm. *Curr Probl Surg*. 2012;49:731-795.

6. Obermeyer ZMD, Lee TH. Lost in Thought — The Limits of the Human Mind and the Future of Medicine. *N Engl J Med*. 2017;377:1209-1211.

7. Tempero MA, Malafa MP, Behrman SW. Pancreatic adenocarcinoma, version 2.2014: featured updates to the NCCN guidelines. *J Natl Compr Canc Netw*. 2014;12:1083–1093.

8. Asare EA, Evans DB, Erickson BA, et al. Neoadjuvant treatment sequencing adds value to the care of patients with operable pancreatic cancer. *J Surg Oncol.* 2016;114:291-295.

9. Lee J, Ahn S, Paik K, et al. Clinical impact of neoadjuvant treatment in resectable pancreatic cancer: a systematic review and meta-analysis protocol. *BMJ*. 2016;6:1-9.

10. Xu CP, Xue XJ, LaingN, et al. Effect of chemoradiotherapy and neoadjuvant chemoradiotherapy in resectable pancreatic cancer: a systematic review and meta-analysis. *J Cancer Res Clin Oncol*. 2014;140:549-559.

11. Andriulli A, Festa V, Botteri E, et al. Neoadjuvant/preoperative gemcitabine for patients with localized pancreatic cancer: a meta-analysis of prospective studies. *Ann Surg Oncol*. 2012;19:1644-1662.

12. Versteijjne E, Vogel JA, Besselink MG, et al. Meta-analysis comparing upfront surgery with neoadjuvant treatment in patients with resectable or borderline resectable pancreatic cancer. *BJS*. 2018;105: 946-958.

13. Tonelli MR, Shirts BH. Knowledge for Precision Medicine Mechanistic Reasoning and Methodological Pluralism. *JAMA.* 2017;318:1649–1650.

14. Abbod MF, Hamdy FC, Linkens DA, et al. Predictive modeling in cancer: where systems biology meets the stock market. *Expert Review of Anticancer Therapy*. 2014;9:867-870.

15. Velikova M, Scheltinga JT, Lucas PJF, et al. Exploiting causal functional relationships in Bayesian network modeling for personalized healthcare. *Int Journal of Approximate Reasoning*. 2014;55:59-73.

16. Verduijn M, Peek N, Rosseel PMJ, et al. Prognostic Bayesian networks I: rationale, learning procedure, and clinical use. *Journal of Biomedical Informatics.* 2017; 609-618.

17. Hashimoto DA, Rosman G, Rus D, et al. Artificial Intelligence in Surgery: promises and perils. *Annals of Surgery*. 2018;286:70-76.

18. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920-1930.

19. Skinner BF. The behaviour of organisms: An experimental analysis. New York, NY: Appleton-Century-Crofts; 1938.

20. Sutton RS, Barto AG. Reinforcement learning: An Introduction. Vol. 1. Cambridge: MIT Press; 1998.

21. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med.* 2014;11:e1001744.

22. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med*. 2009;3:123-130.

23. deGus SW, Evans DB, Bliss LA, et al. Neoadjuvant therapy versus upfront surgical strategies in resectable pancreatic cancer: a Markov decision analysis. *Eur J Surg*. 2016;42:1552-1560.

24. Sharma G, Whang EE, Ruan DT, et al. Efficacy of neoadjuvant versus adjuvant versus adjuvant therapy for resectable pancreatic adenocarcinoma: a decision analysis. *Ann Surg Oncol*. 2015;22:1229-1237.

25. VanHouten JP, White RR, Jackson GP. A Decision Model of Therapy for Potentially Resectable Pancreatic Cancer. *J Surg Res*. 2012;174:222-230.

26. Smith BJ, Mezhir JJ. An interactive Bayesian model for prediction of lymph node ratio and survival in pancreatic cancer patients. *J Am Med Inform Assoc*. 2014;21:203-211.

27. Hayward J, Alvarez SA, Ruiz C, et al. Machine learning of clinical performance in a pancreatic cancer database. *Artificial Intelligence*. 2010;49:187-195.

28. Walczak S, Velanovich V. An evaluation of Artificial Neural Networks in predicting pancreatic cancer survival. *J Gastrointest Surg*. 2017;21:1606-1612.

29. Moons KG, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:375.

30. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLoS Med*. 2012;9:e1001221.

31. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ*. 2001;323:224–228.

32. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:1432–1435.

33. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001:134;663–694.

34. Reitermanov́a Z. Data Splitting. WDS'10 Proceedings of Contributed Papers, Part I. 2010; 31–36.

35. Bartosch- Härlid A, Andersson B, Aho U, et al. Artificial neural networks in pancreatic disease. *Br J Surg*. 2008;95:817-826.

36. Marcus G. Deep learning: a critical appraisal. *arXiv*. 2018; 1-27.

37. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Research and Clinical Practice*. 2017;36:3–11.

38. Zhang X, Perez-Stable EJ, Bourne P, et al. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century. *Ethnicity & Disease*. 2017; 27:95–106.

39. Seker, H. A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. *IEE Transactions on Information Technology in Biomedicne.* 2003;2:114-122.

40. Catto JW, Linkens DA, Abbod MF, et al. Artificial intelligence in predicting bladder cancer outcome: a comparison of neuro-fuzzy modelling and artificial neural networks. *Clin Cancer Res*. 2003;9:4172-4177.

41. Abbod, M.F. Linkens DA, Catto JWF, et al. Comparative study of intelligent models for the prediction of bladder cancer progression. *Oncology*. 2006;15:1019-1022.

42. Catto JW, Abbod MF, Linkens DA, et al. Neurofuzzy modelling to determine recurrence risk following radical cystectomy for nonmetastatic urothelial carcinoma of the bladder. *Clin Cancer Res*. 2009;15:3150-3155.

43. Reis MA, Ortega NR, Silveira PS. Fuzzy expert system in the prediction of neonatal resuscitation. *Braz J Med Biol Res*. 2004;37:755-765.

44. Ghavidel AA, Javadikasgari H, Maleki M, et al. Two new mathematical models for prediction of early mortality risk in coronary artery bypass graft surgery. *J Thorac Cardiovasc Surg*. 2014;148:1291-1298.

45. Pereira RDMA, Salgado CM, Dejam A, et al. Fuzzy modelling to predict severely depressed left ventricular ejection fraction following admission to the intensive care unit using clinical physiology. *The Scientific World Journal*. 2015;1-9.

46. Lucas PJF, Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine.* 2004;201-214.

47. Schoot R, Kaplan D, Denissen J, et al. A gentle introduction to bayesian analysis: applications to developmental research. *Child Dev.* 2013;85:842-860.

48. Hampson LV, Whitehead J, Eleftheriou D, et al. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Stat Med.* 2014;33:4186-4201.

49. Johnson SR, Tomlinson GA, Hawker GA, et al. Methods to elicit beliefs for Bayesian prior: a systematic review. *J Clin Epidemiol.* 2010;63:355-369

50. Gursel G. Healthcare, uncertainty, and fuzzy logic. *Dig Med.* 2016;2:101-112

51. Barro S, Marín R, editors. Fuzzy Logic in Medicine. Vol. 83. Berlin Heidelberg, Germany: Springer; 2002.

52. Dweiri FT, Kablan MM. Using fuzzy decision making for the evaluation of the project management internal efficiency. *Decis Support Syst.* 2006;42:712-726.

53. Grossi E. Medical concepts related to individual risk are better explained with "plausibility" rather than "probability". *BMC Cardiovasc Disord.* 2015;5:31

54. McNeill FM, Thro E. Fuzzy Logic: A Practical Approach. Cambridge, MA: Academic Press Professional; 1994

55. Bouchon-Meunier B, Zadeh LA. Fuzzy logic and its applications. Paris, France: Addison-Wesley; 1995.

56. Roychowdhury A, Pratihar DK, Bose N, et al. Diagnosis of the diseases-using a GA-fuzzy approach. *Inf Sci.* 2004;162:105-120.

57. Pratihar DK, Deb K, Ghosh A. A genetic-fuzzy approach for mobile robot navigation among moving obstacles. *Int J Approx Reason.* 1999;20:145-172.

FIGURE LEGENDS

FIGURE 1. PRISMA Flow Chart

Supplementary Digital Content

Supplementary Digital Content 1.doc: List of search terms used

Tables:

TABLE 1. Summary of Common Methods of Machine Learning			
Method	**Application**	**Strengths**	**Limitations**
Bayesian Network (BN)	Decision support Risk Assessment Prediction[14-16,46]	Allows for incorporation of individual patient data, disease progression and impact of different treatment options on the predicted outcome.[15,16] Facilitates prognosis updating and scenario testing.[16] Provides information on process and outcome variables therefore predict outcomes pertaining to quality and not just amount of survival time.[16,46] Uses probabilistic inference when data is limited and can still make predictions based on global averages of the patient population.[16,46]	Accurate use of data in elicitation of priors is an area of ongoing investigation and debate.[46-49] An over reliance on machine-learned network structures, could mean fundamental causal relationships well established in medical knowledge are lost hence limiting the applicability.[46-49] Can only model linear dependencies.[14]
Artificial Neural Network (ANN)	Modeling Prediction Image interpretation Classification[14]	Models non-linearity and complex relationships.[14,35] Handles high-dimension problems.[14,35] Can generalize.[35] Does not impose any restrictions on the input variables.[14,35]	Heavy data requirements with long training times requiring many design decisions.[14,35] May not generalize well to other data sets.[14,35] Lacks transparency.[14,35]
Fuzzy Logic (FL)	Modeling Prediction Classification[14]	Models non-linearity.[14] Handles uncertainty and complexity.[49-52] Enables prediction to move from probability to plausibility.[53] Transition to a contiguous value is gradual rather than abrupt reflecting human decision-making processes.[50,54,55] Can assess more observed variables yet fewer values are required.[54,55] Transparent.[14]	Extensive expert knowledge of the system to be modeled is required.[56] Requires more fine-tuning and simulation prior to being operational.[57] Cannot model high-dimension problems.[14]

TABLE 2.Summary of Included Studies

Study	Participant Population	Method	Outcome Measure	Main Limitations
deGus et al, 2016[23]	Synthesized data from phase II trials and cohort studies	Markov decision-analysis	Survival in months and quality adjusted life months for upfront surgery versus neoadjuvant therapy	Use of single electronic database of journals Synthesized small underpowered studies with high level if heterogeneity Relied heavily on retrospective cohort studies
Sharma et al, 2015[24]	Synthesized data from phase II trials	Markov decision-analysis	Survival in months and quality adjusted life months for upfront surgery versus neoadjuvant therapy	Use of single electronic database of journals Synthesized small underpowered studies with high level if heterogeneity
VanHouten et al, 2012[25]	Synthesized data from phase II trials and cohort studies	Markov decision-analysis	Survival in months and quality adjusted life months for upfront surgery versus neoadjuvant therapy	Use of single electronic database of journals Synthesized small underpowered studies with high level if heterogeneity Included borderline resectable cases in neoadjuvant cohort Relied heavily on retrospective cohort studies
Smith & Mezhir, 2014[26]	Cancer Registry (n = 6400)	Interactive Bayesian Model	Survival at 6 months, 1,3 and 5-year survival	Follow-up time unclear Unclear if consecutive sampling used Selection method of candidate predictors not clear Complete base analysis used No external validation
Walczak &Velanovich, 2017[28]	Retrospective single institution database (n = 219)	Artificial Neural Network	Death at 7 months post resection	Consecutive sampling used but unclear all consecutive participants included Selection method of candidate predictors not clear Complete base analysis used No external validation No calibration
Hayward et al, 2010[27]	Retrospective single institution database (n = 91)	Machine learning algorithms including: Bayesian Network, decision trees, *k*-nearest neighbor, and ANN	Survival as time dependent event, Eastern Cooperative Oncology Group (ECOG) quality-of-life scores measured at 6 months	Unclear if consecutive sampling used Complete base analysis used No external validation No calibration

Figure 1

PRISMA 2009 Flow Diagram

Identification

Records identified through database
searching
(n = 89,959)
PubMed/Medlin: n = 708
Cochrane Library: n = 68
Embase: n = 56,523
GoogleScholar: 32,660

Duplicates removed
(n = 66,862)

Screening

Records screened after duplicates removed
(n = 23,097)

Records excluded after review
of title and abstract as not
relevant to research question
(n = 22,838)

Records screened
(n = 259)

Eligibility

Full-text articles assessed
for eligibility
(n = 6)

Included

Studies included in
qualitative synthesis
(n = 6)

Studies included in
quantitative synthesis (n =
6; meta-analysis not
conducted)