

Creating Bayesian Network Models for Breast Cancer Risk Prediction

Omolola I. Ogunyemi, PhD,^a Rowan Chlebowski, MD, PhD,^b Ellen Matloff, MS,^c Freya Schnabel, MD,^d Richard Orr, MD,^e Nananda Col, MD, MPP, MPH, FACP^a

^aDecision Systems Group, Brigham & Women's Hospital, Boston, MA; ^bHarbor-UCLA Research and Education Institute; ^cCancer Genetic Counseling, Yale Cancer Center; ^dColumbia Presbyterian Medical Center; ^eSpartanburg Regional Medical Center

BACKGROUND. Current breast cancer risk models are limited to risk factors that were known at the start of their observation period. Bayesian networks enable inclusion of newer risk factors by incorporating expert opinion and empirical evidence. Relationships between risk factors and outcomes can be learned and modeled by observing population-based data, but when data are lacking, subjective probabilities from domain experts can be used to create a Bayesian network. These probabilities can be augmented over time with observational data. We sought to develop a Bayesian network for breast cancer risk prediction that incorporates currently known risk factors for breast cancer.

METHODS. Through a literature search and expert consultation, we attempted to identify all breast cancer risk factors, model causal and associational relationships among these risk factors, and estimate conditional probabilities. We developed separate risk models according to age: 20–39 and 40–70 years. We designed Web-based questionnaires to elicit subjective probabilities from domain experts.

RESULTS. We identified 33 risk factors for women aged 20–39 and 39 for those 40–70 and logically clustered these factors to simplify probability elicitation. Clusters include factors related to genetics, family history, estrogen exposure, biopsy histology, and personal health status. We elicit conditional probabilities using questionnaires that include background information on probability elicitation methods and Bayesian networks, cognitive biases that affect risk estimation, and graphical structures of the Bayesian networks. Experts are asked to estimate the likelihood of [someone] developing cancer in the next 5 years given the presence or absence of specific risk factors by linking qualitative expressions to quantitative values. For example, one question related to family history risk is:

How likely is it that a woman will develop breast cancer within five years if she has one first degree relative under 50 with breast cancer, one second degree relative under 50 with breast cancer, one first degree relative over 50 with breast cancer, one second degree relative over 50 with breast cancer, a male family member with breast cancer and is 70 years or older?

Certain (almost)	Probable	Likely	Uncertain	Doubtful	Improbable	(Almost) impossible
100%	85%	75%	50%	25%	15%	0%

Since a large number of probabilities need to be specified for each Bayesian network, we created questionnaires with a total of almost 1,000 questions.

DISCUSSION. Bayesian networks allow for risk probabilities to be specified based on subjective assessments (“expert opinion”), empirical evidence, or a combination of both. Incorporating expert opinion is important because, often, there are insufficient data to learn and model relationships between risk factors and outcomes using population-based data (“machine learning”). When there is a paucity of data, domain expert opinion can be used to create Bayesian networks. Expert-derived probabilities can be improved over time with observational data from multiple sources, obviating the need for a single data repository that contains all known risk factors. Risk predictions for an individual draw on:

(i) relationships observed in populations studied; (ii) expert opinion; and (iii) the individual’s risk factors.

REFERENCES

Judea Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, 1988.

M. J. Druzdzel and L. C. van der Gaag. Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information, In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, p. 141–148, 1995.

L. C. van der Gaag, S. Renooij, C. L. Witteman, B. M. Aleman, and B. G. Taal. How to Elicit Many Probabilities, In: *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufman, p. 647–654, 1999.