STEM Fellowship Inter-University Big Data Challenge 2022

Bayesian Diagnostic Pipeline and Trauma Response

Utilizing Bayesian Networks to Optimize Diagnoses and Trauma Response in Canadian Healthcare Spending

Isaac W. R. Bahler[1], Samantha B. Chong[2], and Neel S. Shah[3]

[1]Dalhousie University
[2]University of Waterloo
[3]Ontario Tech University

August 17, 2022

Abstract

The publicly-funded Canadian Healthcare system is inefficient and time-intensive. By utilizing the adaptive nature of Bayesian Networks, we aim to suggest a method for improved diagnostic efficiency and trauma response. Using BayesiaLab-10, two interactive questionnaires were proposed to interface with a Bayesian Network. The questionnaires input data regarding patient ailment and output result indicating a course of treatment. Similar pipelines were created for both trauma response and general diagnosis. Due to the unique aspects of healthcare, standard analysis methods are less accurate than in other fields. However, Bayesian Networks provide a unique method of creating adaptive and grounded solutions, which aids in voiding this issue. They accomplish this by using both empirical data that has been collected as well as theoretical data. The argument presented is that by using Bayesian Networks of the proposed variety Canadian healthcare could be made more efficient, aiding in it remaining economically sustainable.

Keywords
Bayesian Networks (BN), Emergency Room (ER), Diagnoses, Cost-optimization

1 Introduction

Canada has remained committed to universally available healthcare since it's implementation in the 1960's. However, to maintain what Canadians have come to regard as a fundamental component of our national identity the system must be modernized and costs lowered to ensure Canadian healthcare has an economically sustainable future. In 2018, 23.4% of the federal government's spending was on the healthcare system, accounting for 186.5 billion dollars or 4 910 CAD per person.[1] As a portion of GDP healthcare spending has increased across all provinces though with different varying magnitudes. In addition to increased spending wait times across Canada are

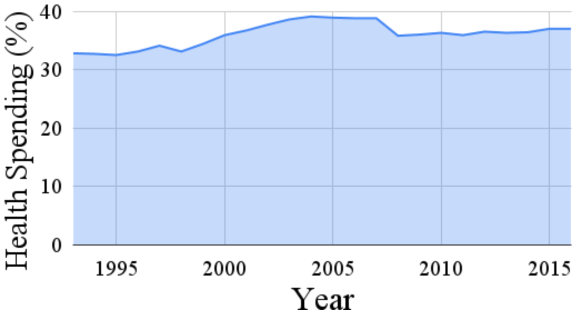

Figure 1: Health Spending as a proportion of provincial GDP vs Year [2]

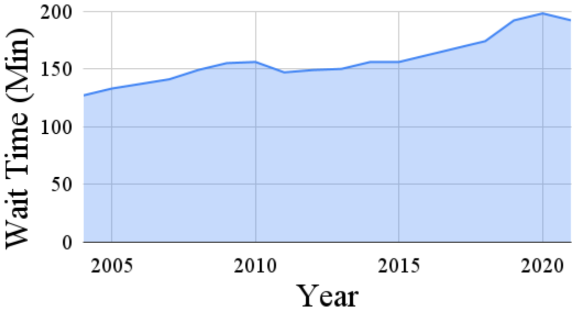

Figure 2: Average Waiting Time in Canadian Hospitals from 2004-2021 [3]

also on the rise as are unneeded emergency room visits. Tracked between 2004 and 2021 as seen in figure 2 the change becomes even more drastic.

The first point of contact within the healthcare system is primary care; these include family physicians, emergency rooms and walk-in clinics. In primary care settings, basic healthcare is provided when continuing professional care is unnecessary. It is also in these settings where it is decided if a patient needs to continue to a secondary care setting. Secondary care settings have become more specialized care to deal with less common ailments.[4] Emergency rooms are one of the most common primary care centers. The main objective of emergency rooms is to aid patients in critical condition who need rapid medical aid.[5] Patients at greater risk are given priority, while those requiring less urgent attention wait, a process called triage.

It is at these initial stages of care that

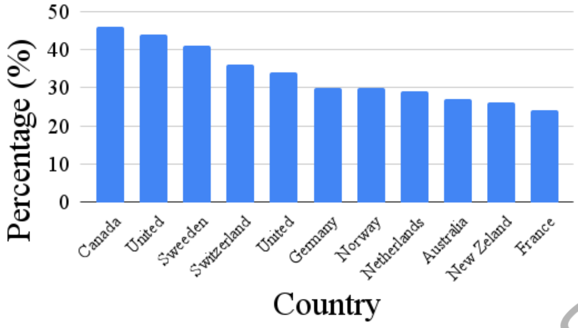

Figure 3: Percentage of Emergency Room Visits where the Patient Thought their Regular Physician Could Have Provided Care if Available [10]

we aim to decrease inefficiency resulting in improved spending and long term sustainability with the Canadian Health system.

1.1 Emergency Rooms

Canada has much higher emergency room usage compared to other countries of similar calibre and status. Many individuals who were found to visit the emergency room were there for ailments their primary care physician could easily care for.[6] In 2019, there were 18.35 million visits to Canadian emergency rooms.[7, 8] Each emergency room visit costs the hospital approximately $ 304 without including the compensation paid to physicians, totalling $ 5.6 billion on emergency room visits.[7] In 2014, the Canadian Institute for Health Information (CIHI) found that 20% of emergency room visits were unnecessary, accounting for an extra 3.67 million emergency room visits costing 1.1 billion dollars.[9]

46% of patients in Canada reported that their last emergency room visit could have been handled by their primary care physician. Comparatively the British system which is similarly funded by taxes and run by the government only has a rate of 34%, 12% less than Canada as seen in figure 3. In this group of nations with similar socioeconomic levels

Canada has the highest percentage of patients who thought they could have gone to another care provider instead of the emergency room.

1.2 Diagnosis

In order to provide a patient with a diagnosis a physician or medical personal must follow a number of steps, many of which are repetitive. A medical history must be recorded including the history of the patients family members and a physical examination conducted. Once these two steps have been completed additional tests may be ordered. These tests must then be completed and interpreted, if a diagnosis is found the patient then moves onto treatment. However if no diagnosis is found portions of the process may be repeated until a diagnosis is found or symptom's resolve themselves.[11] Often this process involves a patient going from specialist to specialist who will each complete the previously mentioned steps, something which is time inefficient.

1.3 Why Bayesian Inference

According to the Canadian Institute for Health Information (CIHI), the Canadian government's total health spending was on track to reach more than $308 billion in 2021.[12] In comparison, the Government of Canada only invests about $10 million in public health research.[13] Less than half a percent of the healthcare spending goes to scrutinizing the effectiveness of the methods used by the Canadian Healthcare system.

Humans are intuitive from an evolutionary standpoint making cause and effect decisions, something which has served us well. Unfortunately, when decision become complex and murky, such as complex multifaceted medical diagnoses, our intuition can lead us astray.[15] Bayesian inference is ideal for reasoning without neglecting the uncertainty of real-life events, eliminating human biases, and providing a mathematical formula to measure probabilistic reasoning.

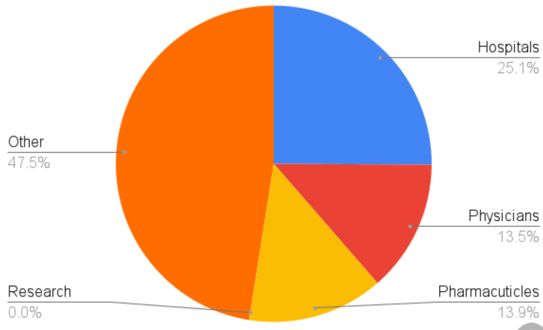

Figure 4: Pie Chart of Canadian Health Spending [13, 14]

The math for evaluating conditional probability is already known, however it remains utilized in law courts, Intelligence agencies and healthcare institutes, where the most critical decisions continue to be made by humans.[16] The lack of adoption has mainly been caused by difficulty reconciling Bayesian reasoning with traditional human instincts.[17] This leaves healthcare systems under utilized and open to improvements.

1.4 Problem Statement

Canada lacks efficiency within their primary care diagnostic process. Additionally, emergency rooms are overrun with patients who do not require their services. It is imperative that the diagnosis process is streamlined and that unnecessary patients are transitioned out of emergency rooms and into other primary care facilities.

2 Materials & Methods

2.1 Methods

A map of analytic modelling was created to summarize and depict the ideas

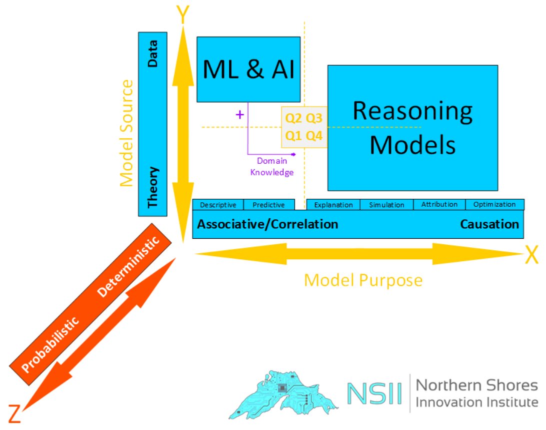

Figure 5: 3 Dimensional Domain Model of Bayesian Network

of Breiman and Shmueli.[18, 17] It contains 2-axes where the X-axis represented the model purpose, and the Y-axis represented the model source. Using this framework the diagram is depicted in figure 5.

For clarity we would like to detail the two model types we propose to address the problems outlined in the introduction. The model types withing our suggested tools are predictive and explanatory. A predictive model uses machine learning with neural networks; the explanatory model uses standard regression. The problem is that most effective models are rarely both predictive and explanatory. Bayesian Networks overcome this deficit by being constructed from both theoretical knowledge and data. Bayesian Networks remain valid even after adding causal assumptions, meaning they are much more reactive than standard models. Thus BN would be true in the entire XY plane of the depicted diagram.

2.2 Materials

Bayesian inferences can be applied in many ways; parameters like joint probability, entropy and Markov blanket can either be calculated manually or using software. We outline our proposed tool's design and how they could be used using BayesiaLab-10. BayesiaLab-10 is a desktop application that uses the Bayesian Network paradigm. A few of the software's core features are knowledge modelling, machine learning, diagnosis, prediction, simulation, adaptive questionnaires, web simulation and visualization in 2D, 3D & VR.[19]

3 Results

Purpose of the BN-Diagnostic Pipeline

The principal objective of the BN-Diagnostic pipeline is not to predict if a patient has a severe disease but to use the available resources and identify the perfunctory and cost-effective procedure to diminish any uncertain aspects regarding the diagnosis.

The aim is to use a general Bayesian Network to correctly identify which specialist, test or existing diagnosis system will be of most benefit to the patient.

3.0.1 Initial Data Processing and Collection

For such a system to work optimally patient information would be collected from specific health departments. Patient information would be turned into a dataset containing indicators of health like: history, physical examinations, laboratory data, observed symptoms and preliminary test results. The data collection process would be automated and done in real-time. The dataset could then be stored as any SQL-driven database, and the data tables can subsequently be input into the BayesiaLab software using Java Database Connectivity (JDBC) drivers.

Nodes 1				
	States	Intervals		Discretization
8- Wrist girth	<=15.4	13.19999981	15.39999962	Asked: R2-GenOpt* - 8 Obtained: R2-GenOpt* - 3
	<=17	15.39999962	17.0	
	>17	17.0	19.5	

Figure 6: Data Split into Intervals with their Threshold Values upon Discretization

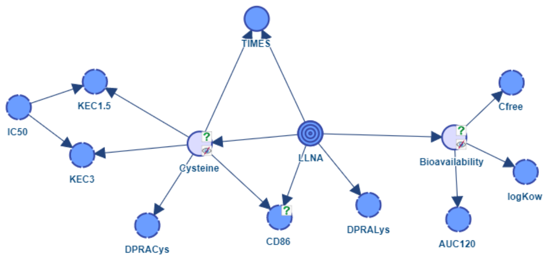

Figure 7: Visual Representation of a Bayesian Network

3.0.2 Pipeline Bayesian inferences

Once the data is imported into the BayesiaLab software, the first step is quantifying the variables into numerical or categorical. Then, an automatic data discretization would occur, using an R-Square GenOpt Algorithm. Discretization is required for two reasons. First, to decide the number of intervals for the continuous variables, and second, to determine the threshold values in order to determine the different intervals of the different variables.

When discretization is complete, all the variable classes will be defined via importing data-specific libraries. Following that a a target node that is a related feature for running the diagnostic decision analysis, is decided. As inferences are created from determining the target node, the supervised learning option will help mould a probabilistic characterization. This step provides the model with a predictive element.

Supervised learning has various options in terms of algorithms. Markov Blanket's (MB) and Augmented Markov Blanket's (ABM) would be used for this case. MB would split the model into Parent (which will transact information from incidental variables, predecessors of Target Node), Children (which will transact information from incidental variables, the successors of Target Node) and Spouses (which act as a conditional filter in transacting information between Parent and Children). ABM would highlight the dependencies between the selected nodes, increasing the model's accuracy.

$$H(X) = H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Shannon entropy

Figure 8: Shannon Entropy Formula

Upon achieving the first supervised learning model, the 'Structural Coefficient' of the model can be altered to either get a better fit or decrease the complexity of the network. Once a model is chosen a critical step will be to check 'Network Performance'. Three factors can judge the majority of Network Performance; total precision (number of correct predictions), confusion matrix (compare occurrences & reliability) and gain curves (Receiver Operating Curve – ROC curve).

BayesiaLab can calculate entropy for each variable and mutual information for numerous combinations of variables. Entropy scales the uncertainty using the Shannon Entropy Formula, as seen in figure 8, whereas mutual information is simply the entropy's combined value when multiple variables need to be examined in combination. Mutual information or conditional entropy is calculated using the conditional probability formula, figure 9.

Upon examining the information, the supervised model can be computed into an 'Adaptive Questionnaire,' taking input for the variables prioritized in the supervised learning process. Each input

$$I(X,Y) = H(X) - H(X|Y)$$

Figure 9: Formula of Conditional Probability

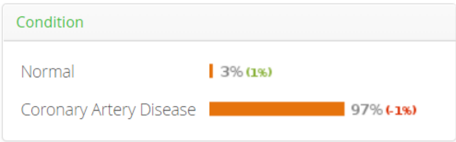

Figure 10: Final Output Based on all Inputs of Adaptive Questionnaire

will change the percentage of conditions in the Target Node. Once all questions are answered, it will show a final percentage of the target node, providing a indication of to what extent the model supports or rejects the proposal.

3.0.3 Availability and Generalization of Diagnostic Features

BayesiaLab's adaptive questionnaire feature can be exported and made accessible through being deployed on a health care providers website. It would be an interactive questionnaire that anyone could access even if the software does not exist on the device they are using.

The interactive questionnaire would only be available to general healthcare practitioners and support staff as it would allow them to confirm the result. Even though the algorithm is proven accurate in the eyes of mathematics, healthcare is often influenced by factors beyond the Diagnostic Pipeline's control. Allowing experts to supervise and support the implementation of this tool is therefore beneficial.

Figure 11: Adaptive Questionnaire Using a Mock Dataset[19]

3.1 Trauma Response

Purpose of the BN-Trauma Response

Trauma response is vital in cases of serious accident or injury. A severely injured patient has 25 percent less risk of dying if they receive trauma specific care.[20] To effectively respond to trauma, the Trauma Association of Canada has formulated Injury Severity Score (ISS) based on a tool called the Abbreviated Injury Scale (AIS).[21] To summarize, ISS uses a variety of factors and situation assessor to determine what kind of response is required for different types of Trauma.[22]

Currently trauma patients primarily arrive at the emergency room in two ways, ambulance or personal choice. Trauma incidents are usually called in summoning first responders with the needed training and equipment. The team arrives at the accident location and performs field triage. If the ISS exceeds 15 then that person is taken to the trauma center for further treatment. The field triage is based on standard parameters which include vital signs, location and severity of injury, mechanism of the accident/incident, the impact of the incident (physical as well as mental) and personal information like age of the person or disability.

Ideally a triage should be conducted by trained medical professionals. However, there are alternative ways to gather the information needed to complete the triage examination. By using Bayesian Networks we aim to perform an automatic self-triage to judge if the event requires a trauma-centered response or a non-trauma one.

3.1.1 Data Processing and Collection

The specific mechanism for collecting the patient information is beyond the scope of this paper. However, this could be done using any form of autonomous information detector which would collect information about incidents like fires, crashes or workplace injuries. Utilizing a tool of this nature would allow for first responders to enter situations more informed than without.

To set up a BN-Mediated Trauma Response Network we will use previous trauma datasets, with information like how the injury was sustained and how severe it was. Using those datasets and a mechanism similar to what we did for BN-Diagnostic Pipeline a predictive model would be created using the BayesiaLab software. Upon receiving the live data from the location of the event, the data would be treated as a test dataset and inputted into the BayesiaLab software which would use the network based on the previous datasets. As the live data automatically inputs values into the network model, the predictive component would generate an output determining whether the person requires a trauma response or not. The output would be in terms of percentage and would increase the efficiency of standard trauma response and allow for EMS to prioritize calls.

Additionally, the system could be adapted to interface with regular citizens by using the same mechanism as we did for BN-Diagnostic Pipeline. Using specific injury question formatted by medical professionals patients at home could determine whether they actually require trauma response. By using a Markov Blanket and Augmented Markov Blanket algorithms in the unsupervised learning model we could create another adaptive questionnaire. This time the questionnaire would be made open to both medical professionals as well as patients. The objective in this instance would be to reduce the number of walk-in patients coming to the emergency room who do not need it. In situations where it is difficult to determine if urgent medical aid is required; people can use that adaptive questionnaire which would assist them to decide if they should go to an ER.

4 Discussion

4.1 Impact

When considering the results section of the report, it is essential to understand the impact of the proposed BN-Pipeline and BN trauma response systems. By reducing inefficiencies in diagnostic pathways within healthcare and decreasing unnecessary ER visits, the proposed Bayesian Network solutions would reduce unnecessary Canadian healthcare spending and increase the sustainability of the our healthcare system. These methods would ensure that Canada can uphold its long-standing practice of providing publicly-funded healthcare. The findings indicate that implementing Bayesian Networks would certainly be a viable method of instrumenting the desired change. The most significant barrier to utilizing the medical tools we propose is the human adaptation in acquiring the data. Without significant change to the current healthcare system, we will be unable to utilize the adaptive solutions that Bayesian Networks provide. The dual nature solution addresses the source of excess cost and inefficiency. Given the possible economic end humanitarian benefit of utilizing large-scale research initiatives like the one proposed here, it seems the impact would be increasingly practical.

4.2 Methodology and Materials Justifications

The decision was made to utilize BayesiaLab 10 as their algorithms are already being used by organizations like NASA, the U.S. Army, Samsung, Stanford University, Intel, Microsoft. The system has garnered significant support from experts in the field. The report was limited, due to limits on procuring confidential patient data we were confined to conducting an entirely theoretical proposal and analysis. Bayesian analysis methods as a whole were specifically focused in upon for this project as their unique ability to deal with uncertainty makes them an ideal candidate for healthcare analysis.

4.3 Ethical Considerations

The ethical considerations surrounding Bayesian Network use, particularly in a medical setting, are of considerable importance when analyzing the potential impact of BN systems. While our findings indicate that BNs could be a viable method of reducing unnecessary ER visits and improving diagnostic speed and efficiency, implementing such a system is not without weakness. As we move towards a future where Neural Networks and Predictive Algorithms can outperform specialists, widespread healthcare changes must be made to take advantage of the increasing power of our tools.

To implement tools like BN systems ethically, we need to introduce informed patient consent with ethically sourced data. Without patients agreeing to share data and doctors being willing to change how they have traditionally practiced, these revolutionary tools will remain sidelined in the future. The potential upsides far outweigh the risks, providing implementation ensures ethical considerations are taken into account.

Conclusions

Throughout this report, we design and argue in favour of implementing Bayesian Networks to facilitate economically sustainable Canadian healthcare by lowering the number of unnecessary ER visits and diagnostic errors, both of which increase healthcare cost and inefficiency. By implementing a general Bayesian Network pipeline in the early stages of patient diagnosis, we argue that patients could be quickly directed to experts, Narrow Breadth Bayesian Networks, and existing computer tools. This pipeline is an automated method of decreasing the bureaucratic inefficiencies and human error that cause time and funding to be wasted on an incorrect course of treatment. The second utilization of Bayesian Networks in an online adaptive questionnaire which would allow patients considering the emergency room to determine if their visit was required. The questionnaire should reduce the massive unnecessary cost of unneeded emergency room visits in a year. When combined, the two proposed implementations of Bayesian Networks should allow for decreased wait times and less spending revitalizing, the degrading system. Due to the constraints of confidential data, this system could not be tested within the bounds of this report. Despite this however all research conducted by the members of this team indicate that implemented together with the support of physicians these systems could drastically increase efficiency ensuring that Canadian Healthcare operates sustainably well into the future. The small scale and narrow breadth tests have already been conducted and implemented, all that remains is to put the tools into the hands of all doctors and civilians as we outlined here.

4.4 Future Areas of Research

There exist many future areas of research concerning the use of Bayesian Networks in healthcare that remain unanalyzed. These topics include, but are not limited to:

- Using Bayesian Networks to predict how individual patients respond to individualized treatment options.[23]
- Using Bayesian Belief Networks to support policy decisions in novel and uncertain scenarios, like risk intensive the onset of epidemics and pandemics.[24]
- Specific research into facilitating the adoption of Bayesian Networks by overcoming barriers like data inadequacies and lack of clinical credibility.[17]

Acknowledgements

We would like to thank the Northern Shores Innovation institute for supporting our entry. Additionally, the resources and mentorship provided by STEM Fellowship and the IUBDC team were excellent and appreciated in creating this submission.

References

[1] Statistics Canada. Three-fifths of total federal, provincial, territorial and local spending went to social protection, health care and education in 2019.

[2] Canadian Institute for Health Information. National health expenditure trends, 1975 to 2018, 2018.

[3] Canadian Institute for Health Information. Nacrs emergency department visits: Volumes and median lengths of stay, 2003–2004 to 2020–2021 — supplementary statistics, 2021.

[4] Government of Canada. Canada's health care system.

[5] Anabela Mesquita Glória Campello Cláudia Dias Cristina Granja Francisco Ferreira, João Andrade. The emergency room–analysis and evaluation of an organizational model. *Rev Port Cardiol*, 27(7-8):889–900, July-August 2008.

[6] Danielle Larouche Danièle Roberge, Raynald Pineault and Léo-Rock Poirier. The continuing saga of emergency room overcrowding: Are we aiming at the right target? *Healthc Policy*, 5(3):27–39, February 2010.

[7] Canadian Institute for Health Information. Hospital spending: Focus on the emergency department, October 2020.

[8] Data Comons. Canada.

[9] Canadian Institute for Health Information. Sources of potentially avoidable emergency department visits, November 2014.

[10] Health Quality Ontaio. Under pressure : Emergency department performance in ontario.

[11] Canadian Medical Protective Association. The diagnostic process.

[12] Canadian institute for Health Information. National health expenditure trends.

[13] Government of Canada. Government of canada makes significant investment in research to transform public health for canadians.

[14] Hannah Jackson. Record spending: Canada expected to dole out more than $300b on health care in 2021. November 2021.

[15] Christian Lebiere and John R. Anderson. Cognitive constraints on decision making under uncertainty. *Front. Psychol*, November 2011.

[16] J Franklin. The science of conjecture: Evidence and probability before pascal ; with a new preface. *The Johns Hopkins University press*, 2015.

[17] Galit Shumueli. To explain or to predict? *Statistical Science*, 25(3):289–310, August 2010.

[18] Leo Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2001.

[19] Bayesialab user guide.

[20] Gregory J. Jurkovich Avery B. Nathens Katherine P. Frey Brian L. Egleston David S. Salkever Ellen J. MacKenzie, Frederick P. Rivara and Daniel O. Scharfstein. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*, 354:366–378, January 2006.

[21] Abbreviated injury scale (ais).

[22] Akkamahadevi Patil Madhu Srinivasarangan Suraj Rameshbabu Homanna Javali, Krishnamoorthy and Sriharsha. Comparison of injury severity score, new injury severity score, revised trauma score and trauma and injury severity score for mortality prediction in elderly trauma patients. *Indian J Crit Care Med*, 23:73–77, Febuary 2019.

[23] Coley Y et a Zeger SL, Wu Z. Using a bayesian approach to predict patients' health and response to treatment, 2020.

[24] Dragan Savic David Butler Raziyeh Farmani, Hans Jørgen Henriksen. An evolutionary bayesian belief network methodology for participatory decision making under uncertainty: An application to groundwater management. *Integrated Environmental Assessment and Management*, 8, March 2011.