

# Bayesian Networks in Medicine: Presenting Query Response Uncertainty for Decision Support

Janneke H. Bolt<sup>1,2,\*</sup>, Anna Berghuis<sup>3</sup>, Arjen Hommersom<sup>2</sup>, Marike Lombaers<sup>3</sup>,  
Johanna M.A. Pijnenborg<sup>3</sup> and Silja Renooij<sup>1</sup>

<sup>1</sup>*Utrecht University, Department of Information and Computing Sciences, Utrecht, The Netherlands*

<sup>2</sup>*Open University of the Netherlands, Faculty of Science, Heerlen, The Netherlands*

<sup>3</sup>*Radboud University Medical Center, Department of Obstetrics and Gynaecology, Nijmegen, The Netherlands*

## Abstract

Despite their good characteristics, tools based on Bayesian networks are not yet widely used in medical decision support. Meeting the needs of the intended users is crucial for the acceptance of these tools, and clinical involvement in their development is thus required to promote their use. During the development of a Bayesian network-based tool for the prediction of lymph node metastases in patients with endometrial cancer, as one of the users needs a measure of query response uncertainty was put forward. In this paper, we sustain meeting this need by exploring options for the presentation of query response uncertainty. We consider the level of detail, one-sided versus two-sided intervals and two different 'look-ahead' options. The different options are illustrated through a small example network.

## Keywords

Bayesian networks, Decision support, Query response uncertainty, Explainable AI

## 1. Introduction

In healthcare, many decisions with often far-reaching consequences have to be made. It is therefore not surprising that, already from the early development of computer-based tools for decision making, such systems were built for the medical domain (e.g. [1, 2]). Among the computer-based tools for decision making are the well-known Bayesian networks [3]. Bayesian networks have great advantages for clinical decision making such as a quite intuitive visualization of the problem domain by means of their graph structure, and their ability to provide predictions for incomplete and/or conflicting patient findings [4]. Despite their suitable characteristics, Bayesian networks are not yet widely used in clinical practice. Kyrimi et al. [5] performed a comprehensive literature review to analyze this low adoption and provided an overview of “benefits, barriers and facilitating factors” for the use of Bayesian networks in healthcare. As one of the barriers, clinicians’ resistance is mentioned, with clinical involvement as one of the facilitating factors. This paper presents research related to employing this facilitating factor. More specifically, in a joint team of AI researchers and clinicians, we investigate the need for explanation types and forms in the context of a Bayesian network tasked with the challenging preoperative prediction of lymph node metastasis (LNM) for patients with endometrial cancer.

Given the diagnosis of endometrial cancer, the preoperative prediction of the presence of LNM has a great impact on the proposed treatment strategy. However, although there are several biomarkers related to LNM, none of them are currently used in daily practice; the tumor grade is used as most important predictor. This clinical problem resulted in an initiative within the European Network for Individualized Treatment of Endometrial Cancer (ENITEC) for the development of a Bayesian network for the prediction of LNM [6]. This network, ENDORISK, includes nine preoperative variables and has as main outcome measures LNM and 5-year disease-specific survival.

---

*HC@AIxIA 2024: 3rd AIxIA Workshop on Artificial Intelligence For Healthcare*

\*Corresponding author.

✉ j.h.bolt@uu.nl (J. H. Bolt); anna.berghuis@radboudumc.nl (A. Berghuis); arjen.hommersom@ou.nl (A. Hommersom); marike.lombaers@radboudumc.nl (M. Lombaers); hanny.ma.pijnenborg@radboudumc.nl (J. M.A. Pijnenborg); s.renooij@uu.nl (S. Renooij)



© 2024 Copyright 2024 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

One of the needs put forward by the intended clinical users of ENDORISK is that the system provides information on the uncertainty in query responses. In particular, the network can be used to compute the probability of LNM for a specific patient, but there is no indication of the reliability of this probability.

This wish for information on the uncertainty of query responses relates to the concept causability, a notion used within Explainable AI (XAI), which encompasses measurements for the quality of explanations [7]. In this context it is particularly important that a system provides an indication of the uncertainty in its predictions [8]. Causability relates to a user-oriented perspective on explainability of AI models, which is in contrast with the majority of XAI methods that focus on providing technical insights into the inner workings of these models [9].

In this paper we propose and discuss different options for the presentation and explanation of query response uncertainty, where we base the latter on an existing method for approximating the variance of a Bayesian network query [10].

## 2. Preliminaries

### 2.1. Bayesian networks and naive Bayesian networks

A Bayesian network is a model of a joint probability distribution  $\Pr$  over a set of stochastic variables  $\mathbf{V}$ , consisting of a directed acyclic graph and a set of conditional probability distributions. In this paper we presume that all  $A \in \mathbf{V}$  are discrete. Variables will be denoted by upper-case letters ( $A$ ), and their values by lower-case letters ( $a_i$ ); sets of variables by bold-face upper-case letters ( $\mathbf{A}$ ) and their instantiations by bold-face lower-case letters ( $\mathbf{a}$ ). Given a binary variable  $A$ ,  $A = a_1$  is often written as  $a$  and  $A = a_2$  is often written as  $\bar{a}$ . Each variable  $A \in \mathbf{V}$  is represented by a node  $A$  in the network's graph<sup>1</sup>. (Conditional) independence between the variables is captured by the graph's set of arcs according to the d-separation criterion [3]. The strength of the probabilistic relationships between the variables is captured by the set of conditional probability distributions  $\Pr(A | \mathbf{p}(A))$ , where  $\mathbf{p}(A)$  denotes a combination of values, or instantiation, of the parents of  $A$ . The probabilities that define these distributions are called the network *parameters*. The joint probability distribution is then defined by:

$$\Pr(\mathbf{V}) = \prod_{A \in \mathbf{V}} \Pr(A | \mathbf{p}(A))$$

Using the formula above, a Bayesian network can yield any probability of the modeled problem domain.

A naive Bayesian network is a Bayesian network with a specific graph structure. The graph consists of just a single target node with a set of observable feature child nodes and captures the assumption that the observable features are all mutually independent given the target node. For a naive network with a binary valued target node  $H$ , and feature nodes  $\mathbf{F}$ , the probability of  $\Pr(h | \mathbf{e})$ , with  $\mathbf{e} = \{e_1, \dots, e_n\}$  an instantiation of  $\mathbf{E} \subseteq \mathbf{F}$  is computed from the network parameters as follows

$$\Pr(h | \mathbf{e}) = \frac{\prod_{i \in \{1, \dots, n\}} \Pr(e_i | h) \cdot \Pr(h)}{\sum_{h_j \in \{h, \bar{h}\}} \prod_{i \in \{1, \dots, n\}} \Pr(e_i | h_j) \cdot \Pr(h_j)} \quad (1)$$

### 2.2. Query response uncertainty

The standard query response from a Bayesian network is a single outcome probability, computed from the network's parameters. Since in practice the parameters of a network are estimates, based on population samples and/or uncertain expert opinions, the query response, even using an exact inference algorithm, will be an estimate as well. Van Allen et al. [10, Theorem 2] provide a formula to approximate the variance of the query responses of a Bayesian network, resulting from the uncertainty in the parameter probabilities. Below we give the formula we used to approximate the variance of

<sup>1</sup>The terms node and variable will be used interchangeably.

the outcome  $\Pr(h | \mathbf{e})$ , of a naive Bayesian network with target variable  $H$  and  $n$  observed features  $\mathbf{e} = \{e_1, \dots, e_n\}$ , in which the parameters are estimated from data, based on [10, Theorem 2]:

$$\tilde{\sigma}^2(\Pr(h|\mathbf{e})) = \sum_{\text{par} \in \text{PAR}} \frac{(w_{\text{par}}^2 - p_{\text{par}}) - (w_{\text{par}} - p_{\text{par}})^2}{m_{\text{par}}}$$

where

- $\text{PAR} = \{\Pr(h), \Pr(e_1|h), \dots, \Pr(e_n|h), \Pr(e_1|\bar{h}), \dots, \Pr(e_n|\bar{h})\}$ ;
- $w_{\text{par}}$  is the partial derivative of Equation (1) with respect to  $\text{par}$ ;
- $p_{\text{par}}$  is the estimate of  $\text{par}$  based on relative frequency in the data, and
- $m_{\text{par}}$  is the number of cases involved in computing  $p_{\text{par}}$  plus 1.

To represent the higher-order uncertainty in the outcome probability, we define a distribution over  $\Pr(h|\mathbf{e})$  with variance  $\sigma^2 = \tilde{\sigma}^2(\Pr(h|\mathbf{e}))$  and mean  $\mu = \Pr(h|\mathbf{e})$ . Following Van Allen et al. [10] we use a beta distribution, which has the advantages of being confined to  $[0, 1]$  and being able to model skewed distributions. The parameters  $\alpha$  and  $\beta$  that control the shape of the beta distribution are related to its variance  $\sigma^2$  and mean  $\mu$  as follows [10, Equation 4]:

$$\alpha = \mu \cdot \left[ \frac{\mu \cdot (1 - \mu) - \sigma^2}{\sigma^2} \right]$$

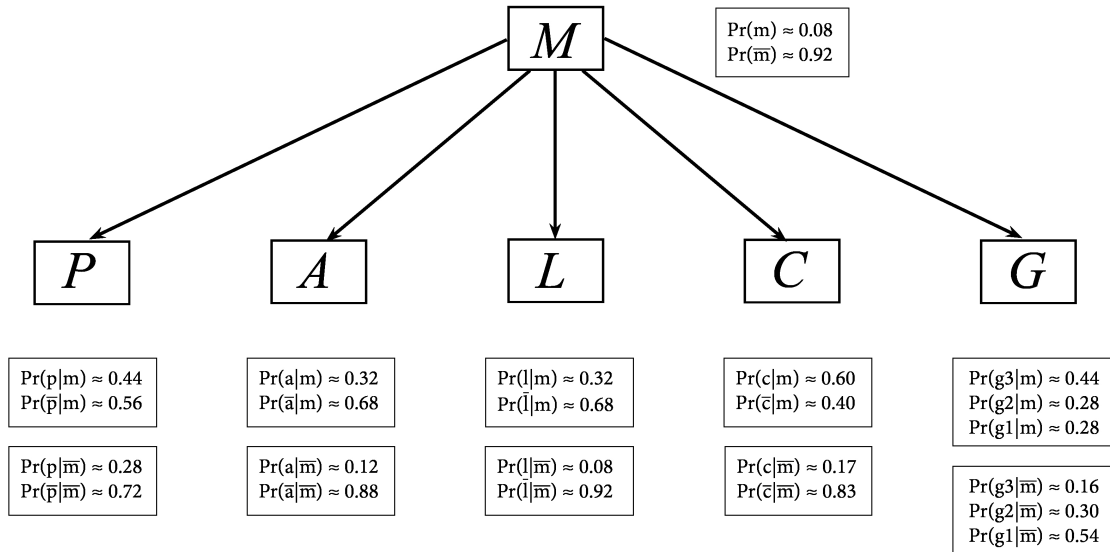
$$\beta = (1 - \mu) \cdot \left[ \frac{\mu \cdot (1 - \mu) - \sigma^2}{\sigma^2} \right]$$

This beta distribution in turn can be used to compute Bayesian credible intervals for  $\Pr(h|\mathbf{e})$ , that is, an interval that contains  $\Pr(h|\mathbf{e})$  with a given confidence. In this paper we consider the 95% credible interval. In computing such interval, the lower and upper end of the interval have to be chosen such that the area under the curve of the probability density function (PDF) of the distribution within this interval is 0.95. This can be done in more ways than one. In this paper we consider two-sided intervals with 2.5% of the area below its lower end and 2.5% above its upper end, and one-sided intervals with 5% of the area above its upper end. We take a one-sided interval with an upper end, since it is important to know whether the risk of LNM is with a given confidence below some threshold value. Note that for the one-sided interval a lower end of 0 can be used since the area under the curve below 0 will always be 0. Note moreover that, for the same distribution, the upper end of the one-sided 95% interval will be lower than the upper end of the two-sided 95% interval.

### 3. Example network

The ENDORISK network is a detailed Bayesian network in which nine clinical and molecular biomarkers are integrated to predict lymph node metastasis (LNM) and 5-year disease-specific survival [6]. To simplify both computations and illustrations, we simplified this network to the naive Bayesian network depicted in Figure 1, with one target variable,  $M$ , with the values  $m$  for LNM is present (confirmed by lymphadenectomy) and  $\bar{m}$  for LNM is absent (confirmed by lymphadenectomy or by the absence of recurrence without therapy), and five diagnostic variables. Further details of the network are given in Figure 1. For the estimation of the parameters of the naive network we used data provided by Radboud University Medical Center, collected within the ENITEC. From the 952 provided cases we only used the 328 cases with complete observations for the variables in the example network. For clarity, in Figure 1 just the mean values of the parameters are given, and not the underlying numbers of cases. For the discretization of the continuous variables we used the same threshold values as were used in the ENDORISK network except for the variable  $P$ . Our adapted threshold for  $P$  made this variable more informative than the threshold used in the ENDORISK network. The values of the binary variables are chosen such that the observations  $p, a, l$  and  $c$  are indications that LNM will be present and the observations  $\bar{p}, \bar{a}, \bar{l}$  and  $\bar{c}$  are indications that LNM will be absent. In our illustrations we will indicate

M: Positive nodes including follow-up m is yes,  $\bar{m}$  is no.  
P: PR expression percentage p is < 50%,  $\bar{p}$  is  $\geq$  50% of tumor cells with nuclear staining  
A: p53 expression preop a is aberrant,  $\bar{a}$  is wild type  
L: L1CAM\_expression\_preop l is  $\geq$ 10%,  $\bar{l}$  is < 10% of tumor cells with membraneous staining  
C: CA125\_PREOP\_bi c is  $\geq$  35 iu/ml,  $\bar{c}$  is < 35 iu/ml  
G: Grade preop g3 is grade 3, g2 is grade 2, g1 is grade 1



**Figure 1:** Example naive Bayesian network for the prediction of lymph node metastasis in patients with endometrial cancer.

for the patient favorable observations with the color green and unfavorable observations with the color red.

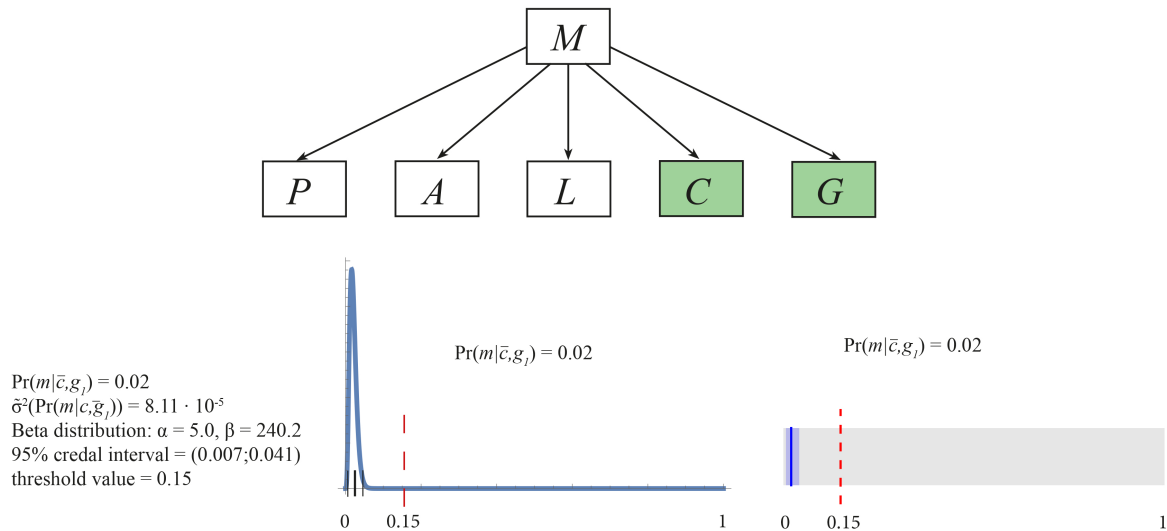
We note that the simplified network is merely used to compute and illustrate the different options for the presentation of query response uncertainty. The actual output probabilities may be different from those computed from the original network; however, here we are not interested in the performance of the network, but in options for presenting uncertainty. For the purpose of illustrating our general approach, the simplified version of the network suffices just as well.

## 4. Options for presenting query response uncertainty

The regular outcome of a standard query for our example network from Figure 1 is the probability of LNM given the observations for a patient entered into the network. Recall that our physicians expressed the wish for a measure on the confidence in this outcome. To provide for such a measure, we implemented the method to compute the variance on the outcome, finding a suitable beta distribution over the outcome and computing a Bayesian credible interval over this outcome, as described in Section 2. However, given the extra information on query response uncertainty, there are several options regarding *which* information is presented to the users and also *how* this information is presented. The choices made may have a considerable impact on the usability of the system and moreover can influence the interpretation of the results. In the following sections we explore several options: varied levels of detail, one-sided intervals versus two-sided intervals, and two different 'look-ahead' options.

### 4.1. Level of detail

Suppose that the observations  $\bar{c}$  and  $g_1$  are entered. We find that the probability  $\Pr(m|\bar{c}, g_1) = 0.02$  with an approximated variance of  $8.11 \cdot 10^{-5}$  and a beta distribution over  $\Pr(m|\bar{c}, g_1)$  with  $\alpha = 5.0$  and  $\beta = 240.2$ . From this beta distribution in turn the two-sided 95% credible interval (0.007; 0.041) is computed. Figure 2 shows, with a decreasing level of detail, three different ways to report over these



**Figure 2:** Three options for reporting on the uncertainty in the outcome  $\Pr(m|\bar{c}, g_1) = 0.02$  with, from left to right, a decreasing level of detail.

data. The first option just states all outcomes numerically. The other two options are more visual. The second provides the mean outcome value, gives the graph of the beta distribution, and indicates the mean and the 95% credible interval with vertical bars on the x-axis. The third option provides the mean outcome value and visualizes both the mean and the 95% credible interval in blue in the gray bar that represents the entire 0-1 interval. In all options, we also provide an illustrative threshold value of 0.15 that indicates the posterior outcome probability above which the patient will be considered for a certain treatment trajectory. In the view of our team, the clarity of the output gains with the decrease of detail.

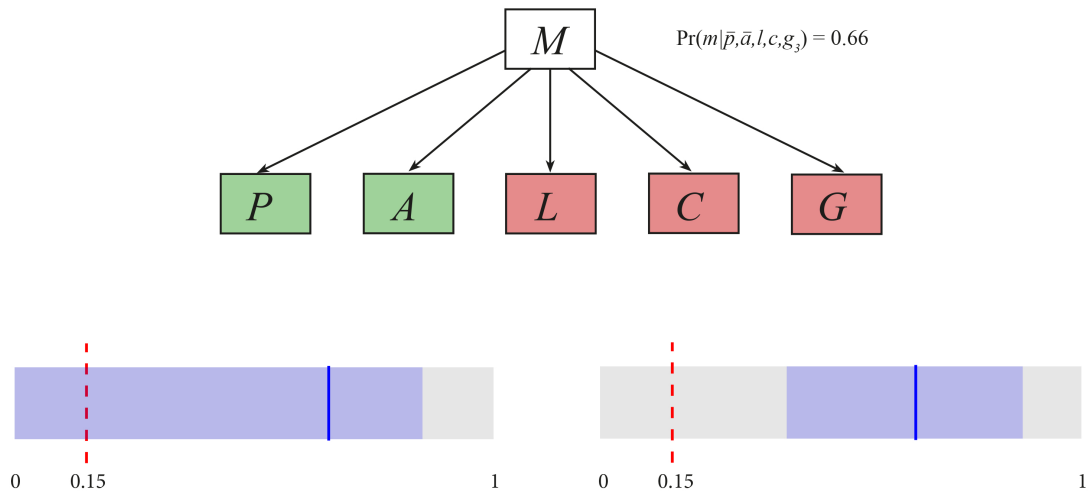
#### 4.2. One-sided or two-sided intervals

In providing credible intervals for query responses, as described in Section 2.2, one has a choice in the upper and lower end of the credible interval. In the previous section a two-sided 95% credible interval (0.007; 0.041) for  $\Pr(m|\bar{c}, g_1) = 0.02$  is given, however, also the one-sided 95% interval (0; 0.037) could have been reported. In the choice for a treatment strategy, the upper end of the credible interval is the one that will influence the clinical decision most, since the treatment will be different in case the risk of LNM is below some threshold value; in that case, the risk of LNM might be regarded as 'low enough'. This argument supports the choice of a one-sided interval. However, with a one-sided interval the level of uncertainty of the outcome is somewhat obscured, while these intervals were added to provide information on this level of uncertainty in the first place. Such obscuration is illustrated in Figure 3, which displays the one-sided and the two-sided 95% credible intervals for LNM present given the observations  $\bar{p}, \bar{a}, l, c$  and  $g_3$ . We find that  $\Pr(m|\bar{p}, \bar{a}, l, c, g_3) = 0.66$ , with one-sided interval (0; 0.852) and two-sided interval (0.388; 0.878). The much larger one-sided interval suggests more uncertainty which is not correct since the two 95% credible intervals are based on the same data.

#### 4.3. Look-ahead: one step look-ahead or full scenarios

From the discussions in our team we noted that the intervals reported in the query response might be confusing. The upper end of the credible interval can be interpreted as the maximal shift in outcome that can still occur when more diagnostic information becomes available, which is incorrect. As described in Section 2.2, this upper end marks the end of the interval that contains the probability of LNM with a given confidence given the *current* diagnostic information.

To avoid confusion, we consider two options that provide insight into the shift in outcome that may occur given more diagnostics: the 'one step look-ahead' option and the 'full scenarios' option. In the



**Figure 3:** A visualization of the one-sided (left) and two-sided (right) credible intervals of the outcome  $\Pr(m|\bar{p}, \bar{a}, l, c, g_3)$ .

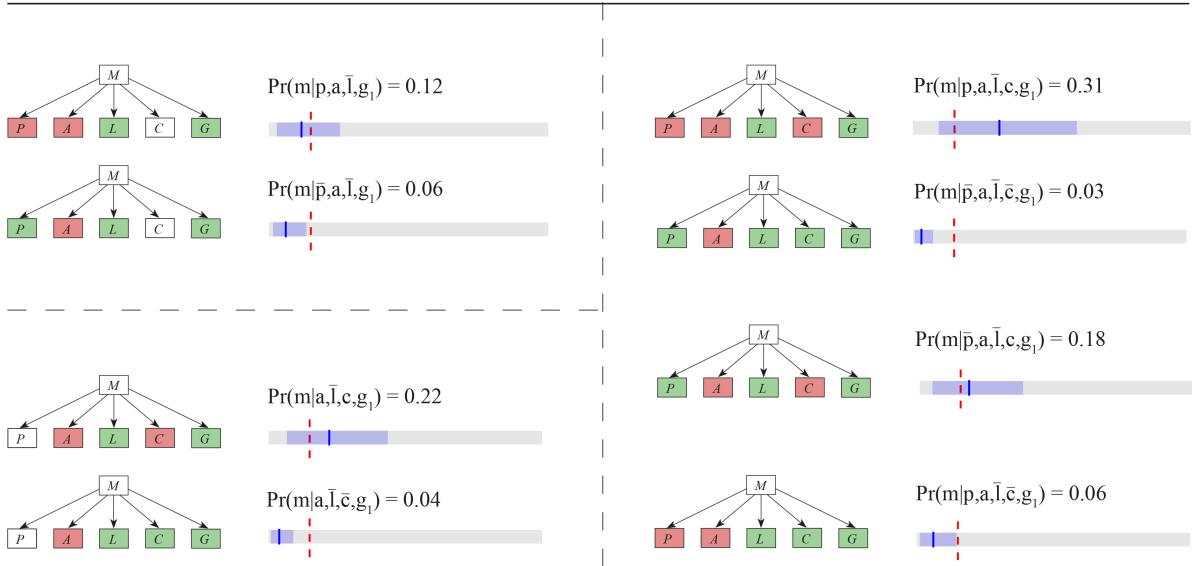
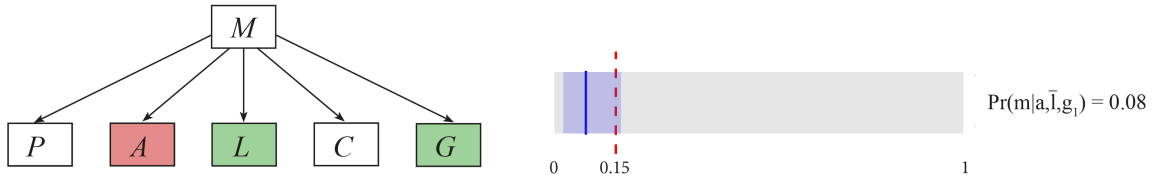
one step look-ahead option, for all unobserved diagnostic variables separately, the system consecutively fills in all possible outcomes and subsequently computes the probability of LNM given the observations already available plus this additional information. In the full scenario option, the system does the same, but then for all unobserved variables combined. All this is illustrated in Figure 4 for an example case with observations  $a, \bar{l}$ , and  $g_1$  and variables  $P$  and  $C$  not yet observed. We find that  $\Pr(m|a\bar{l}g_1) = 0.08$ , with a 95%-credible interval of which the upper end is just above the threshold of 0.15. The left side illustrates the one step look-ahead option, by showing the effect on the probability of LNM of observing a value for either  $P$  or  $C$ :  $p, \bar{p}, c$  and  $\bar{c}$  are shown from top to bottom. The right side illustrates the full scenarios option by showing the effect on LNM of observing all combinations for  $P$  and  $C$ :  $pc, \bar{p}\bar{c}, \bar{p}c$  and  $p\bar{c}$  are shown from top to bottom.

We observe that the one step look-ahead option clearly shows that the observation of  $C$  gives a greater shift in the probability of LNM than the observation of  $P$ . This information might be used for subsequent test selection, which is an advantage of this option. Moreover, the number of possibilities grows linearly with the number of variables not yet observed, while in the full scenarios option this number grows exponentially. This also is in favor of the one step look-ahead option. The full scenarios option, on the other hand, shows the worst and the best case, which information may also be valuable.

Both options give the insight that a physician might need more information in this case since the upper end can still reach values distinctly above and distinctly below the threshold.

## 5. Discussion, conclusions and future research

In this paper we explored different options for presenting query response uncertainty in Bayesian networks to end users. We considered a varied level of detail, one-sided versus two-sided credible intervals and two different 'look-ahead' options. In future research we want to elaborate on these options by a further involvement of more intended users. We want to investigate which level of detail is preferred in what circumstances, or by which type of user. In addition, we will study how users actually interpret the credible intervals and whether they indeed prefer two-sided intervals over one-sided intervals. Also, we will further investigate whether the look-ahead options are indeed useful for test-selection. Note that the full scenarios option might yield to many scenarios to present. In that case it might be an alternative to present just a subset of these scenarios. Which subset is most useful for the user of the system is subject for further research. Moreover we note that the full scenario option is related to the concept of same-decision probability (SDP), which is, as stated by the authors: "the probability that we would have made the same threshold-based decision, had we known the state of some hidden variables pertaining to our decision" [11]. The SDP can be viewed as a kind of summary



**Figure 4:** An illustration of the two options for a look-ahead: the one step look-ahead option (left) and the full scenario option (right).

of the full scenario option, adding up the probabilities of outcomes leading to the same decision. We will investigate how to incorporate such additional information. Providing such information may also be an alternative in case the full scenarios option results in too many scenarios. Another interesting question, regarding all options, is what the actual impact is of the use of intervals in the query response on the decisions made. From a more technical perspective we would like to investigate whether there are situations in which the approximation method yields results that deviate too much from the true values. Moreover we are interested in investigating the best ways to include expert opinions.

## 6. Acknowledgements

This publication is part of the project PersOn with file number P21-03 of the research programme Perspective which is (partly) financed by the Dutch Research Council (NWO).

The data used were collected within the European Network of Individual Treatment in Endometrial Cancer (ENITEC) and were provided by Radboud University Medical Center.

## References

- [1] E. H. Shortliffe, B. G. Buchanan, A model of inexact reasoning in medicine, *Mathematical Biosciences* 23 (1975) 351–379.
- [2] E. H. Shortliffe, *Medical informatics and clinical decision making: the science and the pragmatics*, *Medical Decision Making* 11 (1991) S2–S14.
- [3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.*, Morgan Kaufmann, 1988.
- [4] P. J. Lucas, L. C. van der Gaag, A. Abu-Hanna, *Bayesian networks in biomedicine and health-*

- care, *Artificial Intelligence in Medicine* 30 (2004) 201–214. doi:<https://doi.org/10.1016/j.artmed.2003.11.001>.
- [5] E. Kyrimi, K. Dube, N. Fenton, A. Fahmi, M. R. Neves, W. Marsh, S. McLachlan, Bayesian networks in healthcare: What is preventing their adoption?, *Artificial Intelligence in Medicine* 116 (2021) 102079. doi:<https://doi.org/10.1016/j.artmed.2021.102079>.
- [6] C. Reijnen, E. Gogou, N. C. M. Visser, H. Engerud, J. Ramjith, L. J. M. van der Putten, K. van de Vijver, M. Santacana, P. Bronsert, J. Bulten, M. Hirschfeld, E. Colas, A. Gil-Moreno, A. Reques, G. Mancebo, C. Krakstad, J. Trovik, I. S. Haldorsen, J. Huvila, M. Koskas, V. Weinberger, M. Bednarikova, J. Hausnerova, A. A. M. van der Wurff, X. Matias-Guiu, F. Amant, ENITEC Consortium, L. F. A. G. Massuger, M. P. L. M. Snijders, H. V. N. Küsters-Vandeveld, P. J. F. Lucas, J. M. A. Pijnenborg, Preoperative risk stratification in endometrial cancer (ENDORISK) by a Bayesian network model: A development and validation study, *PLoS Medicine* 17 (2020) e1003111. doi:[10.1371/journal.pmed.1003111](https://doi.org/10.1371/journal.pmed.1003111).
- [7] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller, Causability and explainability of artificial intelligence in medicine, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9 (2019) e1312.
- [8] C. Combi, B. Amico, R. Bellazzi, A. Holzinger, J. H. Moore, M. Zitnik, J. H. Holmes, A manifesto on explainability for artificial intelligence in medicine, *Artificial Intelligence in Medicine* 133 (2022) 102423.
- [9] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, R. Ranjan, Explainable AI (XAI): Core ideas, techniques, and solutions, *ACM Computing Surveys* 55 (2023). doi:[10.1145/3561048](https://doi.org/10.1145/3561048).
- [10] T. Van Allen, A. Singh, R. Greiner, P. Hooper, Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference, *Artificial Intelligence* 172 (2008) 483–513. doi:<https://doi.org/10.1016/j.artint.2007.09.004>.
- [11] A. Choi, Y. Xue, A. Darwiche, Same-decision probability: A confidence measure for threshold-based decisions, *International Journal of Approximate Reasoning* 53 (2012) 1415–1428. doi:<https://doi.org/10.1016/j.ijar.2012.04.005>.