Journal of Biomedical Informatics 43 (2010) 485–495

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Comparing risks of alternative medical diagnosis using Bayesian arguments

Norman Fenton *, Martin Neil

Queen Mary University of London, RADAR (Risk Assessment and Decision Analysis Research), School of Electronic Engineering and Computer Science, London E1 4NS, UK
Agena Ltd., 32-33 Hatton Garden, London EC1N 8DL, UK[1]

ARTICLE INFO

Article history:
Received 16 December 2008
Available online 10 February 2010

Keywords:
Bayes Theorem
Bayesian networks
Event trees
Catheter angiagram
MRA scan
Aneurysm
Palsey

ABSTRACT

This paper explains the role of Bayes Theorem and Bayesian networks arising in a medical negligence case brought by a patient who suffered a stroke as a result of an invasive diagnostic test. The claim of negligence was based on the premise that an alternative (non-invasive) test should have been used because it carried a lower risk. The case raises a number of general and widely applicable concerns about the decision-making process within the medical profession, including the ethics of informed consent, patient care liabilities when errors are made, and the research problem of focusing on 'true positives' while ignoring 'false positives'. An immediate concern is how best to present Bayesian arguments in such a way that they can be understood by people who would normally balk at mathematical equations. We feel it is possible to present purely visual representations of a non-trivial Bayesian argument in such a way that no mathematical knowledge or understanding is needed. The approach supports a wide range of alternative scenarios, makes all assumptions easily understandable and offers significant potential benefits to many areas of medical decision-making.

1. Introduction

In a classic and much referenced study [1] the following question was put to 60 students and staff at Harvard Medical School (we shall refer to this as the 'Harvard question').

> "One in a thousand people has a prevalence for a particular heart disease. There is a test to detect this disease. The test is 100% accurate for people who have the disease and is 95% accurate for those who don't (this means that 5% of people who do not have the disease will be wrongly diagnosed as having it).
> If a randomly selected person tests positive what is the probability that the person actually has the disease?"

Almost half gave the response 95%.

The 'average' answer was 56%.

In fact the correct answer is just below 2% (as we shall explain in Section 2). There is much debate about why intelligent people are so poor at answering questions that require simple mathematical reasoning. There is also much debate about the best ways to avoid such errors. For example, Cosmides and Tooby [11] demonstrated that responses to the Harvard question were significantly improved by using language that avoided abstract probabilities. This led them to challenge the widely believed claims [29] that lay people were inherently incapable of accurate probabilistic reasoning. Their view was that it was the Bayesian framework generally used to answer such questions that was the cause of confusion.

It turns out that the issue of how best to present probabilistic Bayesian reasoning was crucial in a recent medical negligence case that we describe in Section 3. We worked as experts (on probabilistic risk assessment) to help the claimant's legal team represent and verify the key probabilistic arguments, and even more importantly, to understand them sufficiently well to be able to present them in court. The High Court accepted the claimant's case and awarded significant damages.

The approach we used in the case to *explain* the argument (a decision tree representation of Bayes) is described in detail in Section 4. With this approach no mathematics is required, and it is still easy to consider a full range of assumptions that incorporate both the claimant and the defence viewpoints.

The case raises a number of very general and widely applicable concerns about the decision-making process within the medical profession and also about how the Bayesian approach can improve this process. Our view is that, in many cases such as this particular one, it is possible to present purely visual representations of a Bayesian argument, in such a way that no mathematical knowledge or understanding is needed. However, where the medical problem involves many variables and interactions, the proposed approach becomes infeasible and an alternative approach (Bayesian networks) is needed. Section 5 explains why we believe Bayesian networks can provide a viable alternative, and also explains how we used them to fully verify the whole argument in the case.

* Corresponding author. Address: School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK. Fax: +44 020 8980 6533.
E-mail address: norman@dcs.qmul.ac.uk (N. Fenton).

[1] www.agena.co.uk.

1532-0464/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2010.02.004

As we explain in the paper, there is nothing new about using either Bayesian reasoning, decision tree representations, or Bayesian networks within the domain of risk assessment in medical diagnosis. Our novel contribution in this paper is both a real (and successful) case study in using all of these techniques and a new approach to how these techniques can be made much more widely accepted.

2. Presenting Bayesian arguments visually

The easiest way to explain the correct result for the Harvard question is to use the kind of visual argument presented in Figs. 1 and 2 (similar approaches have been proposed in [4]). Here, we imagine a large number of randomly selected people (1000 in this case but the argument works for any large number) being tested for the particular disease.

The argument is essentially as follows:

- The disease is prevalent in one in a thousand people, so in a sample of, say, 1000 people we would expect about one to have the disease (in 100,000 people we would expect about 100 to have the disease etc.). This is represented by the black figure in Fig. 1.
- But if you test everybody in the sample then, in addition to the people who do have the disease, we would expect approximately 50 – that is 5% of the other 999 – will be wrongly diagnosed as having it (in 100,000 people this will be approximately 4995, that is 5% of the 99,900 who do not have the disease). This is represented by the grey figures in Fig. 1.
- In other words fewer than 2% of the people who are diagnosed positive (i.e. 1 out of the 51 in the case of 1000 people and 100 out of 5005 in the case of 100,000 people) actually have the disease. This is represented in Fig. 2.

When people give a high answer, like 95%, they are falling victim to a very common fallacy known as the 'base-rate neglect' fallacy [29]; people neglect to take into consideration the very low probability (of having the disease) that forms the vital starting point. In comparison, the probability of a false positive test is relatively high (5% is the same as 50 in a thousand, whereas there is only a one in a thousand chance of having the disease).

The above visual explanation can convince even the most sceptical and mathematical-illiterate observers of both the correct answer and how to calculate it themselves. However, this cannot be said of the standard formal approach to solving such problems. Specifically, the formal way to present the above argument is to use Bayes Theorem, which is generally acknowledged as the standard approach for reasoning under uncertainty. The Bayesian approach enables us to re-evaluate probabilities (in this case the probability that a randomly tested patient has the disease) in the light of new evidence (in this case a positive test result). Specifically, Bayes Theorem is a formula for the revised (posterior) probability in terms of the original (prior) probability and the probability of observing the evidence. Its use in medical diagnostics is far from new as can be seen from publications dating back almost 50 years [35,57]. However, although the formula is straightforward (see the Appendix for both the formula and its application in this case) most people without a statistical/mathematical background do not understand the argument if it is presented in this way [11]. There are two reasons for this:

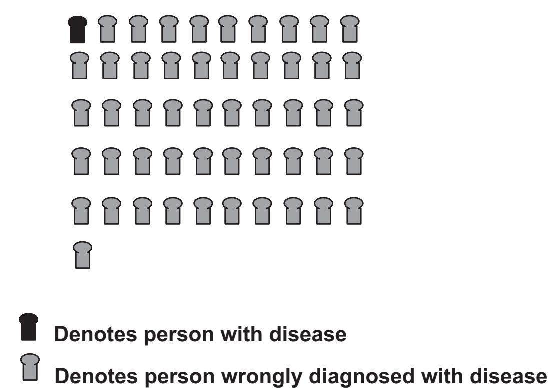

Fig. 2. Only 1 out of the 51 diagnosed with the disease actually have the disease.

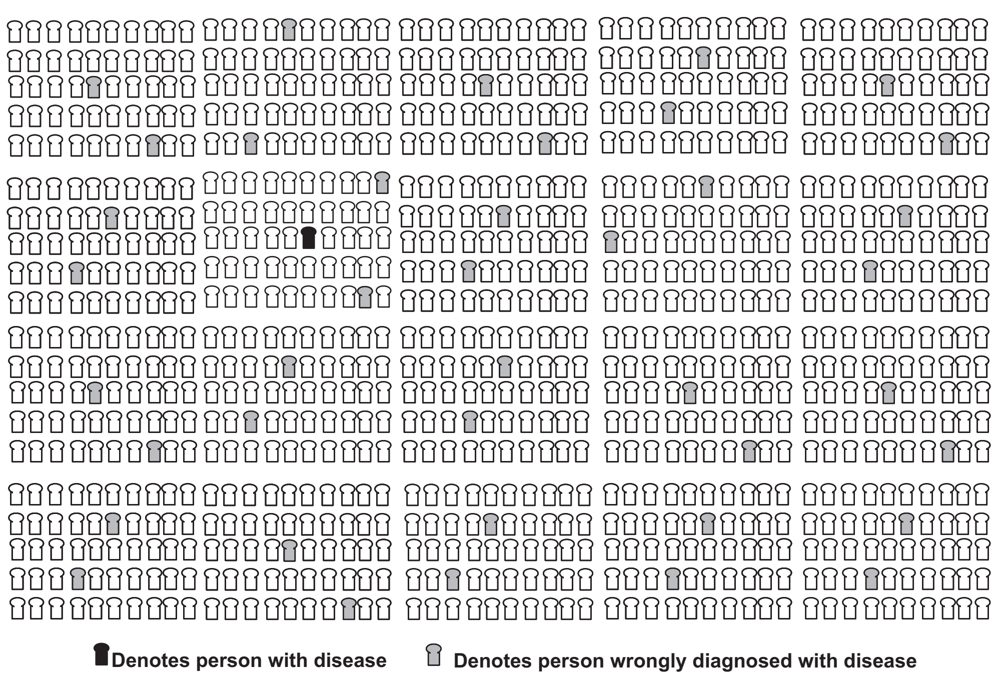

Fig. 1. In 1000 random people about 1 has the disease but about 50 more are diagnosed as having the disease.

1. They simply refuse to either 'examine' or 'believe' a mathematical formula.
2. The results are presented as probabilities rather than frequencies and they find this much harder to contextualise.

Hence, while mathematicians and statisticians assume that Bayesian arguments are simple and self-explanatory, they are not normally understood by doctors [9,17,29,51], lawyers, judges and juries [16,14,21,30]. Because of this the power of Bayesian reasoning has been massively underused.

An approach to presenting the argument that can be viewed as a semi-formal (and more easily repeatable) version of the above visual argument, is to use decision trees (also called event trees). Indeed, this approach is recommended for precisely this type of application in the excellent recent book on medical decision-making [55].

Fig. 3 presents the event/decision tree in this case. Note that, as in the visual argument (and in contrast to the formal Bayesian argument), we start with a hypothetical large number of people to be tested. This simple enhancement has a profound improvement on comprehensibility [11].

Before moving to the particular medical negligence case, it is worth considering the implications of failing to understand the true probabilities arising from medical diagnostic test results. As the Harvard question demonstrated, medical experts tend to believe that a positive test result implies a much greater probability that the patient has the disease than is really the case. Most members of the public would believe the same. But both the experts and the public are demonstrably wrong, as we have shown in Figs. 1–3 (the probability increases in this example but is still very small). In practice such misunderstanding about the true probability has been known to lead not only to unnecessary anguish by the patient but also to further unnecessary tests and even unnecessary surgery (see, for example, [20] which gives a comprehensive analysis of this phenomenon in the case of screening for breast cancer).

3. The background to the medical negligence case

The patient was an insulin-dependent diabetic admitted to hospital suffering from headaches and vomiting. Initial scans were negative, but the patient then developed a (pupil-sparing) 3rd Nerve Palsy. This condition is fairly common and, being pupil-sparing, the cause is normally ischaemic, meaning that the patient makes a full recovery without treatment. Indeed, this particular patient had previously suffered two similar such ischaemic episodes but had fully recovered from them without treatment.

However, in a small percentage of cases, the cause of a pupil-sparing 3rd nerve Palsy can be either an expanding aneurysm or a cavernous sinus pathology (CSP). In either of these cases urgent diagnosis and treatment is required, otherwise the condition can be fatal. The doctor in charge of the patient recommended an MRA (Magnetic Resonance Angiogaphy) scan be performed urgently, since this is a non-invasive test that is reasonably accurate for detecting expanding anuerusyms and CSP. However, it being a Friday evening the hospital could not offer such a test until the following Monday morning. Consequently, the patient was transferred to a specialist hospital that had the equipment to carry out the test immediately.

Contrary to the recommendation of the first doctor, the doctors at the specialist hospital decided to perform an alternative test, called a catheter angiagram (CA). This test is recognised as being more accurate than the MRA scan for diagnosing aneurysms. However, it cannot diagnose CSP at all and, as an invasive test, it carries a known 1% risk of causing a permanent stroke in diabetic patients. The CA test (which was performed early afternoon on the Saturday) indeed caused the patient to suffer a permanent stroke. The cause of the 3rd nerve palsy was subsequently found to be ischaemic (so it is assumed that the patient would have recovered without treatment).

The claim of negligence brought by the patient was based primarily on the notion that the patient was not adequately informed of the relative risks of the alternative treatments; this would have indicated that the sensible pathway was to give the MRA test because, although it carried a very small risk (5%) of non-detection (which could be fatal), this risk was tiny in comparison with the risk (1%) of a stroke from the invasive CA test.

It turned out that, despite the statistical data available, neither side could provide a coherent argument for directly comparing the risks of the alternative pathways. One surgeon advising the claimant's legal team argued that, by using Bayes Theorem, he could 'prove' the risk of the non-invasive MRA test was much less than the CA test. His proof was essentially as follows:

1. The risk from the MRA test is of the patient dying as a result of failing to detect an expanding aneurysm.
2. The prior probability of there being such an expanding aneurysm is very low.
3. Bayes Theorem calculates the posterior probability that the palsy is caused by an aneurysm given that the MRA test is negative. This probability is exceptionally low – much lower than the 1% probability of stroke risk from the CA test.

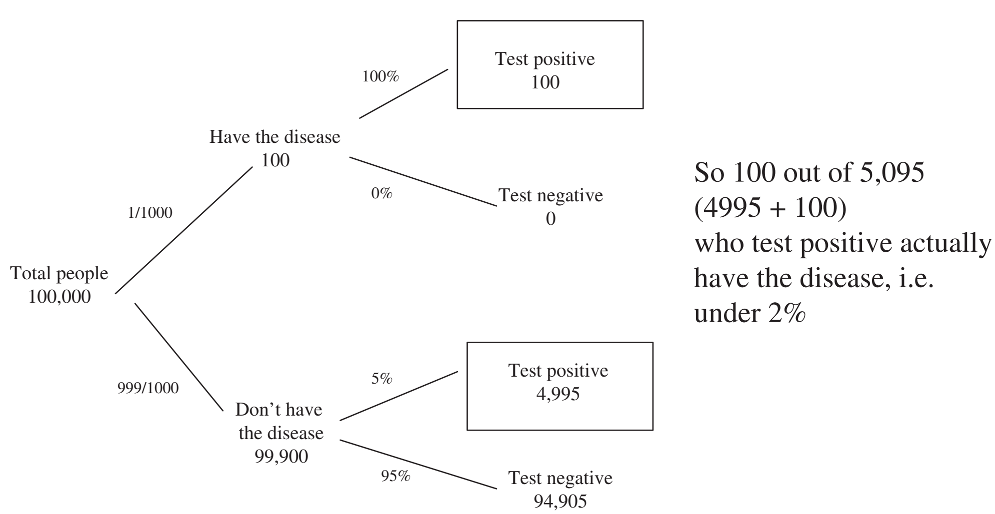

Fig. 3. Event/decision tree explanation.

However, because the surgeon used the Bayes Theorem formula neither the lawyers nor other doctors could understand his 'proof' and they sought some expert advice to both validate the argument and present it in a user-friendly way.

The proof was essentially sound, but was missing some significant variables and interactions between them (such as the different impact and detection rates of large and small aneurysms and the role of CSP as a potential alternative cause of the symptoms).

4. The formal argument in the medical negligence case

Our approach to analysing this problem was to build a causal model (known as a Bayesian network) that contained all of the relevant variables and dependencies (we discuss this in Section 5). We could then run the model in a tool that automatically performs the correct Bayesian calculations under a range of different assumptions and scenarios (including both the claimant and defence figures and differing medical opinions). This was, for us, a necessary validation to convince ourselves of the argument and results, but we were fully aware that such a model and calculations was not appropriate for the lawyers and doctors to understand and present in court.

It turned out that, although the argument was significantly more complex than the classic example presented in Section 2, it was still possible to use the event/decision tree approach. What we had to do was to present the two alternative pathways (MRA test or CA test) as two separate decision trees as shown in Figs. 4 and 5 respectively.

The decision trees show clearly the various assumptions made (in this case by the claimant's legal team). For example, it is assumed that, if an aneurysm really is the cause of the palsy then if this aneurysm is undetected there is a 2% chance that it will burst and bleed (causing death) within 48 h. The relevance of the 48 h is that this was the approximate time before an informed discussion about the patient's condition could take place (on the Monday morning), i.e. this was the time during which the ramifications of the alternative pathways was relevant.

The key thing about the decision trees is that, by starting with a hypothetical million people similar to the patient, it is easy to conceptualise at each stage the number of people who follow the separate treatment paths of the tree. For example, for the MRA test there will be approximately 10 people who die within 48 h as a result of an undetected large aneurysm, whereas with the CA test there will be approximately 9799 people with the relatively harmless ischaemic condition who would suffer a permanent stroke.

Other than an understanding of how to calculate percentages, the decision tree approach is fully understandable without any mathematics. Moreover, (using, an excel spreadsheet version) it also allows us to consider a wide range of different assumptions for example, the defence argued that the 'prior' probability of an aneurysm being the cause of the palsy could be as high as 20% rather than the 1% assumed by the claimant's lawyers, and that the accuracy of the MRA test in detecting large aneurysms could be a low as 80% rather than the 95% argued by the claimant. Such changes are simple to apply. It turned out that, even with these assumptions (and, indeed, others that were most favourable to the defence case) the risk of the invasive surgery (CA) was always significantly higher: from out of the million people the number who would get a stroke/die as a result of the invasive test was much higher than the number who would die as a result of the non-invasive test.

The numerical values resulting from the decision tree approach contrast with the probability outputs that are obtained from the purely Bayesian mathematical version of the same argument. For example, the latter results in a probability of 0.00001 of death within 48 h as a result of an undetected large aneurysm (under MRA test) and a probability of 0.009799 that someone with the

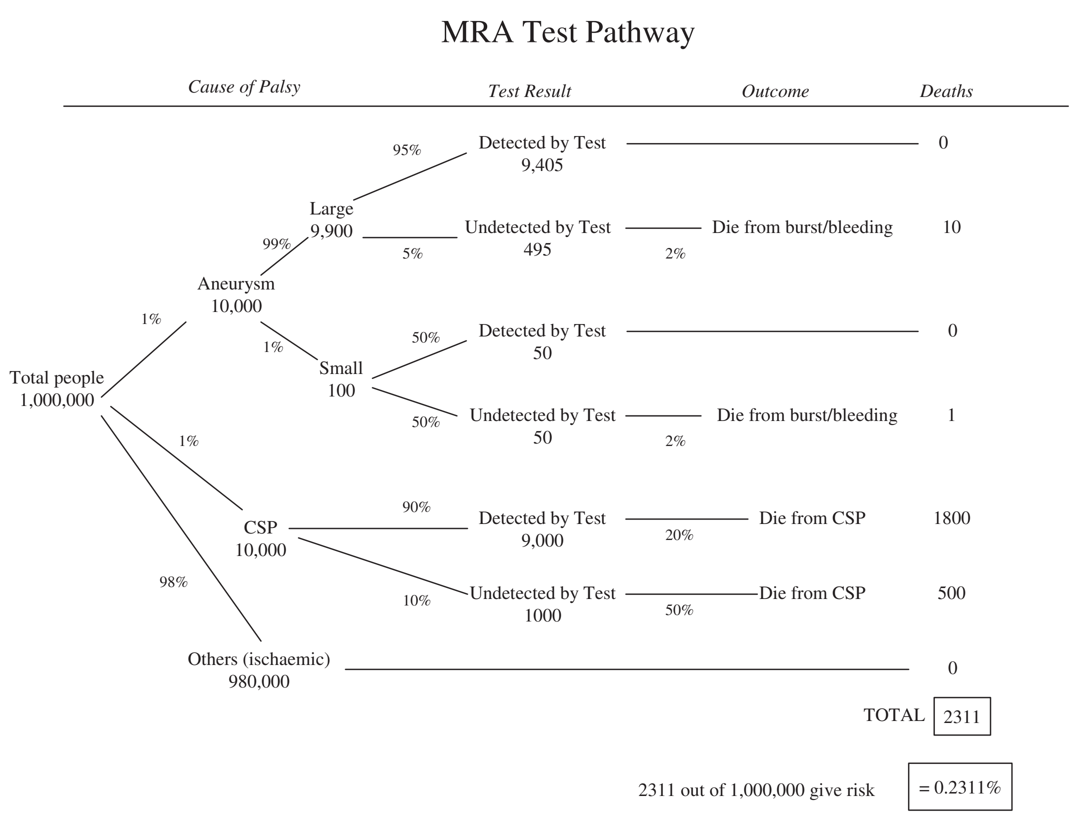

Fig. 4. Decision/event tree for MRA test pathway.

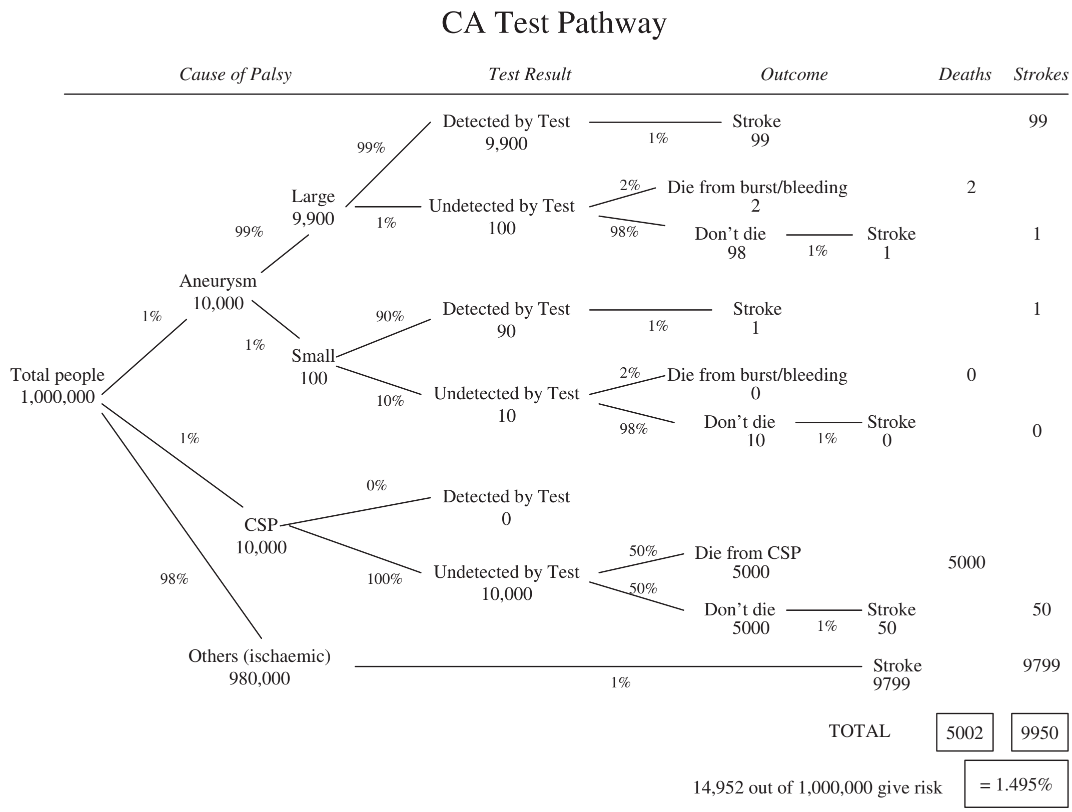

Fig. 5. Decision/event tree for catheter angiagram pathway.

harmless ischaemic condition would suffer a permanent stroke (under CA test). Both probabilities are clearly 'small' and it is difficult for lay people to appreciate the differences between these probabilities [20]. Of course, it is possible to simply turn the argument round and apply the probabilities to the same hypothetical one million patients (as we did in the decision tree approach) but this is generally unsatisfactory because, as we have argued above, lay people generally neither understood nor even believe the probabilities.

The effectiveness of the decision tree approach in the particular example was demonstrated very clearly:

- The only person working on the claimant's legal team who understood the Bayesian formulaic calculations was the expert witness surgeon who had originally provided the calculations. The others who stated that they could not understand the argument at all were: the barrister, the main lawyer working on the case, and two other doctors involved in the case as expert witnesses. Another lawyer supporting the main lawyer had a partial understanding, but insufficient to explain it to his colleagues.
- When presented with the decision tree approach every member of the above legal team said that they now understood the argument. The QC grasped it immediately and described how he would be able to use this explanation in court without having to resort to mathematics.

5. Limitations of the decision tree approach

The decision tree presentation of Bayes worked in this case because there were sufficiently few 'linked variables'.

Generally, decision trees are suitable if the following requirements are satisfied:

- The alternative pathways represented by the different decision trees are independent (in the sense that they do not rely on some common test or action that has not been modelled).
- There are no more than a small number of variables, since even if each variable had only two outcomes there are 2^n different paths for n variables. As a rule of thumb 6 is a reasonable limit.
- Each variable has only a small number of outcomes (as a rule of thumb, less than 5), So, for example, if it does make sense to consider 'size of aneurysm' in terms of a set of outcomes like {'small', 'large'}, or even {'none', 'small', 'medium', 'large'} then this can be accommodated in a decision tree. But if the outcome is on a continuous scale, say 0–1000 in millimetres, then it would not be possible to use a decision tree.
- There are no additional causes, effects and dependencies between the variables.

If these requirements are not satisfied the use of decision trees can become impractical or impossible. The same is, of course, also true of any attempt to explain Bayes from first principles using the formulaic approach; the calculations are beyond even the most experienced mathematicians.

Hence, in such circumstances we believe that the use of Bayesian networks (causal probability models) is inevitable, but raises again the issue of how to present the results in a way that is understandable to lay people. Bayesian networks (referred to subsequently as simply BNs) are graphical models (such as the one in Fig. 6) where the nodes represent uncertain variables and the arcs represent causal or influential relationships (an accessible introductory overview of BNs can be found in [18]). BNs have been fairly widely used in the medical domain since algorithmic breakthroughs in the late 1980s [34,49] meant that large-scale BN models could be efficiently calculated. Indeed, the first commercial BN software arose out of a project to construct a BN model for a particular type of medical diagnosis [3]. Clinical decision-support systems based on BNs were first developed in the late 1980s [7,24]. There have since been many hundreds of BN papers published within the medical domain. Examples include BN models for:

- diagnosis of specific diseases [2,13,23,27,28,40,43,47,57];
- predicting risk of specific diseases [8,32,53,54];
- predicting specific medical outcomes [19,25,31,52,56];
- analysing impact of treatment [12];
- analysing test results [23,39];
- improved medical procedures [6,22,36–38,42];
- cost-effectiveness analysis of different treatments [10,45].

General guidelines on using BNs in medical applications can be found in [41,44,48], while comparisons of BNs with alternative approaches in the medical context can be found in [5,15].

The BN model for the problem we have discussed is shown in Fig. 6.

Like all BNs, the model has two components.

1. The graphical component shown in Fig. 6 that describes the causal structure. Thus, for example, the graphical component tells us that 'death within 48 h' is caused/influenced by the combination of the cause (of the palsy) and whether the test correctly identifies the cause.
2. A probability table associated with each node in the model. For nodes without parents the probability table is simply the prior probabilities for each of the node states. For nodes with parents the probability table is the prior probability for each of the node states given each of the combinations of parent states. For example, the probability table for the node "test correctly identifies cause" is shown in Fig. 7.

By building the BN in a tool (such as AgenaRisk [1]) we can enter observations in the model, such as "MRA test is performed", and run the model. What happens is that the various Bayesian calculations are performed automatically and the probabilities for all of the unknown variables are revised as shown in Fig. 8. We can see, for example, that the probability of "stroke and not death" is 0% compared with just under 0.5% in the initial model, while the probability of "death within 48 h" is about 0.231% compared to 0.366% in the initial model. The figure 0.231% equates to approximately 2311 people in one million – the same result as seen in the decision tree of Fig. 4.

Similarly, Fig. 9 shows what happens in the case of the CA test. We can see, for example, that the probability of "stroke and not death" is now just under 1% compared with just under 0.5% in the initial model (and 0% in the case of the MRA test), while the probability of "death within 48 h" is just over 0.5% compared to 0.365621% in the initial model (and 0.231% in the case of the MRA test). Again these results are essentially the same as in the decision tree of Fig. 9.

So what is needed to accept the BN argument? The Bayesian calculation algorithms in tools like AgenaRisk have an established pedigree, so in principle we should need only the following:

1. *To agree on what the causal structure should be*: There are actually two stages in this process: (i) identifying a minimal necessary set of variables (nodes); and (ii) agreeing on the relevant links between the nodes. There are many books and papers that provide guidelines on both of these steps (see, for example [26,50]). Our experience suggests that they are best achieved with a BN expert and a domain expert (medical in this case) working together. If there is more than one stakeholder then genuine disagreements can be accommodated by producing alternative models (in many cases, for example, experts will not agree of the direction of the links but the resulting alternative models may still produce exactly the same results; what differs are the probabilities that need to be elicited). The BN expert can advise on which structures capture appropriate assumptions of independence and dependence between variables, and also on which structures are computationally infeasible (recommending equivalent feasible structures where appropriate).

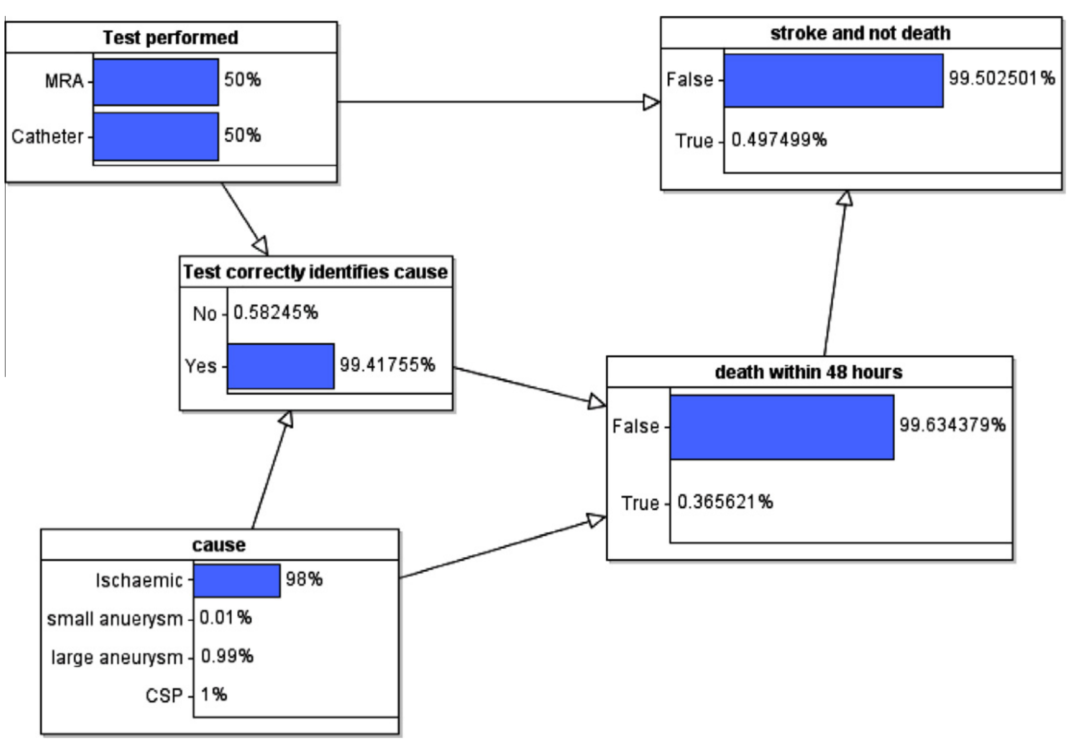

Fig. 6. Bayesian network model with initial probabilities.

Test performed	MRA				Catheter			
cause	Ischaemic	small anuerysm	large aneurysm	CSP	Ischaemic	small anuerysm	large aneurysm	CSP
No	0.0	0.5	0.05	0.1	0.0	0.05	0.01	1.0
Yes	1.0	0.5	0.95	0.9	1.0	0.95	0.99	0.0

Fig. 7. Probability table for the node "test correctly identifies cause".

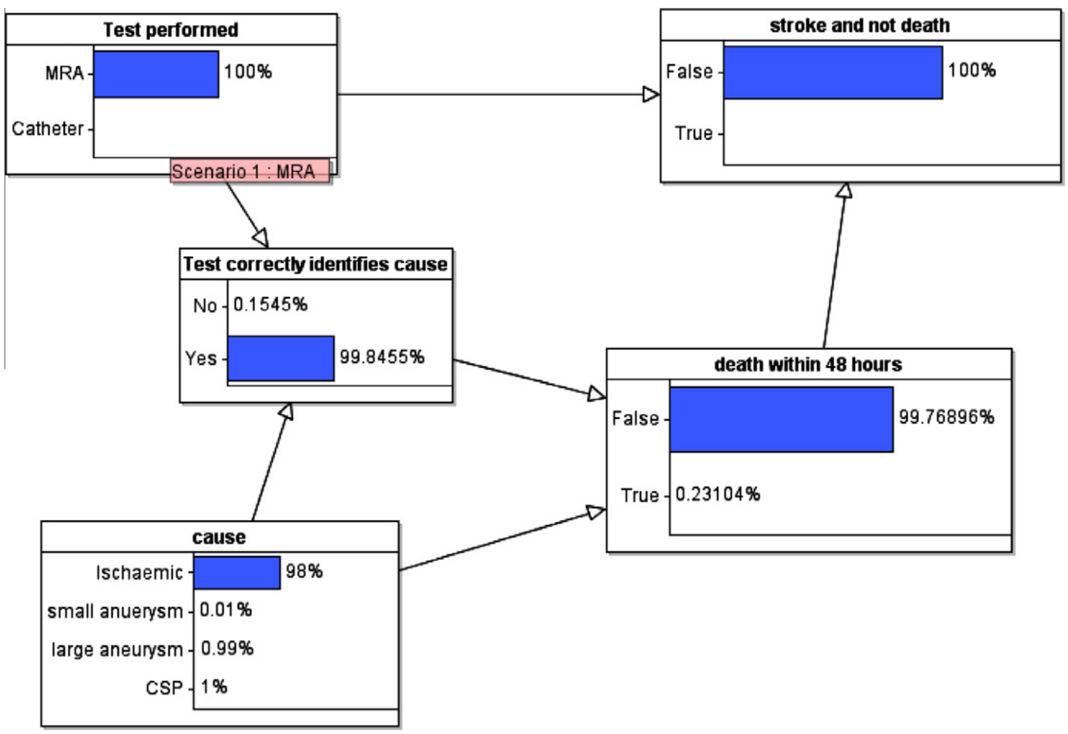

Fig. 8. Revised probabilities with MRA test performed.

2. *To agree on the values for the probability tables*. Specifically, for each node with parents we have to specify the probability of that node's states given every combination of the parent nodes' states. For nodes without parents we have to specify the prior probabilities for each state. In the example in this paper all of the probability values were taken from empirical studies provided by the medical experts (where no such data is available we have to rely on expert judgement). Where different studies provide conflicting probabilities (as in the case of a test correctly identifying a cause) we simply create alternative models representing the different values. In this case we created one version with probabilities representing the most favourable results from the defence perspective and one representing the results of the empirical studies cited by the claimant. Running both models provides results at the two extremes (in both cases the final result supported the claimant's main argument). A single BN model can also be used to test a wide range of different assumptions by creating different 'what-if' scenarios involving a range of different state observations on particular nodes. The same approach can also be used to perform sensitivity analysis. The overall impact of these methods is to lessen the dependence of individual assumptions.

It is important to note that the above required assumptions are no different from what is needed to produce the decision tree argument. However, the *real* challenge is that, whereas you might convince mathematically competent people that the Bayesian calculations (that they understood in the simple case) scale up and are calculated correctly in the tool, medical and legal professionals are reluctant to accept this. Such professionals normally expect some simple argument to lead them to the final result in all cases, and they are not convinced until the *whole* calculation is clear to them. This is, of course, impossible. It is also irrational, given the established and (mathematically) universally agreed pedigree of Bayes Theorem; the same people would surely not reject the use of calculators to perform long division on the basis that it is too difficult to understand the underlying sequence of actions that take place at the hardware circuit level. Nevertheless, the concern is real and has impeded the adoption of Bayes in both the medical and legal profession. Three factors have perpetuated the problem:

1. There is a misunderstanding of the nature of Bayes Theorem. Since the theorem is a formal mechanism for revising subjective beliefs in the light of evidence, members of both the legal and medical professions have perceived it as infringing on the role of the jury and doctors, respectively.
2. On the rare occasions where Bayes has been introduced into court, experts have attempted to explain the calculations from first principles rather than simply presenting the results of the calculations. Moreover, these first principle arguments have attempted to use the formulaic approach rather than the alternatives discussed here. In doing so they confused the jury, judge and lawyers [16].
3. Whereas there have been many prominent campaigns by statisticians and others to promote acceptance of Bayes Theorem, there have been none to our knowledge to promote acceptance of the Bayesian calculation algorithms necessary for all but the simplest problems.

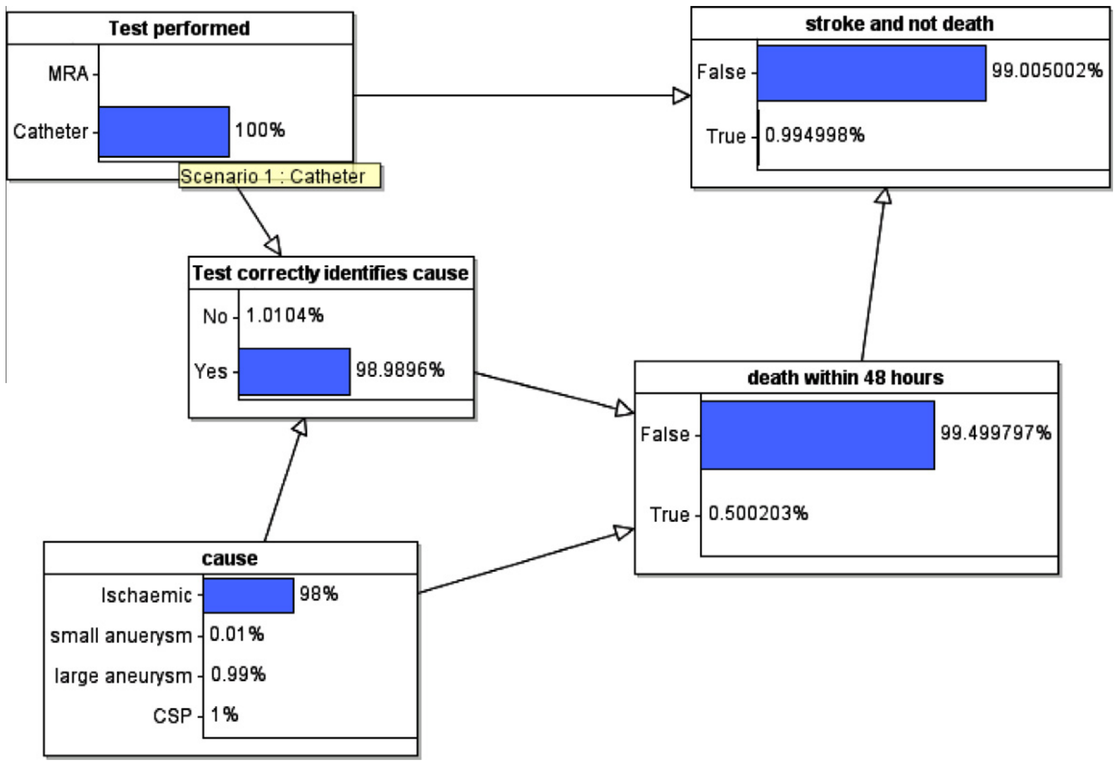

Fig. 9. Revised probabilities with catheter (CA) test performed.

In all but the simplest situations (where something like the decision tree works) it is unreasonable to expect lay people to understand the Bayesian reasoning. But this should not stop us from presenting the results of calculations from a BN model. And such results should only be challenged on the basis of the prior assumptions (causal structure and probability tables), not on the Bayesian calculations that follow from the prior assumptions.

There is one important additional issue that needs to be considered when presenting the results of a complex Bayesian calculation. Compare the following statements:

1. Out of one million people 1000 are likely to die from treatment A, but only 10 are likely to die from treatment B.
2. The probability of dying from treatment A is 0.001, but the probability of dying from treatment B is 0.00001.

Although the statements are equivalent, numerous studies have shown that statement 1 is more easily understandable to most people than statement 2 (the reference [20], for example, describes a number of such studies). Indeed, the failure of both doctors and lawyers to fully understand a statement like 2 was the reason why the original Bayesian presentation was considered unacceptable in the medical negligence case. Although it is straightforward to 'transform' a statement like 2 into a statement like 1, such a transformation should be done within the BN model rather than outside. This means incorporating numeric variables like 'number of deaths' into the model. Until very recently no BN tool was able to incorporate numeric nodes accurately (a fact which, in itself has been an impediment to more widespread use of BNs in practice). However a recent breakthrough algorithm for dynamically discretising numeric nodes [46] (described in overview form in Appendix B) makes such accurate computations possible and this algorithm is implemented in the latest version of AgenaRisk. As shown in Fig. 10 we can simply insert relevant numeric nodes with the appropriate formulas for their probability tables.

6. Summary and conclusions

We have shown how simple decision trees were used effectively to distinguish the different levels of risk of alternative diagnostic tests in a real medical negligence case. The decision tree provides a simple and clear visual explanation of an application of Bayes Theorem. Whereas lay people are known to have problems understanding the Bayesian argument when presented in the mathematical way, their understanding is radically improved by the visual representation. This method is widely generalisable.

For more complex situations involving more causal variables and dependences, the decision tree approach is not feasible. However, another visual modelling approach, namely Bayesian networks, provides an elegant solution in which all calculations are done automatically. The BN approach offers a number of advantages:

- The causal structure concretely represents legal/medical pathways that otherwise get contorted by natural language.
- The separate medical pathways are all captured in the same model (for decision trees you have to create separate trees for each pathway).
- The models can be built with different prior probability assumptions. Hence, in this case we were able to run different scenarios using both the defence and claimant assumptions. The model showed that, even using the defence's own assumptions, the risk of the CA test was greater than that of the MRA test.

The challenge is to convince medics and lawyers to accept the calculations that result from such a BN model and to focus their attention purely on the initial probability assumptions in such models.

The kind of modelling we have used also helps address a number of general and widely applicable concerns about the decision-making process within the medical profession. These concerns include:

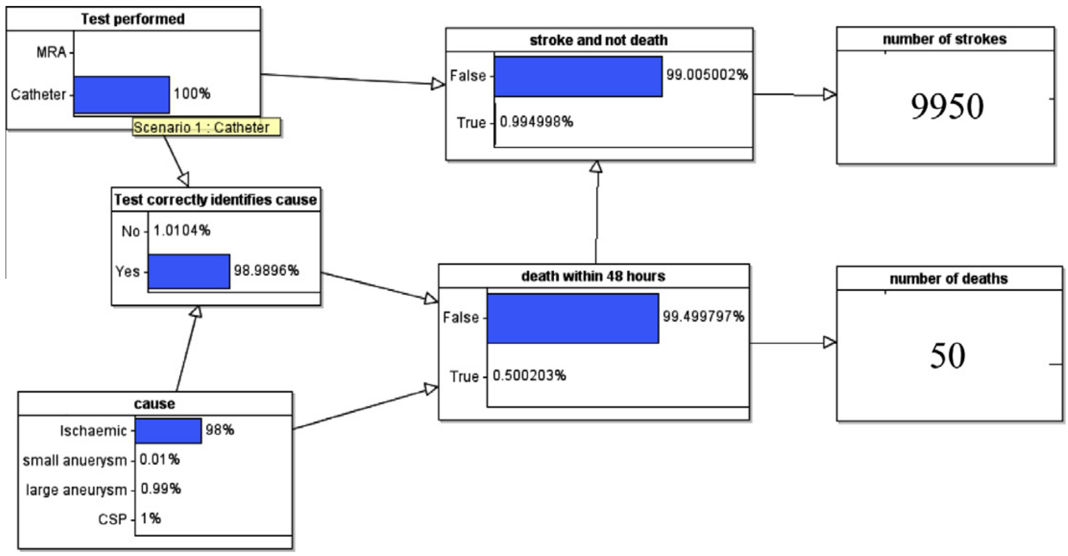

Fig. 10. Bayesian network with additional nodes for number of people (out of 1,000,000) affected.

- *the ethics of informed consent*: the results of both the decision tree and BN approaches could be used by both doctors and patients to help them make more informed decisions;
- *patient care liabilities when errors are made*: the BN models can provide proper quantification of the impact of such errors;
- *faulty research*: both the decision tree and BN approaches expose the research problem of focusing on 'true positives' while ignoring 'false positives'.

We envisage a future where doctors will have immediate access to BN-based decision-support systems that automatically provide the quantified risks of choosing alternative diagnostic test pathways for any type of condition based on 'live' data of: prior probabilities for the condition (including patient-specific data), the various test accuracy and sensitivity, and test outcomes. Moreover, the decision-support systems will be able to present the results in a form that is easily understandable to the patient as well as the doctor. While the decision as to which pathway to take ultimately still rests with the doctor and not with a computer, at least this way the doctor and patient will be properly informed of the relative risks.

Acknowledgments

We would like to thank Professor Max Parmar, whose recommendation led to our introduction to the case described in this paper. We also acknowledge the contribution of the lawyers and doctors involved in the case who provided us with many of the details and insights described. The work has also benefited from our involvement in the EPSRC project DIADEM (EP/G001987/1) and in particular colleagues William Marsh, George Hanna and Peter Hearty. Finally, we gratefully acknowledge the anonymous referees for their incisive and helpful comments, which have led to an improved paper.

Appendix A. Bayes Theorem

Let A be the event 'person has the disease'.

Let B be the event 'positive test'.

We wish to calculate the probability of A given B, which is written p(A|B).

By Bayes Theorem this is:

$$P(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} = \frac{p(B|A) \cdot p(A)}{p(B|A) \cdot p(A) + p(B|\text{not } A) \cdot p(\text{not } A)}$$

Now, in the example of Section 2, we know the following:
p(A) = 0.001.
p(not A) = 0.999.
p(B| not A) = 0.05.
p(B|A) = 1.

Hence:

$$P(A|B) = \frac{0.001}{0.001 + 0.05 \cdot 0.999}$$

which is equal to 0.1963.

Appendix B. Handling numeric nodes in Bayesian networks using dynamic discretisation

Handling continuous numeric nodes in BNs is generally difficult because (except in very special cases) there is no analytic method for computing the necessary Bayesian calculations. Consequently, continuous nodes have to be 'discretised'. For example, a node representing size of an aneurysm in millimetres cannot simply be declared to be 'in the range 0–1000'; it must be defined in terms of finite discrete intervals such as 0–10, and 10–20 etc. The standard approach to working with such continuous numeric nodes in BN tools is to use *static discretisation*, whereby the set of discretisation intervals is defined by the user in advance of any computations and do not change regardless of the evidence entered into the model. But this process is both complicated and inaccurate. You must guess the state ranges before running the calculation, thus pre-supposing that you know the resulting probability distribution of the results beforehand. In simple cases this may be quite easy, but in others it will be difficult or even impossible. The dynamic discretisation algorithm [46] addresses the problem in general by using entropy error [33] as the basis for approximation. In outline, the algorithm follows these steps:

- Convert the BN to an intermediate structure called a Junction Tree (JT) (this is a standard method used in BN algorithms and is described in, for example, [34]).
- Choose an initial discretisation in the JT for all continuous variables.

- Calculate the Node Probability Table (NPT) of each node given the current discretisation.
- Enter evidence and perform global propagation on the JT, using standard JT algorithms.
- Query the BN to get posterior marginals for each node, compute the approximate relative entropy error, and check if it satisfies the convergence criteria.
- If not, create a new discretisation for the node by splitting those intervals with highest entropy error.
- Repeat the process by recalculating the NPTs and propagating the BN, and then querying to get the marginals and then split intervals with highest entropy error.
- Continue to iterate until the model converges to an acceptable level of accuracy.

This dynamic discretisation approach allows more accuracy in the regions that matter and incurs less storage space over static discretisations. In the implementation [1] of the algorithm, the user simply specifies the range of the variable without ever having to worry about the discretisation intervals. Default settings are provided for the number of iterations and convergence criteria. However, the user can also select the number of iterations and convergence criteria, and hence can go for arbitrarily high precision (at the expense of increased computation times).

References

[1] Agena Ltd. AgenaRisk; 2010. Available from: http://www.agenarisk.com.
[2] Alvarez SM, Poelstra BA, Burd RS. Evaluation of a Bayesian decision network for diagnosing pyloric stenosis. J Pediatr Surg 2006;41(1):155–61. discussion 155–61.
[3] Andreassen S, Woldbye M, Falck B, Andersen SK. MUNIN: a causal probabilistic network for interpretation of electromyographic findings. In: 10th International Joint Conference on Artificial Intelligence, Milan, Italy; 1987. p. 366–72.
[4] Ancker JS, Senathirajaha Y, et al. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc 2006;13:608–18.
[5] Aronsky D, Fiszman M, Chapman WW, Haug PJ. Combining decision support methodologies to diagnose pneumonia. Proc AMIA Symp 2001:12–6.
[6] Athanasiou M, Clark JY. A Bayesian network model for the diagnosis of the caring procedure for wheelchair users with spinal injury. In: Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07); 2007. p. 433–8.
[7] Beinlich I, Suermondt G, Chavez R, Cooper G. The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. Second European Conf. Artif Intell Med 1989;38:247–56.
[8] Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. Radiology 2006;240(3):666–73.
[9] Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. N Engl J Med 1978;299:999–1001.
[10] Cooper N, Sutton A, Abrams K. Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section. Stat Methods Med Res 2002;11(6):491–512.
[11] Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition 1996;58:1–73.
[12] Cowell RG, Dawid AP, Hutchinson TA, Spiegelhalter DJ. A Bayesian expert system for the analysis of an adverse drug reaction. Artif Intell Med 1991;3:257–70.
[13] Cruz-Ramirez N, Acosta-Mesa HG, Carrillo-Calvet H, Alonso Nava-Fernandez L, Barrientos-Martinez RE. Diagnosis of breast cancer using Bayesian networks: a case study. Comput Biol Med 2007;37(11):1553–64.
[14] Dawid AP. Bayes's theorem and weighing evidence by juries. In: Swinburne R, editor. Bayes's Theorem: proceedings of the British academy, vol. 113. Oxford: Oxford University Press; 2002. p. 71–90.
[15] Dexheimer JW, Brown LE, Leegon J, Aronsky D. Comparing decision support methodologies for identifying asthma exacerbations. Study Health Technol Inform 2007;129:880–4.
[16] Donnelly P. Appealing statistics. Significance 2005;2(1):46–8.
[17] Eddy DM. Probabilistic reasoning in clinical medicine; problems and opportunities. In: Kahneman D, Slovic P, Tversky A, editors. Judgment under uncertainty: heuristics and biases. Cambridge University Press; 1982. p. 249–67.
[18] Fenton NE, Neil M. Managing risk in the modern world: Bayesian networks and the applications, London mathematical society, knowledge transfer report; 2007. www.lms.ac.uk/activities/comp_sci_com/KTR/apps_bayesian_networks.pdf.
[19] Gevaert O et al. Predicting the outcome of pregnancies of unknown location: Bayesian networks with expert prior information compared to logistic regression. Hum Reprod 2006;21(7):1824–31.
[20] Gigerenzer G. Reckoning with risk: learning to live with uncertainty. London: Penguin Books; 2002.
[21] Goodman J. 'Jurors' comprehension and assessment of probabilistic evidence. Am J Tr Advoc 1992;16:361.
[22] Haddawy P, Kahn Jr CE, Butarbutar M. A Bayesian network model for radiological diagnosis and procedure selection: work-up of suspected gallbladder disease. Med Phys 1994;21(7):1185–92.
[23] Hamilton PW, Anderson N, Bartels PH, Thompson D. Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast. J Clin Pathol 1994;47(4):329–36.
[24] Heckerman DE, Horvitz EJ, Nathwani BN. Towards normative expert systems: part I – the pathfinder project. Methods Inf Med 1992;31:90–105.
[25] Hoot N, Aronsky D. Using Bayesian networks to predict survival of liver transplant patients. AMIA Annu Symp Proc 2005:345–9.
[26] Jensen FV, Nielsen T. Bayesian networks and decision graphs. New York: Springer-Verlag Inc.; 2007.
[27] Kahn Jr CE, Laur JJ, Carrera GF. A Bayesian network for diagnosis of primary bone tumors. J Digit Imaging 2001;14(2 Suppl. 1):56–7.
[28] Kahn Jr CE, Roberts LM, Shaffer KA, Haddawy P. Construction of a Bayesian network for mammographic diagnosis of breast cancer. Comput Biol Med 1997;27(1):19–29.
[29] Kahneman D, Slovic P, Tversky A. Judgement under uncertainty: heuristics and biases. New York: Cambridge University Press; 1982.
[30] Kaye DH, Koehler JJ. Can jurors understand probabilistic evidence? J R Stat Soc Ser A 1991;154(1):75–81 [Statistics in society].
[31] Kazi JI, Furness PN, Nicholson M. Diagnosis of early acute renal allograft rejection by evaluation of multiple histological features using a Bayesian belief network. J Clin Pathol 1998;51(2):108–13.
[32] Kline JA, Novobilski AJ, Kabrhel C, Richman PB, Courtney DM. Derivation and validation of a Bayesian network to predict pretest probability of venous thromboembolism. Ann Emerg Med 2005;45(3):282–90.
[33] Kozlov AV, Koller D. Nonuniform dynamic discretization in hybrid networks, In: Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI) Providence, Rhode Island; 1997. p. 314–25.
[34] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). J. R. Stat. Soc. B 1988;50, No. 2:157–224.
[35] Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 1959;130(3366):9–21.
[36] Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting hospital admission for Emergency Department patients using a Bayesian network. AMIA Annu Symp Proc 2005:1022.
[37] Lehmann H, Shortliffe E. Thomas: building Bayesian statistical expert systems to aid in clinical decision making. Comput Methods Programs Biomed 1991;35(4):251–60.
[38] Lin J-H, Haug PJ. Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. J Biomed Inform 2008;41(1):1–14.
[39] Lucas PJ, van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artif Intell Med 2004;30(3):201–14.
[40] Luciani D et al. Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. Emerg Med J 2007;24(3):157–64.
[41] Mani S, Valtorta M, McDermott S. Building Bayesian network models in medicine: the MENTOR experience. Appl Intell 2005;22(2):93–108.
[42] Marshall AH, McClean SI, Shapcott CM, Hastie IR, Millard PH. Developing a Bayesian belief network for the management of geriatric hospital care. Health Care Manag Sci 2001;4(1):25–30.
[43] McKendrick IJ, Gettinby G, Gu Y, Reid SW, Revie CW. Using a Bayesian belief network to aid differential diagnosis of tropical bovine diseases. Prev Vet Med 2000;47(3):141–56.
[44] Montironi R, Whimster WF, Collan Y, Hamilton PW, Thompson D, Bartels PH. How to develop and use a Bayesian belief network. J Clin Pathol 1996;49(3):194–201.
[45] Negrín M, Vázquez-Polo F. Bayesian cost-effectiveness analysis with two measures of effectiveness: the cost-effectiveness acceptability plane. Health Econ 2006;15(4):363–72.
[46] Neil M, Tailor M, Marquez D. Inference in hybrid Bayesian networks using dynamic discretization. Stat Comput 2007;17(3):219–33.
[47] Nikiforidis G, Sakellaropoulos G. Expert system support using Bayesian belief networks in the prognosis of head-injured patients of the ICU. Med Inform (Lond) 1998;23(1):1–18.
[48] Nikovski D. Constructing Bayesian networks for medical diagnosis from incomplete and partially correct statistics. IEEE Trans Knowledge Data Eng 2000;12(4):509–18.
[49] Pearl J. Fusion, propagation, and structuring in belief networks. Artif Intell 1986;29:241–88.
[50] Pearl J. Causality: models reasoning and inference. Cambridge University Press; 2000.
[51] Piatelli-Palmarini M. Inevitable illusions: how mistakes of reason rule our minds. John Wiley & Sons Inc.; 1994.

[52] Sakellaropoulos GC, Nikiforidis GC. Development of a Bayesian network for the prognosis of head injuries using graphical model selection techniques. Methods Inf Med 1999;38(1):37–42.
[53] Sanders D, Aronsky D. Prospective evaluation of a Bayesian network for detecting asthma exacerbations in a Pediatric Emergency Department. AMIA Annu Symp Proc 2006:1085.
[54] Sanders DL, Aronsky D. Detecting asthma exacerbations in a pediatric emergency department using a Bayesian network. AMIA Annu Symp Proc 2006:684–8.
[55] Sox HC, Blatt MA, et al. Medical decision making. ACP Press; 2007.
[56] Verduijn M, Rosseel PM, Peek N, de Jonge E, de Mol BA. Prognostic Bayesian networks II: an application in the domain of cardiac surgery. J Biomed Inform; 2007;40(6):619–30.
[57] Warner HR, Toronto AF, Veasy LG. Experience with Baye's Theorem for computer diagnosis of congenital heart disease. Ann NY Acad Sci 1964;115:558–67.