PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Probabilistic Expert Systems and Graphical Modelling: A Case Study in Drug Safety

David J. Spiegelhalter, A. Philip Dawid, Tom A. Hutchinson and Robert G. Cowell

Phil. Trans. R. Soc. Lond. A 1991 **337**, 387-405
doi: 10.1098/rsta.1991.0131

Probabilistic expert systems and graphical modelling: a case study in drug safety

By David J. Spiegelhalter[1], A. Philip Dawid[2], Tom A. Hutchinson[3] and Robert G. Cowell[2]

[1]*MRC Biostatistics Unit, 5 Shaftesbury Road, Cambridge CB2 2BW, U.K.*
[2]*Department of Statistical Science, University College London, U.K.*
[3]*Royal Victoria Hospital, Montreal, Canada*

Probabilistic expert systems are intended to provide reasoned guidance in complex environments characterized by extensive uncertainty. An explicit 'causal' model is constructed for the process being observed, in which an acyclic directed graph is used to express conditional independence assumptions about variables, and probability assessments specify a full joint probability distribution. The resulting graphical structure can cope with a range of issues that arise in realistic modelling. Here we consider a particular example of assessing the chance that a suspected adverse reaction is due to a particular drug under suspicion. The background biological knowledge provides an appropriate model and probability assessments are obtained from expert microbiologists. The model allows a variety of interpretations for 'causality'. Details of the graphical and computational algorithms used to perform efficient calculations of conditional probabilities on complex graphical structures are provided and illustrated with the example. Further developments should allow updating of the risk parameters in the light of a series of case reports, and may form the basis for a flexible expert system for causality assessment and post-marketing surveillance.

1. Introduction

Expert systems are computer programs that are intended to provide reasoned guidance in complex situations. Uncertainty pervades many applications, due to missing information or an inability to build a purely logical model for the process being studied, and a variety of approaches have been suggested for dealing with this problem: fuzzy logic, certainty factors, non-monotonic reasoning, belief functions have all been suggested; see Shafer & Pearl (1990) for a wide collection of papers. The probabilistic approach has been hampered by apparent computational difficulties, but recent developments have now made it feasible to deal with real applications using specialist software for probabilistic reasoning.

In this paper we select a particular example to illustrate the methodology, which emphasizes the use of background scientific knowledge as a basis for a model that can be used for handling individual cases in a coherent way. The problem is in the context of routine post-marketing surveillance of pharmaceutical products, which requires the processing of reports of suspected adverse reactions. The accurate identification of the cause of the reaction is of great importance both to the pharmaceutical industry and to regulatory authorities, and yet the poor quality of the received data can make such 'causality assessment' exceptionally difficult.

Phil. Trans. R. Soc. Lond. A (1991) **337**, 387–405
Printed in Great Britain

In the next section we briefly review current methods of causality assessment, including a probabilistic procedure, and in §3 we introduce a specific problem concerning the development of a diarrhoeal reaction (pseudomembranous colitis (PMC)) following antibiotic therapy. In §§4 and 5 we step through the stages in representing the PMC problem as a 'causal network', showing how the basic structure can be extended in a modular fashion to provide an increasingly realistic and complex model. Our description emphasizes the exploratory nature of our model building but, while certainly not claiming to have a definitive representation of the process under study, we hope to demonstrate the concise manner in which complex phenomena can, with care, be modelled by this approach. In §6 we summarize technical aspects of handling causal network representations of disease processes that have been developed in the context of expert systems, emphasizing the intuitive graphical structure and its re-representation for computational purposes.

We then give an example based on an actual case report, and demonstrate sensitivity analysis of the conclusions to assumptions about the context in which the reaction developed.

The conclusions that the model draws about the drug responsible for a reaction in a specific patient depend crucially on the assumptions concerning the potential of that drug to initiate the chain of events that led to the reaction, and these assumptions are open to question. Thus, while identifying the cause in individual cases is certainly important, it is even more vital to use the reports to learn more about the propensity of a drug to cause harm. This inverse process is difficult in view of the mainly highly selected nature of the reports that are received, but some tentative suggestions are made in our concluding section.

The general framework of our approach is currently being explored in a variety of applications, which include the diagnosis of congenital heart disease (Spiegelhalter & Cowell 1991), monitoring in anaesthesia (Beinlich *et al.* 1989), diagnosis of neurological disorders (Andreassen *et al.* 1989) and decision making in diabetes (Andreassen *et al.* 1991). Each of these applications requires probabilistic identification of an underlying causal mechanism that will best explain the observed constellation of findings on individual cases; they are not concerned with general identification of causal mechanisms for populations. This paper emphasizes the technical issues: Hutchinson *et al.* (1991*b*) and Cowell *et al.* (1991), respectively, discuss our problem from the medical and computing points of view.

2. Causality assessment: background and an example

The drug surveillance group of a pharmaceutical company or regulatory authority will regularly receive reports of adverse clinical events that are suspected to be reactions to previous therapy. The reports will usually give some background information on the patient, the date and type of reaction, the dates on which the suspected drug was taken, and possibly the effects of stopping the drug ('dechallenge'). However, vital information is often missing, such as other drugs being taken and the clinical context in which the reaction occurred.

Four basic methods for processing these reports can be identified (Hutchinson & Lane 1989). First, the global approach requires the assessor to examine the evidence and make an intuitive overall judgement as to whether the reaction was certainly/probably/possibly/unlikely to have been caused by the drug in question. The irreproducibility of such an approach has been shown in observer agreement

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

studies (Karch *et al.* 1976), and this has led to two more structured procedures. The algorithmic, or flowchart, method requires the assessor to step through a sequence of questions that logically lead to a conclusion. For example, a U.S. Food and Drug Administration algorithm (Turner 1984) first asked if the current has a reasonable temporal association from the drug, and if so it asked if there was a dechallenge from the drug (i.e. did the reaction cease when the drug was stopped); if this information is not available, the algorithm immediately classified the causal relationship as 'possible' and asked no more questions.

Such algorithms seem unnecessarily rigid and make limited use of available information; this has led to scoring systems in which points are accumulated from aspects of the case evidence, and if the total points lie in a certain interval then the assessment is 'probable', etc. (Venulet 1984). Hutchinson & Lane (1989) have pointed out that such scores are arbitrary and ignore the relative strengths of the items of evidence, and with other authors (Lane *et al.* 1987; Lane 1989) they have suggested a bayesian probabilistic approach, in which the probability of causation is explicitly calculated. We emphasize that, in the following development, we do not claim that the idea of 'causation' can be uniquely defined. A variety of interpretations are possible, but we shall try to be precise as to which we are using in any context. As we shall find in §5, one of the major strengths of the bayesian graphical approach is that it allows a number of quantities to be calculated, each of which could be argued to be the probability of 'causation'.

Suppose there are two competing causes, A and B, of a specified type of reaction E, where B could represent 'natural causes'. Let $A \rightarrow E$ denote the event that E occurred and that A alone 'caused' E, which we take to mean that if A alone had been given then E would still have occurred exactly as it did (in the following sections we consider assigning joint responsibility to two drugs simultaneously). Let the background information (type of patient, drug dosage and schedule, class of reaction, time horizon within which the occurrence of the specific clinical problem is considered as a reaction) be denoted by M. Then the odds on A against B causing the reaction, before incorporating the occurrence of specific details of the reaction, are

$$p(A \rightarrow E \,|\, M)/p(B \rightarrow E \,|\, M). \tag{1}$$

After detailed findings F are obtained (e.g. timing of the reaction relative to time of taking the drugs, effects of dechallenge, rechallenge, etc.), the revised posterior odds are, by Bayes's theorem,

$$\frac{p(A \rightarrow E \,|\, F, M)}{p(B \rightarrow E \,|\, F, M)} = \frac{p(F \,|\, A \rightarrow E, M)}{p(F \,|\, B \rightarrow E, M)} \times \frac{p(A \rightarrow E \,|\, M)}{p(B \rightarrow E \,|\, M)}, \tag{2}$$

or

$$\text{posterior odds} = \text{relative likelihood} \times \text{prior odds}.$$

The relative likelihood is an assessment of how much more likely is the specific case evidence were A to have caused the reaction as against B. Figure 1 displays a 'high-level' representation of the process being modelled.

The arrows indicate dependencies to be taken into account when considering the probability of any random (elliptical) node in such a graph (rectangular nodes indicate fixed context). Thus figure 1 represents terms in expressions (1) and (2): the probability of each possible cause depends directly on case parameters (M), while the probability of case findings (F) depends on both cause and background parameters. The crucial element is that while a graph may be constructed by thinking causally,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

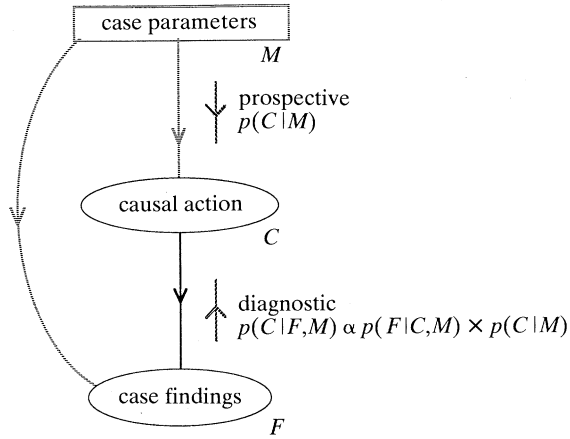

Figure 1. Overview of process being modelled: prospective modelling of causal mechanism given case parameters, and specific case findings given case parameters and cause. After observing F, Bayes's theorem 'reverses the arrow' between C and F, and allows the calculation of the posterior probability of cause given M and F.

Table 1. *Background and reaction details of patient* (*timings relative to start of drug A*)

background information M	
patient	female, aged 29
drugs given	antibiotic A day 0–day 10 antibiotic B day 12–day 22
reaction observed	pseudomembranous colitis (PMC)
details (F) of reaction	
timing	occurred on day 27

when specific case findings are actually observed this diagnostic evidence needs to be propagated back to the causal action node. Figuratively the arrow from causal action to case findings needs to be reversed; technically this is precisely the role of the likelihood terms used in Bayes's theorem (2).

In many examples it may be reasonable to assume the items of evidence in D are conditionally independent given the true cause, and then the relative likelihood breaks into a product of terms that may be individually assessed. Examples of such analyses are given in the Proceedings of a Drug Information Workshop (1986). However, we now introduce an example which appears to demand additional sophistication in its probabilistic modelling.

3. A specific example of a reaction

We now consider an actual case report received by the drug surveillance department of a pharmaceutical company. The relevant background and details of the reaction are given in table 1.

In a typical scoring system each drug is assessed individually without directly comparing their likelihoods of having led to the event. In contrast, a probabilistic approach as described in the previous section would first assess the prior odds that A rather than B would cause a reaction within some time horizon that defines the occurrence of a reaction, and then assess the relative likelihood of the reaction

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

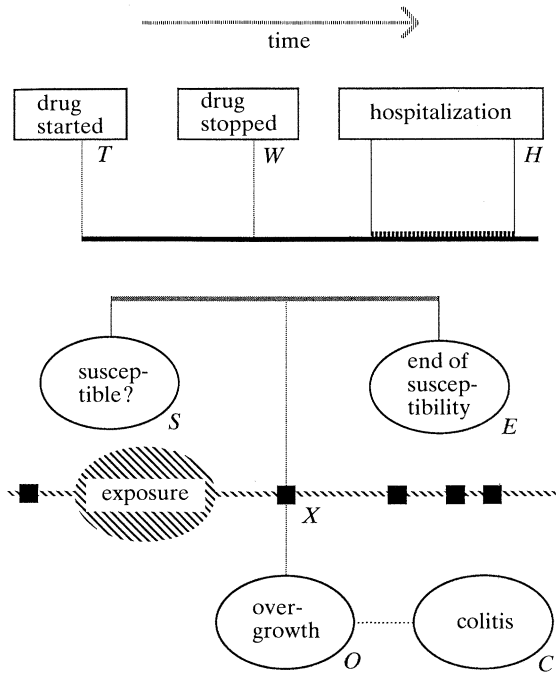

Figure 2. Schematic diagram for process under study. The top level represents the timing of the treatment of the patient, and the possible 'window of susceptibility' is shown below. Next comes the exposure process to *C. difficile*, which is shown as being intensified while the patient is in hospital. If exposure occurs within the window of susceptibility then overgrowth is assumed to happen, followed a variable number of days later by clinical evidence of colitis.

occurring on day 27 (the only detailed information), given it occurred within the specified horizon. See Hutchinson *et al.* (1991*a*) for an example of such an analysis using a spreadsheet program.

Such an assessment is, however, somewhat unsatisfactory since a considerable amount is known about the mechanism by which an antibiotic can lead to the development of PMC, and a conscientious assessment should take this knowledge into account, explicitly and quantitatively. In collaboration with an expert microbiologist, Dr Douglas Burdon, we have developed a formal model to describe this process. Specifically, an antibiotic depletes the bowel flora in such a way that, if the patient is then exposed to the particular virulent bacterial organism *Clostridium difficile* (abbreviated to *C. difficile*), this organism may colonize the gut. It is this overgrowth that after a few days will lead to clinical signs of PMC-induced diarrhoea.

The process is shown in figure 2. A drug is started at time T (taken as 0 in the computer analysis) and taken until time W. Soon after T there is a chance that a 'window of susceptibility' is opened up due to modification of gut flora (time S), and this stays open until time E, which is some unknown period after the drug is stopped. In all our analysis we have assumed that $S = T+1$ if the patient becomes susceptible; i.e. if the drug is going to alter the gut flora, it does so by the end of the first day of treatment. Meanwhile, and quite independently, there is an 'exposure process' going on in which each day there is a certain chance of being exposed to *C. difficile*. The day of first exposure after time $T+1$ is the 'relevant' exposure day (X). If this occurs within the window of susceptibility, then on this day overgrowth occurs

(O), and soon after colitis is observed (C). The diagram shows that if the patient is in hospital, particularly in a ward where there is an epidemic of *C. difficile*, then the exposure rates may increase dramatically.

Various additional complexities may be relevant. First, after two drugs have been given, then even if they act independently both may have their windows of susceptibility open when an exposure occurs and so the unique cause is not ascribable, i.e. if either drug had not been given the reaction would still have occurred, but equally if either drug had been given alone the reaction would have occurred. Secondly, hospitalization and the existence of an epidemic may change the daily exposure rate. Thirdly, it is possible that certain antibiotics, particularly when given intravenously, may actually 'block' overgrowth while they are being administered since they may also act on *C. difficile*, and hence relevant exposure can only take place after the drug has been stopped. Finally, most of the relationships described in the model are stochastic and their parameters can only be assessed subjectively by expert microbiologists on the basis of limited experience: examples include the daily chance of exposure of *C. difficile* in the community, the extent to which susceptibility continues after the drug is stopped, the chance an individual becomes susceptible at all and the lag between overgrowth and development of colitis.

The complexity of the problem strongly suggests that intuitive assessments of likelihood ratios may not be reliable, and hence we have investigated a formal analysis using a graphical approach that uses recent developments within the general context of expert systems.

4. A bayesian graphical representation of the causality assessment problem

Systems intended to help clinicians classify patients into diagnostic or prognostic states exist in similar degrees of development to those in causality assessment. There are global judgments in which only structured information is available, algorithms, developed either informally or based on data analysis, and scoring systems with their weights again either assessed informally or through statistical discriminant analysis. Recent work in artificial intelligence in medicine has, however, emphasized the use of causal networks, in which an attempt is made explicitly to model the qualitative medical knowledge that is available concerning the disease processes under consideration (Szolovits *et al.* 1988). These models are represented by directed graphs in which links represent direct influences, and absent links indicate conditional independence assumptions. Attached to these links are conditional probability tables that quantitatively express the uncertainty on the links, and the aim is to be able to observe fragmentary evidence, possibly sequentially, on parts of the graph, and propagate the effect of this evidence through the graph to revise the probability of the nodes of direct interest. Such graphical representations have a long history, beginning with Wright (1934) and continuing in the areas of structural modelling in the social sciences, while there is also a clear relationship with the pedigree graphs used in genetics (Thompson 1986).

In our context we need to develop a directed graphical representation for the qualitative mechanism by which a reaction might occur when one or two antibiotics had been previously taken, and provide assessments of the necessary conditional probability distributions. The theory for probability calculations on such networks has been developed by Pearl (1986, 1988) and Lauritzen & Spiegelhalter (1988), and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

in §6 we describe the computational algorithm by which these initializing conditional probabilities may be manipulated to allow efficient calculation of the required outputs once evidence on a case has been obtained; the required outputs will generally be the probabilities of causation and the evidence will generally be the day on which colitis is noted, although the computational procedures are not restricted to these circumstances.

There are four main stages in constructing a network and turning it into a representation ready to receive evidence, and two further stages in processing evidence as it arrives.

(*a*) *Representation of the qualitative knowledge as a directed acyclic graph* (DAG)

The nodes of the graph are random variables and direct influences of a node v are 'parents' of v (denoted $\mathrm{pa}(v)$) in the graph. Formally, the graph represents the assumption that the joint distribution of the whole set V can be expressed as

$$p(V) = \Pi_v\, p(v \mid \mathrm{pa}(v)), \tag{3}$$

i.e. the joint distribution is made up of terms which model the conditional distribution of each node given its direct influences. This is equivalent to a conditional independence assumption that, if we are told the values of $\mathrm{pa}(v)$, then v is independent of all other nodes in the graph except direct descendants of v (see Lauritzen *et al.* (1990) for a full description of the conditional independence statements that can be read off a DAG). Such assumptions follow naturally in a context such as genetics (Thompson 1986), in which a child's genotype is only directly influenced by its parents, and so if the parents' genotypes are known then no other individual in a pedigree provides information about the child's genotype (apart from descendants of the child). The essence of our approach is to make similar conditional independence assumptions within other domains, and then exploit (3) to allow the full joint distribution to be obtained by a set of assessments of individual terms $p(v \mid \mathrm{pa}(v))$.

We first consider the simplified situation in which only one drug has been taken; this is appropriate if we were to assume either A or B was the cause of the reaction, with no allowance made for simultaneous causation. Thus in each one-drug model we are assuming we know the causative agent, but we may use the model to calculate prior and likelihood terms for A and B individually, to be processed using (2).

Figure 3 shows a network representation of the essential elements of the complex processes summarized in figure 2. Each elliptical node represents a random quantity, specifically the time, including ∞ to represent 'never' where appropriate, of the critical events defined in the legend. The boxes represent fixed contexts defined by the background information on the case. The 'expert' biological knowledge allows us to make the conditional independence assumptions represented by the graph; for example, that given the day on which overgrowth occurs, the day on which colitis first appears is independent of all other events in the network. Each such assumption translates into a formal conditional probability statement, for example, $p(C \mid O,$ other events$) = p(C \mid O)$, and we obtain from (3) an expression for the joint distribution of all the unknown quantities:

$$p(C, O, E, S, X \mid M = \{H, T, W\}) = p(C \mid O)\, p(O \mid E, S, X)\, p(E \mid S, W)\, p(S \mid T)\, p(X \mid H), \tag{4}$$

for which we need only obtain assessments of the distribution of each quantity conditional on its direct influences (parents) in the graph. Once we have such

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

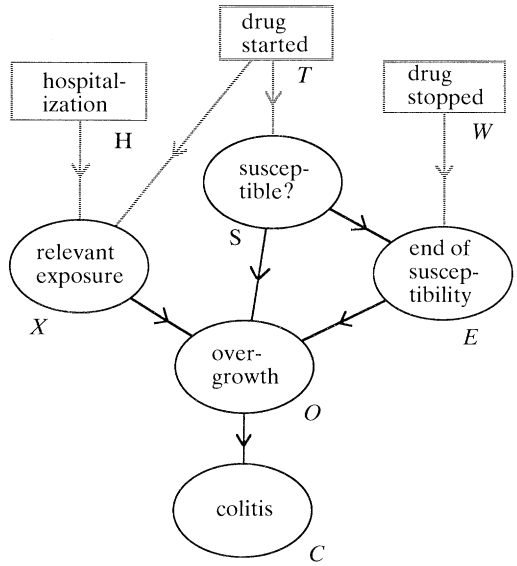

Figure 3. Graphical model of process when only one drug is being taken. Case parameters: H, days of entering and leaving hospital; T, day treatment started; W, day treatment withdrawn. Internal variables: X, day of first relevant exposure to *C. difficile*; S, day of start of susceptibility to overgrowth; E, day susceptibility ends; O, day overgrowth occurs. Case findings: C, day colitis appears.

assessments the joint distribution can be represented in a suitable decomposed form and, given any evidence on any node in the graph, conditional probabilities of other nodes can be obtained by the procedures outlined in §6.

(*b*) *Quantitative conditional probability distributions*

This brings us to the second stage of specifying quantitative conditional probability distributions for each variable v given each configuration of its parent nodes. In genetics these may be generated by known inheritance laws, but in general they may be considered parameters of the system that need to be either estimated from data or subjectively assessed. It is clearly essential to use relevant data if it is available, but in this context no analytic study has apparently been performed. We therefore consider specific models for each of the terms in (4), and provide some numerical assessments based on the judgement of Dr Burdon.

In our application the assessments have been specified to point values as if they were precisely known, but Spiegelhalter & Lauritzen (1990) show how the conditional probabilities can themselves be considered as random quantities that form an additional layer on the graph, and formal bayesian analysis then allows updating of the distributions on those probabilities as data accumulates over a set of cases. In general we would view subjective judgements as essentially providing a 'head-start' in order to begin a system, and Spiegelhalter & Cowell (1991) show that, if the subjective assessments are assumed to be fairly imprecise then they are rapidly overwhelmed by accumulating data. The possibility of adopting such a 'learning' procedure within adverse drug reactions is touched upon in our conclusions.

Node X: day of first relevant exposure to C. difficile. We assume that the first possible relevant day of exposure is $T+1$ and, in the absence of evidence to the contrary, assume that thereafter daily exposures are independent events with probability $\lambda(d)$

on day d, which may depend on whether the patient is in hospital (H). Hence the distribution of the day of first exposure is assumed to be a geometric distribution

$$p(X = d \,|\, H) = \lambda(d) \prod_{t=T+1}^{d-1} \{1-\lambda(t)\} \quad (d = T+1, T+2, \ldots).$$

Initial values for the daily risks have been set at

$$\lambda(d) = \begin{cases} 0.0002, & \text{if patient is outside hospital on day } d, \\ 0.01, & \text{if patient is in hospital on day } d, \\ 0.9, & \text{if there is a known epidemic of } C.\ \textit{difficile} \text{ on day } d \text{ in the hospital.} \end{cases}$$

Node S: day of susceptibility of overgrowth. We assume that susceptibility to overgrowth occurs with probability σ, and if it does occur then is on day 1. Hence the distribution for S is

$$p(S = T+1 \,|\, T) = \sigma,$$
$$p(S = \infty \,|\, T) = 1-\sigma.$$

A typical value for σ might be 0.9. This quantity is the essential determinant of the adverse risk associated with the drug, and we shall later discuss the issue of learning about this crucial parameter from epidemiological data.

Node E: the day susceptibility ends. If susceptibility has occurred, then it is thought to be certain to continue for a period, say g days, after the drug has been stopped (day W). After this, the susceptibility may or may not continue for some time. We have modelled this by a geometric distribution that results from assuming there is a chance m, known as the susceptibility stop-rate, of ceasing to be susceptible on any day. Thus the distribution for E is given by

$$\text{if susceptible;} \quad p(E = d \,|\, S = T+1, W) = m(1-m)^{d-W-g-1} \quad (d = W+g+1, W+g+2, \ldots);$$
$$\text{if not susceptible;} \quad p(E = \infty \,|\, S = \infty, W) = 1.$$

After extensive discussion, values of $g = 5$ and $m = 0.05$ have been assessed, which gives an expectation of about 3.5 weeks for susceptibility continuing after the drug is stopped.

Node O: the day overgrowth occurs. This is assumed to follow logically as soon as relevant exposure occurs while the patient is susceptible, and to happen on the day on which this combination of circumstances first occurs. Thus O is defined logically in terms of its influences X, S and E, and to have distribution as follows:

$$\text{if susceptible;} \quad p(O = d \,|\, X = d, S = T+1, E) = \begin{cases} 1 & \text{if} \quad T+1 \leqslant d \leqslant E, \\ 0 & \text{if} \quad E < d. \end{cases}$$
$$\text{if not susceptible;} \quad p(O = \infty \,|\, X, S = \infty, E = \infty) = 1.$$

Node C: day colitis appears. There is a lag of a few days between overgrowth occurring and the symptoms of colitis becoming apparent, and we assume a discrete distribution

$$p(C = d \,|\, O) = p(d-O) \quad (d = O+1, O+2, \ldots).$$

A plausible distribution assumes $p(2) = 0.47$, $p(3) = 0.47$, $p(4) = 0.05$, $p(5) = 0.01$.

5. The two-drug model

We now consider the graphical representation of the two-drug model shown in figure 4 (the background parameters have been left implicit). This has two highlighted components that exactly match the one-drug model considered above, in that an exposure process interacts with a susceptibility process to lead to overgrowth. However, the individual models for the two drugs do meet in two aspects: they have a common exposure process (shown at the top of the graph) and they have a possible common consequence in the development of colitis. We now step through the graph and provide details of the dependencies and any additional numerical assessments. It should be clear that the representation shown below was not immediately apparent, and modelling the process with an appropriate graphical structure took some time and many false starts.

Nodes Y, X_1 and X_2: the exposure process. The graphical formulation of the exposure process expresses that relevant exposure may occur for the first drug before the second is started, but if it does not then any subsequent exposure is relevant to both drugs simultaneously. Hence Y and X_2 are independent variables that summarize first relevant exposure within two disjoint intervals, and X_1 is logically derived in that $X_1 = Y$ if Y is not infinity, and $X_1 = X_2$ otherwise. Formally, we have that

$$p(Y = d \,|\, H) = \lambda(d) \prod_{t=T_1+1}^{d-1} \{1-\lambda(t)\} \quad (d = T_1+1, T_1+2, \ldots, T_2),$$

$$p(Y = \infty \,|\, H) = \prod_{t=T_1+1}^{T_2} \{1-\lambda(t)\},$$

$$p(X_2 = d \,|\, H) = \lambda(d) \prod_{t=T_2+1}^{d-1} \{1-\lambda(t)\} \quad (d = T_2+1, T_2+2, \ldots).$$

If $$Y \neq \infty, \quad p(X_1 = d \,|\, Y = d, X_2) = 1.$$

If $$Y = \infty, \quad p(X_1 = d \,|\, Y = \infty, X_2 = d) = 1,$$

which simply expresses that $X_1 = \min\{Y, X_2\}$. The numerical values for the exposure process are as in the one-drug model.

Nodes $S_1, E_1, O_1, S_2, E_2, O_2$: the individual drug models. These are as described for the one-drug model adapted to the particular times for taking each drug, but possibly using drug-specific assessments for the susceptibility σ and the delay m in losing susceptibility.

Nodes O, I_1 and I_2: the overgrowth process. Nodes O_1 and O_2 represent the potential for causing overgrowth; to model the actual overgrowth that leads to symptoms we need to define O as the minimum of O_1 and O_2. The indicator variables I_1 and I_2 then show which of the drugs was responsible for the overgrowth, allowing for the fact that it is feasible that both windows of susceptibility were open at the common time of relevant exposure $X_1 = X_2$; hence $O_1 = O_2$ and both drugs could be held equally responsible. Formally, we have that

$$O = \min\{O_1, O_2\}; \qquad I_1 = \begin{cases} 1 & \text{if } O = O_1, \\ 0 & \text{otherwise}; \end{cases} \qquad I_2 = \begin{cases} 1 & \text{if } O = O_2, \\ 0 & \text{otherwise}. \end{cases}$$

The conditional probability tables for O, I_1 and I_2 are then degenerate with ones in the appropriate positions.

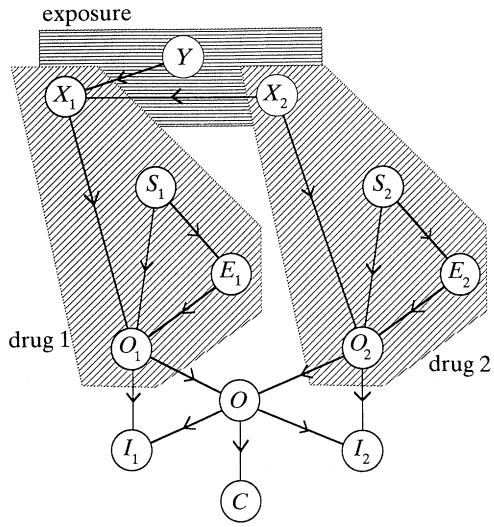

Figure 4. Graphical model of process when two drugs A and B have been administered, starting on days T_1 and T_2 respectively. Exposure: Y, day of first exposure unique to first drug (i.e. between T_1+1 and T_2 inclusive); X_1, day of first exposure relevant to first drug (i.e. after T_1+1); X_2, day of first exposure relevant to second drug (i.e. after T_2+1). First drug: S_1, day of start of susceptibility due to first drug; E_1, day susceptibility due to first drug ends; O_1, day of potential overgrowth due to first drug. Second drug: S_2, day of start of susceptibility due to second drug; E_2, day susceptibility due to second drug ends; O_2, day of potential overgrowth due to second drug. Causation: O, day of actual overgrowth; I_1, actual overgrowth caused by first drug; I_2, actual overgrowth caused by second drug. Case findings: C, day colitis appears.

Node C: the colitis symptoms. This is exactly as for the one-drug model.

We note that this representation allows a number of different interpretations for 'causation' to be simultaneously considered. If we observe colitis to have occurred on a particular day C, and then interrogate, say, I_1, we find the probability of causation by drug A in the sense that exactly the same event would have occurred had drug B not been given. Since $p(I_2 \mid C)$ gives the probability that the exact adverse event would have occurred if B alone had been given, we have that $p(I_1 \mid C)+p(I_2 \mid C)-1$ is the probability of 'joint causation', in the sense that at first relevant exposure both drugs would have led to the window of susceptibility being open. Thus we can decompose the blame into three components:

$1-p(I_2 \mid C)$: the probability that A alone caused the reaction,

$1-p(I_1 \mid C)$: the probability that B alone caused the reaction,

$p(I_1 \mid C)+p(I_2 \mid C)-1$: the probability of joint causation.

Alternatively, the probability that O_1 is non-zero represents the chance that drug A would have led to colitis at some point in the future, even if the actual overgrowth observed was due to drug B. Finally, we could examine the probability of S_1, which represents the event that drug A made the patient susceptible to overgrowth, even if in this case the patient was fortunately not exposed in the appropriate interval. Each of these events could be argued to be relevant to 'causation', and an advantage of a realistic model is that a unique definition need not be sought.

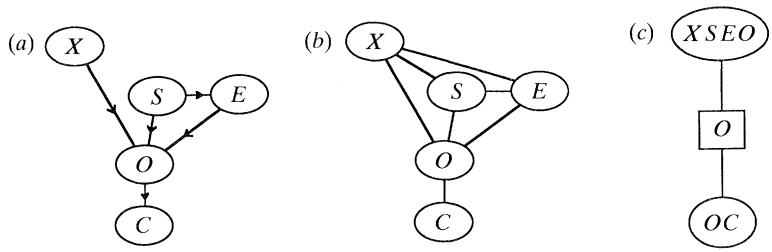

Figure 5. States of graphical re-representation for the one-drug problem. The original network (*a*) is turned into a moral graph (*b*) by adding links between co-parents, and dropping directions. Since this graph is already triangulated, its cliques can be arranged as a clique-tree (*c*).

6. Computational techniques based on graphical models

Having formally specified the joint distribution as the product of terms (4), we need to be able to calculate efficiently quantities of interest, such as the marginal probability of colitis occurring, or the chance of causation by drug A given colitis is observed on a particular day. The representation (4) is not appropriate for such calculations, since it would not be feasible to perform the multiplication and store the full joint distribution in the computer memory. Pearl (1986) showed how, if the graph had the structure of a tree (no cycle in the undirected graph formed by ignoring the directions on the links), then the required results could be obtained using only 'local' computations without specifying the full distribution. Furthermore, there is no need to have an overall controller of the process; each node communicates autonomously with its neighbours in the graph. Problems arise when loops occur in the graph, as in our examples, and Lauritzen & Spiegelhalter (1988) showed that local computations can still be used provided a new representation is found for the joint distribution that is related to an undirected graph, derived from the original DAG, but with a specific property. This stage of graphical re-representation requires five steps:

1. Add an undirected link between all co-parents that are currently unjoined.
2. Drop all directions in the graph. (This forms the so-called 'moral' graph.)
3. 'Triangulate' the graph by filling-in sufficient additional links to ensure that there is no cycle of length 4 or more without a short-cut.
4. Identify the cliques of this triangulated graph as the maximal sets of nodes that are all neighbours of each other.
5. Join the cliques together as a tree, which has the special property that if a node v is contained in any two cliques C and D, then v is also contained in all cliques on the unique path between C and D in the tree.

Those unfamiliar with this area may be somewhat mystified by such a sequence of graph-theoretic operations, and we try to give a short probabilistic justification as we step through the examples: see, for example, Lauritzen & Spiegelhalter (1988), Jensen *et al.* (1990), Andersen *et al.* (1989), Shenoy & Shafer (1988) and Dawid (1991) for further theoretical exposition.

Figures 5 and 6 show these five steps of graphical re-representation carried out to the one-drug and two-drug model. After steps (1) and (2) we are left with an undirected graph whose cliques (sets of nodes who are all joint neighbours) comprise a node and its parents. Referring to expression (3) we see that the joint density is a product of terms defined on this clique set. We therefore have a Markov random field defined on the moral graph, and we may use the independence properties derived

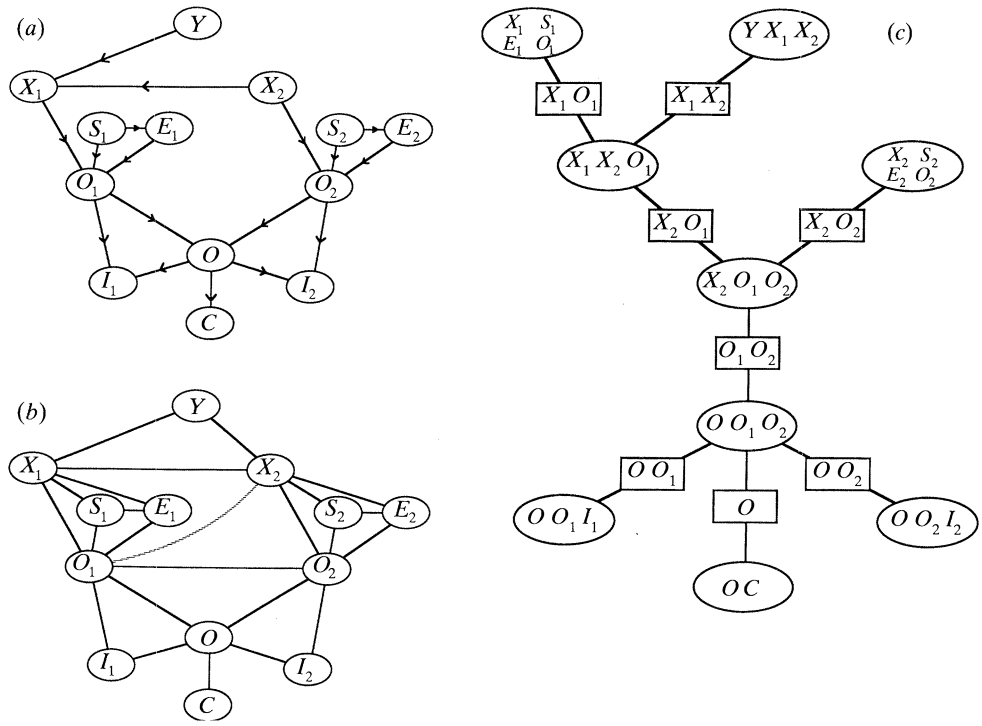

Figure 6. Graphical re-representation of the two-drug model (*a*). In this case an additional link is required to triangulate the moral graph (*b*), which then allows the cliques to be arranged as a clique-tree (*c*).

from undirected graphs that are more usually exploited in applications such as image processing (Isham 1981). More relevantly, there is a growing interest in graphical representations for contingency table models (Edwards & Kreiner 1983). Darroch *et al.* (1980) show how decomposable models (those whose joint distributions can be expressed wholly in terms of the marginal distributions on the cliques) correspond to triangulated graphs in the sense described above, and step (3) (filling in additional links to make triangulated) ensures that from now on it will be sufficient to derive the appropriate clique marginals; essentially we have embedded our original structure in one that is more complex, in that some of the conditional independence assumptions are no longer apparent from the graph, but allows much simpler analysis.

The moral graph of the one-drug model requires no triangulation, whereas the two-drug moral graph has a cycle $\{X_1, X_2, O_2, O_1\}$. To short-circuit this cycle a single link has to be introduced, which could be either X_1–O_2 or X_2–O_1 as chosen.

Triangulated graphs have been extensively explored in the relational database literature (Tarjan & Yannakakis 1984) and it is known that the cliques (step (4)) of such a graph have the powerful property that a join-tree, as defined in step (5), always exists. This means that when evidence arrives concerning any particular node, its implication can be propagated to any other node through the tree without any global supervision to prevent inconsistencies; each clique can act autonomously in receiving and passing on 'messages' from its neighbours. The clique tree for the two-drug model illustrates this property well: node O_1 occurs in five cliques, but these form a chain in the tree. Each clique in the triangulated graph forms a node in

the clique tree, and the separator sets are the clique intersections through which the messages pass. We note that each of the separator sets also separates the original directed graph into two or more components.

Before considering the exact form of the messages, we need to complete the next stage in getting the network ready to receive evidence; whereas the preceding stage involved the qualitative graph, we now need to consider the quantitative aspects. We have seen how the joint distribution was originally expressed as a product of functions on nodes and their parents, and we now need to express it as a product of functions ψ and ϕ defined on the cliques and their separators respectively. These functions are called potentials and we shall see below how this representation facilitates the required computations. Specifically, we require a joint distribution expressed as

$$p(V) = \prod_{C \in \boldsymbol{C}} \psi(v_C) \Big/ \prod_{S \in \boldsymbol{S}} \phi(v_S), \tag{5}$$

where $\boldsymbol{C}$ and $\boldsymbol{S}$ are the sets of cliques and separators respectively, and v_C and v_S indicate nodes in sets C and S respectively. Now, since the model is decomposable we have that a particular potential representation takes the form

$$p(V) = \prod_{C \in \boldsymbol{C}} p(v_C) \Big/ \prod p(v_S), \tag{6}$$

where p indicates the marginal distributions on the cliques and separators. From this representation the marginal distributions on any single node v may be easily obtained from the distribution on any clique which contains v. The essence of our approach is a general algorithm for going from any potential representation (5) to a clique marginal representation (6); once this is available we need only show that the representation (5) is not only easily obtained from the original conditional probability form (3), but also that the form (5) is retained when evidence on any nodes is elicited. We shall first show that the potential representation (5) is present at initialization and is retained under conditioning, and then describe the propagation algorithm for deriving (6) from (5).

Comparison of expressions (3) and (5) shows that the potential representation is straightforward to achieve at initialization: we need only set the original clique potentials (v_C) as the product of conditional probability tables corresponding to any node which, with its parents, lies in the clique. For example, in the one drug model the clique potential on $\{X, S, E, O\}$ is initialized to be $p(X)\,p(S)\,p(E\,|\,S)\,p(O\,|\,X, S, E)$, while in the two drug model the potential on the clique $\{O_1, O_2, O\}$ is set to $p(O\,|\,O_1, O_2)$. For many cliques, for example $\{X_2, O_1, O_2\}$, the initializing potential function can be considered to be unity since no node and its parents lie in that clique. All potentials $\phi(v_S)$ on clique separators are set to unity.

We now have a joint distribution expressed in terms of functions on cliques and their separators, and the next step is to show that this representation will hold whatever evidence is received on the network, i.e. for any observation $W = w$, $p(V \backslash W\,|\,W = w)$ can be expressed in the form (5). This is not difficult to see, since we always have $p(V \backslash W\,|\,W = w) \propto p(V \backslash W, W = w)$; thus to absorb evidence a clique containing the node that has been observed is selected and any potential not defined on the observed value is set to zero. This may occur at various parts of the network before propagation takes place.

We now consider the propagation algorithm. First, a root clique is arbitrarily

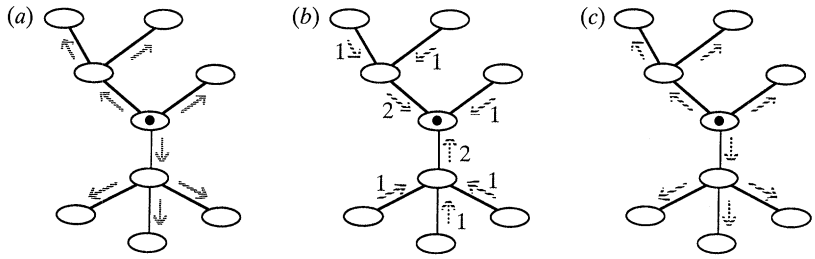

Figure 7. Sequence of message-passing in clique tree for two-drug model. The root-clique sends out a message to 'collect evidence', and then evidence is propagated back to the root. Finally, the collected evidence is distributed through the tree, giving the correct clique posterior distributions which can be marginalized to their constituent nodes.

selected in the clique-tree. It sends out a message 'collect evidence' to its neighbours, who communicate it to their neighbours until it reaches the ends of the tree. Each clique then collects evidence from those neighbours further from the root clique, using the fundamental operations shown below for passing a message from clique C to clique D through separator S.

(i) Marginalize $\psi(v_C)$ to $\phi'(v_S) = \Sigma_{v_{C\setminus S}} \psi(v_C)$.
(ii) Multiply $\psi(v_D)$ pointwise by $\phi'(v_S)/\phi(v_S)$.
(iii) Replace $\phi(v_S)$ by $\phi'(v_S)$.

The process of message-passing is shown in figure 7, which emphasizes that in collecting evidence each clique must await the messages from its more distant neighbours before passing on towards the root. When finally the root clique has collected evidence from its immediate neighbours, it normalizes its new potential to add to unity; this is the correct marginal distribution for the root clique. It then distributes evidence, which involves the same message-passing operation but working back through the tree. When this is complete all cliques and separators will hold their correct marginal distributions from which the current distribution for any node can be easily derived. The algorithm described above is a simplification of that described in Lauritzen & Spiegelhalter (1988), and proofs that these operations lead to this conclusion may be found in Jensen *et al.* (1990); Shenoy & Shafer (1988) provide an axiomatic basis for a general propation scheme, while Dawid (1991) provides numerical examples and generalizes the algorithm to solve many related problems on networks.

The above procedure allows for many facilities, such as easily retracting the items of evidence, exploring influential observations, and planning future questions (Lauritzen & Spiegelhalter 1988). Moreover, some simplifications are possible within specific circumstances. For example, once an initial clique marginal representation has been obtained, the effect of single items of evidence may be propagated using only distribute evidence and hence only a single pass through the network. Also, if only certain nodes of the network are of interest, it may be only necessary to propagate single items of evidence the necessary distance; for example, in the two-drug network, from a clique containing C to the first cliques containing I_1 and I_2. Finally, we could perform sensitivity analyses by varying the quantitative assessments along plausible ranges in order to calculate intervals for the probabilities of interest.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

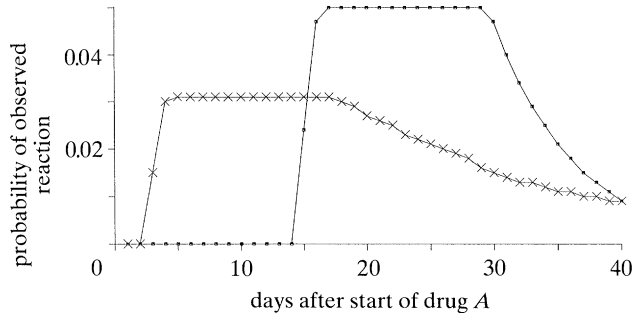

Figure 8. For each drug given considered singly, the probability of colitis occurring on each day, given it occurs at some time. ×, Drug A alone; □, drug B alone.

7. Examples

Our first example was introduced in §3. For illustrative purposes, we assume the two drugs have some distinct characteristics: while drug A is as described in §4, with susceptibility risk $\sigma_A = 0.9$ and susceptibility stop-rate $m_A = 0.05$ per day, drug B has susceptibility risk $\sigma_B = 0.2$ and susceptibility stop-rate $m_B = 0.15$ per day. Otherwise the parameters are taken to be those given in §4. Each drug is first considered separately with the one-drug model, from which we calculate that their prospective probabilities of causing PMC are 0.0058 and 0.0008, respectively, whose ratio 7.25 provides the prior odds on A against B causing a reaction (expression (1) in §2). Conditionally on causing a reaction at some currently unspecified time, the predictive distributions of C, the day of colitis, are shown in figure 8; the distributions are very flat until about a week after the drug is stopped, when they begin to tail off. Clearly, changing relevant parameters could shorten or extend this period. These conditional distributions have ordinates 0.019 and 0.050 for day 27, giving a likelihood ratio of 0.38 for Bayes's theorem (expression (2) in §2). Therefore, if we were to assume that one and only one of the drugs was responsible, our posterior odds on A against B given the case-specific evidence would be $0.38 \times 7.25 = 2.76$, or a probability of 0.73 for A. These posterior odds could also have been obtained directly as the ratio of the unconditional predictive probabilities of colitis occurring exactly on day 27 under each drug given singly.

However, a more realistic analysis exploits the two-drug model to allow for the possibility of joint responsibility. When we condition on knowing $C = 27$, this evidence propagates through the network using the techniques described in the previous section to provide posterior probabilities $p(I_1 \mid C = 27) = 0.862$ and $p(I_2 \mid C = 27) = 0.310$, which decompose into probabilities of 0.690 on A alone, 0.138 on B alone, and 0.172 on joint causation.

These conclusions may, however, be sensitive to information available on the hospitalization of the patient. Suppose, say, that it is known that the patient spent from days 5 to 10 in hospital, although no epidemic of *C. difficile* has been reported. Including this information in the two-drug model leads to the probabilities of causation shown in table 2. This very slightly shifts the blame to drug B. However, this is greatly magnified if we suppose that an epidemic has occurred in the hospital. This change is explained by examining the distribution on other nodes in the network that reveal, for example, that in these circumstances $p(S_1 \mid C = 27) = 0.0003$; i.e. that the first drug almost certainly did not induce susceptibility, even though it was of high risk of doing so. This becomes a more intuitive conclusion when one realizes

Table 2. *Probabilities of causation under different assumptions*

	drug A alone	drug B alone	both A and B
assuming no joint causation (two one-drug models)	0.73	0.27	—
two-drug model	0.690	0.138	0.172
hospitalized days 5–10	0.689	0.139	0.172
exposed to epidemic days 5–10	0.0003	0.9997	0

that, if drug A had induced susceptibility, then in such a high-risk environment overgrowth and colitis would almost certainly have occurred very much earlier than was actually observed.

8. Discussion and conclusions

The models described above have been programmed in C on a personal computer (Cowell *et al.* 1991), and the program allows interactive examination of the effect of changing the context and assumptions concerning the relevant parameters that influence the conditional probability tables. The program is currently being evaluated by the drug surveillance unit of a pharmaceutical company and, although no formal evaluation has taken place, examples such as that above have been presented to both expert assessors of ADRs and to microbiologists, who felt that it mirrored their considered judgement. The aim of such systems is to formalize and make available the expertise of the microbiologist in the routine assessment of cases, and so a formal peer review study is one form of evaluation. A more statistical criterion is whether or not the predictions made by the model accurately reflect the risks of developing the disease, and it is clear that some additional factors will need to be taken into account to make the model adequately realistic.

One issue is blocking, as mentioned in §2, in which a drug may induce susceptibility but might still prevent overgrowth while it is being taken. The tendency for this to occur depends on both the specific drug and the route of administration, with a higher chance of blocking if given intravenously rather than orally. This phenomenon has the effect of making irrelevant any exposure to *C. difficile* while the drug is being taken, and so is easily handled within our framework by introducing a 'blocking' node as a direct influence on the exposure node X. The chance of the blocking node being positive is influenced by the drug and route of administration, and if negative has no effect on X, while if positive shifts to zero the exposure rate (d) for d between day T of starting the drug and day W of withdrawal. The effect of introducing this node is to adjust our probabilities of causation for the possibility that one or other of the drugs was blocking overgrowth.

Adequacy of the model's predictions depends both on the qualitative structure, about which we feel fairly confident, and the assessed parameters, which may be open to more argument. This naturally brings us to the crucial issue of learning about the parameters as data on a number of cases accumulate. A parameter of vital interest to pharmaceutical companies is the susceptibility rate σ, whose size is the main determinant of the safety of the drug with respect to this particular reaction. If we acknowledge the uncertainty concerning this parameter, then it is possible to consider σ a random quantity and include it in the graph as a parent node of S, the day of susceptibility. Spiegelhalter & Lauritzen (1990) describe how one can then

process a case using the current expected value of σ, which is essentially the current 'best guess', but then revise our belief about σ after case evidence has been obtained. This revised belief is then carried over to the next case when a new estimate of σ is used for processing.

While this sequential updating process is attractive and is immediately appropriate in contexts such as clinical diagnosis, there are a number of problems with using it in the context of adverse reactions. The first problem is that we may not have independent confirmation of the cause of the reaction, and cannot be definitive that the drug caused susceptibility. While the formal bayesian updating procedure should still update the susceptibility parameter to the appropriate extent, it might be rather unsatisfactory to rest the condemnation of a drug on a set of cases in which it had not definitely been found to be at fault. The second, and perhaps more important problem, is the fact that a drug surveillance unit only hears of a highly selected sample of cases, and that the reporting process could effect our opinion concerning a particular drug. We are currently exploring the possibility of modelling the reporting procedure, with the aim of examining the sensitivity of our conclusions to reporting biases. One way in which a drug might be naively condemned due to circumstantial evidence is when two drugs are routinely given in the same order, and the second drug then will generally appear to have been more likely to have been the causal agent, even if they had identical susceptibilities. We are confident that a formal learning procedure will not be misled into systematically increasing the susceptibility parameter of the second drug.

The models described in this paper have been developed incrementally, and increasing realism has been accompanied by inevitable additional complexity. However, we have been impressed by the way in which the original network formalism has been able to adapt in a modular fashion to these demands, and we feel confident that problems such as those outlined above can be incorporated into the formalism. In addition, we are now attempting to model other classes of reaction along similar lines. While laying no claim as to the absolute correctness of either the qualitative or quantitative components of the models, we hope that their explicitness points the way to a more coherent analysis of many classes of adverse drug reactions.

We are indebted to Douglas Burdon and Sue Roden for their contribution to the project. Stephen Senn made many valuable criticisms. Financial support was provided by Glaxo Group Research Ltd and the Science and Engineering Research Council.

References

Andersen, S. K., Olesen, K. G., Jensen, F. V. & Jensen, F. 1989 HUGIN – a shell for building Bayesian belief universes for expert systems. In *Proceedings of 11th International Joint Conference on Artificial Intelligence, Detroit*, pp. 1080–1085. San Mateo: Morgan Kaufman.

Andreassen, S., Hovorka, R., Benn, J., Olesen, K. G. & Carson, E. 1991 A model-based approach to insulin adjustment. In *Proc. 3rd Eur. Conf. on Artificial Intelligence in Medicine, Maastricht*.

Andreassen, S., Jensen, F. V., Andersen, S. K., Falck, B., Kjaerulff, U., Woldbye, M., Sorensen, A. R., Rosenfalck, A. & Jensen, F. 1989 MUNIN – an expert EMG assistant. In *Computer-aided electromyography and expert systems* (ed. J. E. Desmedt), pp. 255–277. Elsevier.

Beinlich, I. A., Suermondt, H. J., Chavez, R. M. & Cooper, G. F. 1989 The ALARM monitoring system: a case study with two probabilistic inferencing techniques for belief networks. In *AIME '89: Lecture Notes in Medical Informatics vol. 38* (ed. J. Hunger, J. Cookson & J. Wyatt), pp. 247–258. Berlin: Springer-Verlag.

Cowell, R. G., Dawid, A. P., Hutchinson, T. A. & Spiegelhalter, D. J. 1991 A Bayesian expert system for the analysis of an adverse drug reaction. *Artif. Intell. Med.* **3**, 257–270.

Darroch, J. N., Lauritzen, S. L. & Speed, T. P. 1980 Markov fields and log-linear models for contingency tables. *Ann. Statist.* **8**, 522–539.

Dawid, A. P. 1991 Applications of a general propagation algorithm for probabilistic expert systems. *Statist. Computing*. (In the press.)

Edwards, D. & Kreiner, S. 1983 The analysis of contingency tables by graphical models. *Biometrika* **70**, 553–562.

Hutchinson, T. A. & Lane, D. A. 1989 Assessing methods for causality assessment of suspected adverse drug reactions. *J. Clin. Epidemiol.* **42**, 5–16.

Hutchinson, T. A., Dawid, A. P., Spiegelhalter, D. J., Cowell, R. G. & Roden, S. 1991*a* Computer aids for probabilistic assessment of drug safety. I. A spread-sheet program. *Drug Inf. J.* **25**, 29–39.

Hutchinson, T. A., Dawid, A. P., Spiegelhalter, D. J., Cowell, R. G. & Roden, S. 1991*b* Computer aids for probabilistic assessment of drug safety. II. An expert system. *Drug Inf. J.* **25**, 41–48.

Isham, V. 1981 An introduction to spatial point processes and Markov random fields. *Int. statist. Rev.* **49**, 21–43.

Jensen, F. V., Lauritzen, S. L. & Olesen, K. G. 1990 Bayesian updating in recursive graphical models by local computations. *Comp. Statist. Q.* **4**, 269–282.

Karch, F., Smith, C., Kerzner, B., Mazullo, J., Weintraub, M. & Lasagna, D. 1976 Adverse drug reactions – a matter of opinion. *Clin. Pharmac. Therap.* **19**, 489–492.

Lane, D. A. 1989 Subjective probability and causality assessment. *Appl. stoch. Models Data Analysis* **5**, 53–76.

Lane, D. A., Kramer, M. S., Hutchinson, T. A., Jones, J. K. & Naranjo, C. 1987 The causality assessment of adverse drug reactions using a Bayesian approach. *Pharm. Med.* **2**, 265–283.

Lauritzen, S. L. & Spiegelhalter, D. J. 1988 Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Jl R. statist. Soc.* B **50**, 157–224.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. & Leimer, H.-G. 1990 Independence properties of directed Markov fields. *Networks* **20**, 491–505.

Pearl, J. 1986 Fusion, propagation and structuring in belief networks. *Artif. Intell.* **29**, 241–288.

Pearl, J. 1988 *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.

Proceedings of DIA Workshop 1986 The future of adverse drug reaction diagnosis: computers, clinical judgement and the logic of uncertainty. *Drug Inf. J.* **20**, 455–533.

Shafer, G. & Pearl, J. (eds) 1990 *Readings in uncertain reasoning*. San Mateo: Morgan Kaufmann.

Shenoy, P. P. & Shafer, G. R. 1988 Axioms for probability and belief-function propagation. *Uncertainty in artificial intelligence 4*, pp. 169–198. Amsterdam: North-Holland.

Spiegelhalter, D. J. & Cowell, R. G. 1991 Learning in probabilistic expert systems. In *Bayesian statistics 4*. Valencia University Press.

Spiegelhalter, D. J. & Lauritzen, S. L. 1900 Sequential updating of conditional probabilities on directed graphical structures. *Networks* **20**, 579–605.

Szolovits, P., Patil, R. S. & Schwartz, W. B. 1988 Artificial intelligence in medical diagnosis. *Annals Int. Med.* **108**, 80–87.

Tarjan, R. E. & Yannakakis, M. 1984 Simple linear time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Computing* **13**, 566–579.

Thompson, E. A. 1986 Genetic epidemiology: a review of the statistical basis. *Statist. Med.* **5**, 291–302.

Turner, W. M. 1984 The Food and Drug Administration algorithm. *Drug Inf. J.* **18**, 259–266.

Venulet, J. 1984 The Ciba-Geigy approach to causality. *Drug Inf. J.* **18**, 315–318.

Wright, S. 1934 The method of path coefficients. *Ann. math. Statist.* **5**, 161–215.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES