Subjective Probability and Causality Assessment

by

**David A. Lane
University of Minnesota
Technical Report No. 507
November 1987**

Key words: subjective probability, causality assessment, adverse drug reaction, accessible probability evaluation, coherent decomposition

This work has been supported by grants from the American Medical Association Education and Research Foundation and the John Simon Guggenheim Memorial Foundation.

1. INTRODUCTION

This paper outlines an approach to a particular inferential problem that arises in the public health management of marketed drugs. The ideas that underlie this approach can be applied to a wide class of other problems, arising in medicine, engineering, business and public policy. The characteristics that these problems share can be described as follows: A panel of experts is assembled to recommend an appropriate course of action to some decision-maker; which action is appropriate depends in part on the experts' prediction (or retrodiction) about what will (or did) occur with respect to some phenomenon of interest. There are many different streams of evidence that affect the experts' judgement, relating to different factors and sources of information, involving different modes of knowing, from "hard" data to scientific theory to hunches based on professional lore and personal experience. These streams of evidence may point in different directions. The experts must evaluate the evidentiary significance of each of the streams, and then they must integrate their evaluations into an answer to the problem of primary inferential interest. How are they to proceed?

The challenge is to design a procedure that satisfies two fundamental dicta, articulated by two of the greatest inferentialists of this century: Fisher's "never throw information away!"¹, and de Finetti's "don't contradict yourself!"² According to the Fisher dictum, all the evidence available to the experts should be incorporated into their overall solution, including the experts' opinions about each item's relative relevance and plausibility. And according to de Finetti's dictum, all these evaluations should be merged in a way that satisfies a reasonable set of desiderata for a logic of reasoning under uncertainty.

The procedure described in this paper attempts to implement Fisher's dictum through the construction of an explicit, shared knowledge base incorporating elements from all the various relevant modes of knowing. De Finetti's dictum is implemented by using subjective probability to express the experts' collective uncertainty and then relying on de Finetti's theory of coherent inference to guarantee against self-contradiction as the different streams of evidence are evaluated and merged. Since the measurement of uncertainty in the probability scale can be quite unreliable, as Tversky and his coworkers have documented (see Kahneman, Slovic and Tversky (1982)), an attempt has been made in the work described here to differentiate

¹ Paraphrased, for example from Fisher (1973), p. 113.

² The theme of de Finetti (1972) -- and most of his other works.

between the propositions that experts in the field can measure reliably -- here called accessible -- and those they cannot, and to elicit assessments from experts only with respect to the former and then deduce, through the rules of coherence, those evaluations of the latter that are necessary to the problem at hand.

The next section introduces the problem of causality assessment for adverse drug reactions. The problem is described in terms of its role in postmarketing surveillance; it also comes up in clinical decision-making, setting editorial policies about the publication of reports of adverse effects in medical journals, and liability litigation (see Venulet et al. (1982) and Herman (1984)).

2. CASE REPORTS AND CAUSALITY ASSESSMENT

Serious adverse reactions to a drug usually occur too rarely to be identified in the clinical trials required to demonstrate efficacy before the drug can be put on the market. Instead, they are generally first encountered in the uncontrolled world of everyday clinical practice, and the industrial and national regulatory agencies that are responsible for drug safety must rely for their first indications of a possible drug-adverse event connection on case reports submitted to them by practitioners who observe an occurrence of the event in one of their patients taking the drug.

A typical situation in the work of these agencies finds a group of experts assembled around a table reviewing a small series of case reports (perhaps as small as one!) that link a particular drug with a particular type of adverse event. The first item on the agenda is to determine whether the link is just coincidental -- or, alternatively, whether the drug can in fact cause the event to occur. The experts tackle this question by trying to determine whether, for each case in their series, the available evidence indicates that the drug did cause the particular adverse event in question to occur. This problem is called causality assessment in drug safety circles, although a more accurate name would be retrodictive causality assessment (RCA), since the problem begins with a particular case in which the event has already occurred and looks backward in time to discover the event's unobservable cause.

If the experts have determined that the drug can in fact cause the event to occur, the focus of interest of the drug safety agencies shifts to the second agenda item, which could be called the predictive causality assessment problem: what is the chance that the next patient who takes the drug will subsequently experience the adverse event? This is the critical question for most clinical and regulatory decisions about the use of the drug, since the evaluation of net benefit of drug use that should guide such decision-making depends on its answer.

The purpose of this paper is to discuss a procedure, based on the use of subjective probability, to carry out the expert panel's retrodictive causality assessments. The paper proceeds as follows. First, some of the difficulties of retrodictive causality assessment will be illustrated, using a case of fatal hepatitis that was submitted to the Centre de Pharmacovigilance Bordeaux-Aquitaine, one of the regional agencies responsible for monitoring adverse drug reactions in France. Next, the probabilistic causality assessment procedure will be described and applied to this case. This procedure will then be compared to some alternative possible methods, and some advantages of the procedure emphasized. Finally, some projects for developing the procedure will be briefly mentioned, including two applications that relate the results of probabilistic RCA to probabilistic predictive causality assessments.

3. WHY RCA IS DIFFICULT

RCA problems are difficult for two reasons: there is a lot that experts know that is relevant to a particular problem -- and there is a lot that they do not know. The former results in a proliferation of different factors and sources of information that have to be taken into account, and so causality assessment problems are complicated, while the latter requires the assessors to take account of the extent of their uncertainty in order to obtain valid assessments. More specifically, here are some particular sources of complication and uncertainty:

(1) Profusion of possible confounders and effect modifiers: Almost all drug-associated adverse events have alternative possible nondrug etiologies, and there are rarely tests that can determine whether or not an adverse event is drug-caused. That is, for the most part, fulminant hepatitis is fulminant hepatitis, whether the agent that initiates the pathological process is a virus, a drug or a toxic agent like carbon tetrachloride or alcohol. Most drug-associated adverse events can be caused by many different drugs, so even if it is possible to determine that an event is drug-caused, it can be difficult to decide which drug is responsible when the patient is taking more than one. In addition, there are usually many possible factors -- genetic, behavioral, medical -- that can affect a patient's risk for experiencing the adverse event, from possible drug as well as nondrug causes. In the presence of all these confounders and effect modifiers, it is difficult to sort out what is affecting what, especially when all the relevant clinical information comes from the uncontrolled world of everyday practice.

(2) The ambiguity of "third party" reports: Case reports are submitted by busy practitioners who may not be experts in adverse drug reactions. The experts who analyze the reports do not see the patients whose cases they are trying to understand,

and they cannot control the nature or the quality of the information they review. From their point of view, critical information may be missing, difficult to interpret, or even erroneous.

(3) The complexity of case information: While there is rarely a pathognomonic test that can conclusively solve a causality assessment problem, there are generally many different pieces of case information that can help differentially diagnose the cause of the event: the timing of the event in relation to administration of the drug, the clinical characteristics of the event as it develops through time, and, when they occur, the response to withdrawal (dechallenge) and readministration (rechallenge) of the drug. Often the evidence offered by these different pieces of information points in different directions.

(4) The dispersion of expertise: Academic medicine is not organized around the problem of adverse drug reactions. Many different disciplines and subspecialties can provide critical background information for solving causality assessment problems: laboratory pharmacology, clinical pharmacology, epidemiology, and the "organ specialties" (nephrology, hepatology, and so forth) dealing with the locus of the disease which the suspect drug(s) are supposed to treat and the adverse event itself. But the relevance of what they know to the study of adverse drug reactions may not be obvious to these specialists, unless the proper questions are put to them by workers whose special expertise is the adverse effects of drugs.

I will now illustrate some of these difficulties with a case of fatal hepatitis following the use of the drug amodiaquine as a malarial prophylactic. In the early '80's, amodiaquine was increasingly replacing chloroquine as the drug of choice for Europeans travelling to malarial endemic areas, because of the development of chloroquine-resistant malarial parasites. In this period, the French Centres de Pharmacovigilance and the drug's manufacturer began to receive reports of hepatitis occurring among travellers taking amodiaquine; several such reports had been received when this one arrived at the Centre de Pharmacovigilance Bordeaux-Aquitaine.

CASE SUMMARY

The patient is a 38 year-old French woman who had lived in Gabon since 1981. From 1981 to 1983 she received chloroquine for malaria prophylaxis. On December 15, 1983 (Day 0), she was taken off chloroquine and prescribed amodiaquine 600 mg once per week. Her subsequent clinical course is as follows:

Day 36: Patient tired, vomiting.

Day 48: Jaundice, without fever or arthralgia. AgHBs negative.

Day 50: HEPATIC TESTS: bilirubin 5 x normal; ASAT, ALAT 15, 25 x normal.

DECHALLENGE (patient taken off amodiaquine).

Day 60: Patient felt better, jaundice decreasing. No hepatic testing done.

Day 67: RECHALLENGE (patient received 600 mg amodiaquine).

Day 70: Jaundice and pruritis noted. HEPATIC TESTS: bilirubin 18 x normal; ASAT, ALAT 20, 25 x normal.

Day 82: Patient's condition had so worsened that she was sent back to France. There she was noted to have severe jaundice, without sign of encephalopathy. She was given 100 mg of chloroquine. IMMUNOLOGICAL TESTS: AgHBs, antibody HBC, HVA (IgM) -- all negative.

Day 83: Encephalopathy (grade 1 coma with asterixis). HEPATIC TESTS: bilirubin 20 x normal; ASAT, ALAT 22, 25 x normal.

Day 84: Grade 3 coma with severe respiratory disorder.

Day 87: Patient dies. Postmortem biopsy showed almost complete hepatic necrosis.

Did amodiaquine cause this woman's hepatitis? Almost all of the forty or so physicians comprising the five group of experts with whom I have reviewed this case expressed a strong initial opinion that it did. This opinion is based on the fact that the hepatitis seemed to abate when the amodiaquine was withdrawn and reappeared in fulminant form shortly after readministration. Typically, positive responses to dechallenge and especially to rechallenge are regarded as very strong evidence in favor of drug causation.

But in fact the situation is more complicated than that. First, viral hepatitis is endemic in Africa, and European residents face very high risks of contracting it, a risk that continues to increase over their first several years of residency. Clearly, it is important to consider how high is this risk, compared to the magnitude of the risk for drug-induced hepatitis facing users of amodiaquine. Second, while the evidence from

the immunological testing tells against hepatitis A and hepatitis B, it offers no evidence against nonA-nonB (NANB) viral hepatitis; and NANB is often characterized by a "waxing and waning" course, so that a one or two month delay from onset until fulminance with at least subjective, symptomatic improvement in the middle of this period is not at all an unlikely course for NANB. How unlikely, compared to what would be expected in response to dechallenge and rechallenge for drug-induced hepatitis? Consider that amodiaquine has a long half-life (around a week); would it be reasonable to expect that drug-induced hepatitis would begin to abate in a week and a half? Third, was there really an improvement in the patient's underlying liver disease by day 60? Did she even see her physician on that day, or merely report to him her own impression that she was "feeling better" and that her jaundice was decreasing? The reports gives no information on this point, but the fact that no hepatic testing was done suggests that the patient might not have actually visited her doctor. And the bilirubin value on day 70 is strongly suggestive that the hepatitis had in fact not abated in response to dechallenge: can bilirubin levels increase from less than 5 to 18 x normal in just three days?

These are just a few of the questions that make this case difficult to analyze. Notice that the first point requires epidemiological expertise, the second calls for input from pharmacology and clinical hepatology, while the third is primarily a question of theoretical physiology!

4. A PROBABILISTIC PROCEDURE FOR RCA

4.1 Formulating the Problem

The aim of the procedure is to evaluate the expert panel's *posterior odds* that a reported adverse event E was caused by a particular drug D:

$$(1) \quad \frac{P(D \rightarrow E|B,C)}{P(D \nrightarrow E|B,C)}$$

Here, P measures the panel's degree of belief, interpreted according to the economic metaphor introduced by de Finetti (see de Finetti (1972) and Lane (1981)). The proposition "D $\rightarrow$ E" (D caused E) means that E *would not have happened as and when it did had D not been administered*; "D $\nrightarrow$ E" denies D $\rightarrow$ E. B represents *background information*, including all the knowledge the panel can assemble relevant to the connection between the drug D and events similar to E (from their own clinical experience, collected case series and epidemiological studies, facts and theories from

pharmacology and other basic sciences). The only case-specific information in B is the proposition that a patient with a specified *clinical condition* M who has been administered drug D in a specified way subsequently develops an *adverse event of type* E_i (M and E_i are generic characterizations, explicitly defined by the panel as part of the assessment procedure, that describe essential elements of the patient's condition and the event E, respectively; together with D, they define a kind of *epidemiological reference set* of which the particular patient whose case is under review is a member). C represents case information: details about the particular patient and her adverse event E.

The panelists cannot apply their expertise directly to assess the posterior odds. There seem to be two types of propositions about which these experts can comfortably measure their uncertainty in the probability scale: propositions which bear a frequency interpretation that relates to verifiable events occurring in patient populations similar to those of their own clinical experience or that of their colleagues, reported personally or in the medical literature; or propositions that can be interpreted in terms of an acceptable theoretical model for the phenomenon to which the proposition refers, from which frequency predictions can be calculated.

The proposition " $D \rightarrow E$ " is not of these types. As with any counterfactual proposition, " $D \rightarrow E$ " is essentially unobservable: this patient was prescribed D before she experienced E, so that it is impossible to observe directly what would have happened had she not been administered the drug at this time. Moreover, in contrast to some other diagnostic problems in which additional confirming information can eventually result in certain knowledge about cause, it is rarely the case with adverse drug reactions that any useful case information beyond what is contained in the case report itself will ever be obtained. Thus, it is hard to accumulate clinical experience on "similar" cases that can give a direct interpretation to the proposition " $D \rightarrow E$ " in terms of the frequency of any observable event. And theoretical models for phenomena do not help, since they predict from cause to effect, not the other way around as required for the proposition " $D \rightarrow E$ ".

The strategy that underlies the probabilistic procedure for RCA is to decompose the posterior odds into components that involve propositions that are accessible to the knowledge and experience of the expert panelists. Then, the rules of probability theory are used to merge the solution to these component assessment problems into a solutions to the overall RCA problem. The implementation of that strategy begins with an application of Bayes' theorem, as follows:

$$(2) \quad \frac{P(D \rightarrow E|B,C)}{P(D \nrightarrow E|B,C)} = \frac{P(D \rightarrow E_t|B)}{P(D \nrightarrow E_t|B)} \cdot \frac{P(C|D \rightarrow E,B)}{P(C|D \nrightarrow E,B)}$$

posterior odds prior odds likelihood ratio

Both the posterior and prior odds are calculated conditionally on B and so refer to a patients with clinical condition M who has been administered D and experienced an event of type E_t . However, the identity of the patient to whom the two terms refer is different. In the posterior odds, it is the subject of the case report under review; while in the prior odds, it is a "generic" patient (perhaps the "next" patient) with the three defining characteristics M, D, and E_t . Thus, the prior odds can be regarded epidemiologically, as will be discussed below; on the other hand, the probabilities in the likelihood ratio involve thinking in terms of mechanism, arguing from cause to effect.

It is helpful to consider the information in C in chronological order. A typical chronological sequence is illustrated in Figure 1.

-- FIGURE 1 GOES HERE --

The categories of case information include *Hi*, the patient's history antedating the onset of E; *Ti*, the timing of the onset of E in relation to the administration of D; *Ch*, characteristics of the event from time of onset to time of dechallenge; which can include information about duration, severity, evolution, and laboratory tests; *De*, response to dechallenge; and *Re*, response to rechallenge. The likelihood ratio (LR) is decomposed into factors, corresponding to these chronological categories of case information:

$$(3) \quad LR = LR(Hi) \cdot LR(Ti) \cdot LR(Ch) \cdot LR(De) \cdot LR(Re),$$

where, for example,

$$LR(Ti) = \frac{P(Ti|D \rightarrow E, B, Hi)}{P(Ti|D \nrightarrow E, B, Hi)}$$

Note that the probabilities that appear in each likelihood ratio factor are evaluated conditionally on B and all chronologically preceding case information. The probabilistic approach to RCA is designed to calculate the posterior odds in favor of D-causation, by evaluating the subjective probabilities that appear in the prior odds and the likelihood ratio factors corresponding to the chronological categories of case information.

In the discussion that follows, the procedure will be illustrated with an assessment carried out by a panel of experts from McGill University and the University of Montreal. Members of the panel included Dr. Michael Kramer (Departments of Pediatrics and Epidemiology, McGill), Dr. Tom Hutchinson (Departments of Medicine and Epidemiology, McGill), Dr. Walter Bloom (Department of Medicine, McGill), Dr. Pierre Biron (Department of Pharmacology, Montreal), Dr. Pierre Toussignant (Department of Epidemiology, McGill), and Ms. Marie Mouchbahani (Department of Pharmacy, Royal Victoria Hospital, Montreal). I served as moderator for the panel and Ms. Patti Groome (Clinical Epidemiology Division, Royal Victoria Hospital) coordinated literature searches.

4.2 Implementing the Procedure I: Gathering the Evidence

The purpose of this phase of the procedure is to assemble and organize the background and case information conditionally on which the posterior odds is calculated. The work proceeds by a general discussion of the panel, guided by a moderator who poses a series of questions about *case parameters* and differentially diagnostic *case information*.

First, the panel assesses the values of four case parameters: M and E_t create an epidemiological reference set for the patient under consideration; the *cause list* specifies the alternative etiological candidates; and the *time horizon* determines a period of time to which all considerations about drug-event connection are restricted.

M and E_t : E_t specifies the general type of the adverse event E , and M abstracts out the most important aspects of the patient's condition that determine his risk for events of type E_t from causes other than the drug under consideration. It is important to define these parameters as explicitly as possible.

Sometimes, there will be a question about whether to include certain aspects of the patient's condition or the event in the definitions of these parameters or in the appropriate chronological category of case information. The choice should be guided by the ease of the ensuing assessments and so depends on the assessors' experience and information; roughly, the assessors should choose definitions for M and E_t that make it easiest for them to "think epidemiologically" about the class determined by these definitions. The essential thing is consistency in the course of the assessment: M and E_t are included in B , the background information, and are part of the reference set for *every* calculation.

Cause list: The cause list consists of a set of mutually exclusive propositions about the possible causes of E . Typically, there are a number of possible hypotheses

specifying drug causal candidates -- say $D_1, \dots, D_n$ -- followed by a number of hypotheses specifying nondrug causes.

The requirement that the propositions on the cause list be mutually exclusive is somewhat artificial and requires a bit of care in defining and interpreting the propositions. For example, suppose the assessors believe that an *interaction* between drugs D_A and D_B may have caused E . Then they should have a separate entry on the cause list for this interaction; and if they also believe that either drug alone may also have caused E , they must include entries for " D_A alone" (that is, " E would have happened as and when it did had D_A and not D_B been given, but not if D_B and not D_A had been given") and " D_B alone". Similarly, there is an asymmetry between drug and nondrug causes: because of the definition of the proposition " D caused E " for a possible *drug* cause D , if N is a possible *nondrug* cause, the proposition " N caused E " implies that " E would have happened as and when it did had *none* of the drugs mentioned in $D_1, \dots, D_n$ been given." Finally, since causality assessment focuses on the question of drug responsibility, it is possible to lump together different nondrug causes, as long as they give the same probability to each piece of case data that distinguishes drug from nondrug causation.

Time horizon: B includes the assertion that a patient with condition M experiences an event of type E_t . The time horizon puts an upper limit on the length of time after D -therapy begins in which it is asserted that the event occurs. To choose the time horizon, the assessors should think about the distribution for the onset of an event of type E_t that is caused by D ; the time horizon should be about as long as the support of this distribution. It is easy to see that, for coherent assessors, the exact value selected will not affect the resulting posterior odds. The reason for specifying a time horizon at all is that doing so increases the accessibility of some of the assessment tasks (especially the prior odds and likelihood ratio for timing).

The values for the case parameters assessed by the panel for the example case are given in Figure 2 below.

-- FIGURE 2 GOES HERE --

Next, the panel sorts out which of case data differentiate among the possible causes and lists these by chronological category. The evidentiary significance of each listed item is supported by heuristic arguments that prepare the way for the quantitative phase of the evaluation. Frequently, these arguments depend on conjectures that the panelists feel require support beyond what their own expertise provides, and so specific queries are formulated for literature reviews or elicitation of the opinion of outside experts. (The members of the panel themselves and the

research coordinator Ms. Groome elicited information and opinion from the relevant outside experts.)

The case information considered relevant in the preliminary review by the Montreal panel for the example case is shown in Figure 3 below. The discussions that culminated in the information as presented in Figures 2 and 3 required a single session of two hours.

-- FIGURE 3 GOES HERE --

4.3 Implementing the Procedure II: Pruning Information

The purpose of this phase of the RCA procedure is to carry out a semi-quantitative, order-of-magnitude analysis to eliminate from detailed consideration causes that will have very low posterior probability and case information that will not effectively discriminate between the causes with appreciable posterior probability. It is, after all, pointless to evaluate probabilities that will have no real effect on the overall causality assessment.

The first task is to distinguish between primary and secondary causes. Primary causes are at least two orders of magnitude more likely than secondary causes, and so the detailed quantitative assessment in the next phase of the procedure can be restricted to partitioning probability among the primary causes. Should the detailed assessments reveal that primary causes are in fact less well-supported (or suggest that certain secondary causes are in fact better-supported) than initially believed, secondary causes can be upgraded as necessary.

A useful tactic to determine that a given possible cause is secondary is to find a particular combination of factors or pieces of evidence that tell strongly against the cause relative to another possible cause, determine an order-of-magnitude bound for how strongly, and then check to see that no other factor or piece of evidence sufficiently favors the given cause to allow it to "catch up" with the test primary cause to within two orders of magnitude.

For example: compare chloroquine to amodiaquine with respect to timing and prior. According to the panel's judgement, timing is at least ten times more likely for amodiaquine, because "short onset" (within four months) is at least as likely as "long onset" (from 6 months to several -- say 5 -- years), and if each of these timing distributions is roughly uniform on the appropriate interval, the likelihood ratio between a short and long onset drug is greater than 15. The actual onset distributions, if they were carefully assessed, would tend to increase this ratio in favor of the short onset drug (but it was not obvious to the panelists that the ratio would be greater than

100, or two orders of magnitude). The prior also tells strongly in favor of amodiaquine, since chloroquine has been much more extensively used than amodiaquine (probably two or more orders of magnitude), with even less reports of suspected hepatic reactions. Thus the timing and prior taken together give a comfortable two order of magnitude cushion to amodiaquine over chloroquine, and no other evidence favors chloroquine at all, so it cannot "catch up."

In the example assessment, only two primary causal candidates survived the pruning: NANB and amodiaquine (hereafter D). The panel could not determine whether either dominated the other without a detailed analysis, carried out in the next phase of the analysis.

Case information can be dropped at this stage for two reasons. First, the literature searches and outside expert assessments may not confirm some of the conjectures on which the panel believed some data's evidentiary significance to rest. Second, data that may differentiate between some of the causes on the initial list may fail to differentiate between the causes that remain on the pruned list. Here is an example illustrating both these possibilities: based on their review of the literature on cases of viral and drug-induced hepatitis meeting the definition of E_i (based on drugs whose pathophysiologic mechanisms were regarded as similar to those entertained for amodiaquine, as well as previous cases of amodiaquine-associated hepatitis), the panelists assessed similar distributions for enzyme and bilirubin values for the two primary causes, D and NANB; these distributions did differ from the distribution of values associated with hepatitis B, but because of the immunological test results, this was now a secondary cause.

The pruned list of case information as determined by the Montreal panel is presented in Figure 4 below. This phase of their assessment took about one hour, in addition to the substantial time between meetings spent by the research assistant carrying out the required literature searches.

-- FIGURE 4 GOES HERE --

4.4 Implementing the Procedure III: Assessing the Evidence

This is the quantitative phase of the procedure. The panel must evaluate a series of probabilities that will determine their values for the prior odds, and they must assess their probabilities for each piece of case information on the pruned list, conditional on each primary cause and all chronologically preceding case information. These assessment problems will be discussed in reference to the example case below. First, I will comment on the process under which the assessments are carried out.

There are two essential steps in the assessment process: first, the propositions whose probabilities are to be assessed must be precisely formulated; next, an assessment for each of these propositions that represents the "shared expertise" of the entire panel must be generated. The goal of the formulation step is to produce propositions that are accessible to the knowledge and experience of the panelists. As stated above, we have found that our medical experts can think about propositions formulated in terms of the frequencies with which well-defined events occur among well-defined patient populations that the panelists can readily visualize, on the basis of their own clinical experience or that of their professional colleagues as described anecdotally or in the medical literature; or they can think about propositions referring to predictions derived from theoretical models for possible pathophysiological mechanisms. Thus, the moderator must set each proposition to be evaluated in one of these two contexts.

When panelists disagree in their evaluations for the probabilities of propositions formulated in these two ways, we have found that there are almost always differences in their underlying "knowledge bases" relative to the relevant clinical experience or posited mechanisms. Our approach to such disagreements in probability evaluations is to initiate discussions that explicate these "knowledge base" differences and allow the formation of a pooled knowledge base, in terms of which new propositions are formulated, whose probabilities are then related to the probabilities that are required for the RAC problem by the rules of coherence. For example, if additional possible mechanisms have been introduced, the law of total probabilities can be employed to merge the evaluations conditional on each mechanism in the augmented list; or if someone has contributed new information equivalent to the knowledge of additional relevant occurrence frequencies, Bayesian updating can be used. In our experience, this process of explicit knowledge-pooling generally produces probability assessments that represent a consensus of opinion among the panelists. If disagreements still persist that cannot be resolved in this way, the panel measures their collective uncertainty by means of a distribution around the probability in question.

4.4.1. Evaluating the Prior Odds

According to the definition of the prior odds (equation 2), its value is the panelists' answer to the following question: consider a class of patients with condition M who receive D and subsequently (within the time horizon) experience an event of type E_t ; in what proportion of these patients is the event caused by D? Alternatively, the same problem can be posed from a *prospective* point of view, as follows. Imagine a large

class of patients with M. Suppose half of them are selected at random and receive D, while the rest receive some alternative therapy with the same beneficial effects as D, but which cannot cause events of type E_t . Let $P(E_t|D)$ represent the panel's estimate of the proportion of the patients who receive D that experience events of type E_t , and $P(E_t|D^c)$, the corresponding estimate for the proportion of those who do not receive D. Then,

$$(3) \quad \text{prior odds} = \frac{P(E_t|D) - P(E_t|D^c)}{P(E_t|D^c)}$$

Thus, one strategy for assessing the prior odds is to use what information is available to evaluate the two quantities $P(E_t|D)$ and $P(E_t|D^c)$ as defined above, and then use equation (3) to determine the prior odds. As a result, assessing the prior odds calls for techniques for doing synthetic (or "meta-analytic") epidemiology: combining information from a variety of different sources to infer what the incidences would be in an imaginary "clinical trial". Usually, the resulting opinion about these incidences is sufficiently diffuse that it is best to assess subjective *distributions* for these quantities and then to evaluate the means of these distributions to plug into (3).

There are four kinds of information that can have an effect on an assessor's uncertainty about the incidence of an adverse event E following the use of D. First, there can be knowledge of possible pathophysiological mechanisms related to known actions of the drug, perhaps supported by data from animal toxicology experiments or premarketing studies on human subjects. Second, D may be related pharmacologically or chemically to other drugs, for which data bearing on adverse effects exist, and these related frequencies may suggest plausible values for the incidence of similar adverse effects of D. Third, it is possible to use the number of reports that have been received by the drug's manufacturer or to a national monitoring agency, together with estimates for the fraction of all cases that are reported to the agency and the number of patients in the population exposed to the drug to generate indirect frequencies that provide an estimate of the incidence of E following the use of D. Finally, once a suspicion about the D-E connection has crystallized, epidemiological studies may be carried out to investigate it. The direct frequencies obtained from these studies may have to be adjusted to take into account the different patient populations observed by the different studies, the different variables they measured, and the different research designs they employed. Each of these kinds of information can be converted into distributions for the relevant incidences; typically, mechanistic arguments produce the

most diffuse distributions, with each of the types of frequencies producing sharper distributions in the order in which they are listed. In particular, moderately extensive direct frequency information, when available, generally dominates everything else.

The Montreal panel's evaluation of the prior odds in the example case relied primarily on indirect frequencies to estimate $P(E_t|D)$ and direct frequencies (with substantial adjustment required) for $P(E_t|D^c)$. For succinctness, I will restrict to their point assessments rather than the uncertainty distributions for these quantities; the emphasis of this summary will be on the number of different factors and sources of information they had to take into account in arriving at their answers, and how the evidence from these factors and sources was amalgamated.

The panel estimated $P(E_t|D)$ as follows. At the time of their evaluation, French investigators from industry and the Centres de Pharmacovigilance were carrying out a national survey to identify all cases of liver disease following amodiaquine prophylaxis among Frenchmen who travelled abroad (Lenoir (1987)). Based on 22 ascertained cases, and estimating number of travellers exposed from prescription data, these investigators estimated that 1 in 15000 prophylaxis courses result in liver disease, of which perhaps 60% meet the definition of E_t , giving an incidence for ascertained E_t of 4 per 100,000 courses. How effective was their ascertainment? Given the high probability of hospitalization given E_t and the intensity of their search, the panel assumed that the ascertainment rate was better than 50%, probably close to 75%. Thus, their estimate for $P(E_t|D)$ was between 4 and 8 per 100,000, and they adopted 6 per 100,000 as their point estimate.

Next, they estimated the numerator of equation (3) by considering: how many of these cases were actually caused by D? They decided that most of the cases of liver disease occurring among travellers taking D were actually due to D, for the following reasons. First, the French investigators eliminated any cases that tested positively for hepatitis A or B or could be convincingly attributed to alcohol. Next, temporary travellers to regions where the prevalence of viral hepatitis is high are likely to take gamma globulin shots; other studies of travellers' hepatitis make it unlikely that the incidence for non-drug caused E_t could be much higher than around 1 per 100,000 per year, roughly the overall incidence of NANB among European adults, or about 3 per million over the four month time horizon. Thus, the panel's estimate for the incidence of E_t due to amodiaquine among French travellers taking the drug is 6 per 100,000.

Finally, the panel assumed that the rate of drug-induced liver disease should be comparable for Europeans taking the drug in Africa as for French travellers. Also, the available case series (from France and Switzerland -- Larrey et al. (1986), Neftel et al.

(1986)) gave no grounds for supposing that age and sex modify the risk for amodiaquine-induced liver disease. Thus, their estimate for the attributable incidence of amodiaquine-induced E_t for a patient with clinical condition M is 6/100,000.

To estimate $P(E_t|D^c)$, the panel needed information on the incidence of NANB among middle-aged European women resident in Gabon for 3 years. They could not obtain such information directly, so they constructed the following chain of reasoning based on what data they did find. First, they concluded that among cases of viral hepatitis in adults in Africa, about 20% are not due to hepatitis A and B (so they are due to what I am here calling NANB) -- and that this is for whites as well as blacks, though the data on this point is sketchy (Crocchiolo et al. (1984), Greenfield et al. (1984)). Among African cases of NANB, it is not generally possible to determine the mode of transmission as it is in Europe or America (so the fact that there is no mention in this case of transfusion, parenteral injection, drug use, etc., does not affect the estimated incidence for NANB).

The best information on incidence rates of viral hepatitis among white residents of Africa came from a study on hepatitis A and B among American missionaries and their families (Kendricks ()). These rates climb from years 1 to 3 of residence in endemic areas, then fall. For middle-aged women in their third year of residence in Africa, the panel estimated an attack rate of about 2 per 100. Based on their previous estimate that the incidence of NANB is 25% of the incidence for A and B combined, this yields an estimate of 5 per 1000 per year for NANB.

The panel supposed that missionaries and their families would have a higher rate of viral hepatitis than the typical European resident in Africa, because of greater exposure to the native population and generally lower living standards. How much higher? This question is critical for the assessment, and the panel had no data on which to base their answer; their guesses ranged from 2 to 5 times higher. These guesses yield an estimated rate of NANB for all middle-aged European women in their third year of African residence of 1 to 2.5 per 1000 per year. For a point estimate, the panel settled on 1.5 per 1000 per year. With a four month time horizon, they thus estimated an incidence rate of NANB within the time horizon of 5 per 10,000.

Finally, the panel had to determine the fraction of NANB cases that satisfied their definition of E_t . To answer this question, they used data on NANB presented in a survey paper and in several published case series (Dienstag (1983), Farrow et al. (1981), Kryger et al. (1980)). They concluded from this data that slightly less than half of all NANB cases have bilirubin levels less than 3 x normal and about 40% of these cases have enzyme values less than 5 x normal. Although no data was presented on

the joint distribution of these variables, the panel's internists claimed on the basis of their own experience that low values of bilirubin tend to be associated with low enzyme levels, and they estimated that about half of the NANB cases met the definition of E_t .

Thus, the panel came up with an estimate of 2.5/10,000 for $P(E_t|D^c)$, and so

$$\text{prior odds} = \frac{6/100000}{2.5/10000} = 0.24.$$

4.4.2 Evaluating the Likelihood Ratio Factors

To evaluate the likelihood ratio factors, the probability for each item on the pruned list of case information must be evaluated, given each possible cause and all preceding data. Even though only the probability for the data actually observed plays a role in these likelihood ratio factors, to ensure meaningful and reliable results, it is important to evaluate these probabilities in their proper context. For example, for information that refers to *timing* (time to onset of E, duration of E before dechallenge, time of disappearance of E after dechallenge), it is best to construct entire *timing distributions* for the relevant events, as illustrated in the first example below. Similar considerations hold for probabilities about the results of laboratory tests.

In general, the likelihood ratio factors are easier to evaluate than the prior odds, except in those rare situations in which data from large, well-designed epidemiological studies can be used to provide direct frequency estimates for all the possible causal candidates. The propositions that appear in the likelihood ratio factors are often directly accessible to an individual physician's personal clinical experience (when they refer to observable events that happen relatively frequently) or clearly related to mechanistic ideas that are generally shared by the relevant experts. I will now summarize two of the likelihood evaluations from the Montreal panel's analysis of the amodiaquine-hepatitis case.

LR(Ti): The panel had to assess the probability that the first prodrome (tiredness, vomiting) would begin in the fifth week after beginning amodiaquine therapy (Ti), as a function of cause. Since there is no relation between the onset of viral infection and drug use (given that the patient had already been in Africa for three years when she began taking D), the onset distribution given NANB-causation is uniform on the sixteen-week time interval determined by the time horizon, and so $P(Ti|NANB \rightarrow E) = 1/16$.

To the panel members, the onset distribution given D-causation depends on mechanism. Two possible mechanism types were considered: immunologic and dose-

related (metabolic). The panelists (and other adverse drug reaction experts) associate a standard distributional shape (unimodal, with a gradual right hand tail) with both these mechanisms; the appropriate time scale depends on the event type E_t , the dose schedule and pharmacologic properties of the drug D. To assess the distributions given mechanisms for this problem, the moderator queried the panel members first about the distribution's range (1-14 weeks for the immunologic distribution; 1 to 20 half-lives for the dose-related distribution), the mode (4-5 weeks; 9 to 10 half-lives), and concentration (50% chance of onset between 3.5 and 6 weeks for the immunologic distribution). The resulting distributions are sketched in Figures 5 and 6. There were very little differences among the panelist's individual assessments, and they were easily resolved in a short discussion. The probabilities that the first prodrome would begin in the fifth week of treatment assigned by the two distributions are 0.15 and 0.05, for the immunologic and dose-related mechanisms respectively.

-- FIGURES 5, 6 GO HERE --

To complete the evaluation of $P(T_i|D \rightarrow E)$, the panel had to assess its probability that the reaction is of each of the two posited types. Though this assessment task does not belong to either of the two classes of accessible assessments described above, the experts of this panel (and all the other ones with which I work) seem able to address it, by means of a kind of meta-frequency expert consensus: that is, there seems to be a generally shared opinion about what proportion of all drugs that can cause a particular event type E_t do so by each of a particular set of possible mechanisms. Some of these mechanism distributions are presented in standard texts on drug-induced disease. The Montreal panel initially favored an immunologic distribution 3:1 over a dose-related one. However, some of the data in the case series collected by the French investigators differentiates between the two mechanism-types: in particular, the onset times for the 9 cases in the series that were available to the Montreal group favored the dose-related mechanism over the immunologic one by a likelihood factor of 2.25, while the incidence of fever (more likely with an immunologic than a dose-related mechanism) among these cases slightly favored the immunologic mechanism (likelihood factor = 0.89). Overall, these data favor the dose-related mechanism by a factor of 2, giving posterior odds in favor of the immunologic mechanism of 1.5, corresponding to a posterior probability of 0.6 that the mechanism of amodiaquine-induced hepatitis is immunologic.

Thus, $P(T_i|D \rightarrow E) = (0.15) \bullet (0.6) + (0.05) \bullet (0.4) = 0.11$, and so

$$LR(Ti) = \frac{0.11}{1/16} = 1.8$$

LR(De-Re): For this case, the panel preferred to evaluate the data in De and Re out of chronological order. The data judged differentially diagnostic from the time of dechallenge until the patient's death can be listed as follows:

- (1) the hepatitis is fulminant, resulting in death from hepatic necrosis.
- (2) on day 70 (three days after rechallenge), liver function tests gave the following values: bilirubin 18 x normal; enzymes 20 - 25 x normal.
- (3) hepatic encephalopathy began on day 83.
- (4) on day 60 (10 days after dechallenge) "the patient felt better, jaundice decreasing."

The key to the panel's analysis of this data is the claim that, given data elements (1), (2) and (3) above, datum (4) is not differentially diagnostic. I will just sketch the argument that supports this claim, without reproducing the actual calculations. The basic idea is this: according to the liver experts consulted by the panelists (none of whom were themselves hepatologists), under most circumstances bilirubin can only increase by about 3 x normal per day; not only can a theoretical justification be given for this statement, but none of the experts consulted could ever remember having seen bilirubin increase by more than 4 x normal per day, absent renal failure. If the data element (4) really represents a "real" rather than a merely subjective positive response to dechallenge, then bilirubin must have dropped below 5 on day 60 -- and, since no more drug was administered until day 67, stayed below 5 until day 67. But then, according to data element (2), bilirubin would have had to increase over 13 x normal in just three days, an exceedingly unlikely event. Thus, a "real" positive response to dechallenge implies a nearly impossible "real" positive response to rechallenge. On the other hand, if the patient's improvement on day 60 is merely subjective, or the result of a "real" slight fluctuation in an otherwise progressing case of fulminant hepatitis, then the data element (4) fails to discriminate between D and NANB causation (or in fact slightly favors NANB, since this disease is frequently characterized by a relapsing course).

According to the panelists assessments, which depended to a large extent on data presented in several published case series of fulminant NANB (McNeil et al. (1984), Gimson et al. (1983), Mathiesen et al. (1980) as well as some general information on drug-induced liver disease, data element (1) (the fact that the patient died of hepatic necrosis) failed to distinguish between D and NANB, while the values of the liver

function tests on day 70 slightly favored NANB. However, data element (3), giving the time from onset to encephalopathy, favored D-causation, with a likelihood factor of 3.5.

Overall, then, for the example case, the Montreal panel assessed the value of LR(De-Re) as 3.5, and so

$$\begin{aligned}\text{posterior odds} &= \text{prior odds} \bullet \text{LR(Ti)} \bullet \text{LR(De-Re)} \\ &= (0.24) \bullet (1.8) \bullet (3.5) \\ &= 1.5,\end{aligned}$$

so the posterior probability in favor of D-causation is $1.5/2.5 = 0.6$.

The evaluation phase for this case took the panel about three hours, spread over two sessions two weeks apart. In addition, the group's research assistant spent quite a lot of time between sessions assembling information from the medical literature and the French investigators on the amodiaquine-hepatitis study, several of the panel members interviewed clinical colleagues in Montreal, particularly hepatologists, and the moderator spent time between sessions structuring the elicitation protocol.

It is worth recalling that the result differs from the initial judgement expressed by almost all the panelists: it does not appear to be a definite case of drug-induced hepatitis, but rather nearly a toss-up. The panelist's initial judgement did not seem to take into account the evidence evaluated in the prior odds, particularly the high incidence of NANB in the epidemiological reference set defined by M. Moreover, the initial judgement was driven by the perception that positive responses to dechallenge and rechallenge had occurred, which had to be modified in the course of the probabilistic analysis.

Two other panels performed independent probabilistic analyses of this case. Both consisted of academic, industrial and Pharmacovigilance experts in adverse reactions in France (one in Paris, the other in Bordeaux). The results of the first two phases of the analysis were virtually identical for the three groups. Differences emerged in the quantitative evaluations, although the Paris and Montreal groups obtained similar results (though they took different information into account in evaluating the prior odds). The Paris group contained several experts in liver disease and several of the amodiaquine-hepatitis investigating group; as a result, they arrived at their evaluations of the likelihood ratio terms quite a bit more quickly, with less reliance on literature searches, than did the Montreal group. The Bordeaux group missed the significance of the day 70 bilirubin value, in part because they were reluctant to devalue the literal meaning of the day 60 "positive response" to dechallenge. However, when presented with the Montreal analysis, they agreed with the probability

assessments outlined above for LR(De-Re) and modified their own interpretation of the case accordingly.

4. SOME ADVANTAGES OF THE PROBABILISTIC RCA PROCEDURE

There are two other RCA procedures in current use. The usual approach to a RCA problem is *global introspection*: that is, the expert assessor collects all the facts that he thinks are relevant to the problem at hand, mixes them together in his head -- and then he just decides what his answer is. This answer is usually expressed in terms of a qualitative probability scale: for example, "definite"/"probable"/"doubtful"/ or "unrelated".

Unfortunately, global introspection does not work well. Several groups of clinical pharmacologists have demonstrated how unreliable global introspection is, by comparing their individual evaluations of a series of suspected adverse drug reactions and documenting the extent of their disagreement (Karch et al. (1976), Koch-Weser et al. (1977), Dangomau et al. (1980)). In addition, global introspection is uncalibrated: one assessor's "possible" might mean the same thing as another assessor's "probable". Other problems with global introspection are discussed in Lane (1984) and Kramer (1986).

Because of these problems with global introspection, during the past decade physicians from industry, regulatory agencies and academia have developed more than a dozen standardized assessment methods (SAM) for RCA (see Venulet et al. (1982), Herman (1984) and Steqpens (1985,1987) for reviews and examples of these methods). These SAM range from simple flow charts posing ten or fewer questions to lengthy questionnaires containing up to 84 items. However, they share a common basic structure. They divide the considerations that bear on causality assessment into a number of factors or axes: for example, timing; alternative etiological candidates; previous recognition of the event as a possible adverse reaction to the drug; dechallenge; rechallenge. Information relevant to each factor is elicited by a series of questions, the answers to which are restricted to yes/no/(and for some methods) don't know.

The questions typically call for a direct evaluation of the expert's inferences about the evidence, without explicitly eliciting the scientific or clinical knowledge that underlies the inference. For example, every SAM poses a question like "Was the timing of the event as expected for an adverse reaction to this drug?" To answer this question, the evaluator need not consider his predictive distribution for onset time, nor the meaning of "expected" relative to the spread of such a distribution. Also, he need

not consider whether the timing is more or less "as expected" relative to any other possible cause; the evidential significance of the timing information is determined just by the answer to this question ("yes" favors drug-causation; "no" tells against it).

The evaluator answers to the SAM's questions are then converted to a score for each factor, the factor scores are summed, and this overall score is converted into a value on a qualitative probability scale. As an indication of the way in which SAM process case information, all SAM to which I have applied the example case score it a "definite" for amodiaquine causation, primarily because of the "positive" responses to dechallenge and rechallenge.

The probabilistic RCA has several advantages when compared to these alternative procedures:

(1) Explicitness: Unlike global introspection, and to a greater extent than any SAM, it is explicit: it requires the assessors to define the values of the case parameters that underlie their deliberations, to list all the evidence that they regard as differentially diagnostic, to measure in an interpretable scale their evaluation of the evidentiary significance of each separate listed item.

(2) Completeness: Unlike SAM, with their preset series of question and limited range of possible answers, the probabilistic procedure allows assessors to bring to bear any line of argument that they believe may differentiate between the possible causes of the event. In particular, as was illustrated with the example case, the probabilistic procedure, unlike global introspection or SAM, can directly incorporate quantitative information from case series and epidemiological studies. And, of course, the probabilistic procedure allows assessors to take into account their uncertainty about what actually happened in any particular case and about the theories and assumptions that they use to analyze the case material.

(3) Localization: When the probabilistic procedure is successfully implemented, the assessment tasks that require the actual exercise of expert judgement are sufficiently localized that the assessors' knowledge and experience can be directly applied. Experts know about frequencies and theoretical mechanisms, not about how to establish the relative importance of various streams of evidence that may point in different directions. There is an additional value to localization: it is much easier to identify and then to resolve the sources of disagreement between experts on the local than the global level.

(4) Coherence: The method by which the various sources of evidence are merged into an overall, global RCA is completely inexplicit in global introspection and thus not open to critical examination. For SAM, the methods for converting answers to localized

subproblems into overall probability ratings are completely arbitrary. In contrast, the probabilistic method combines local judgements into global causality assessments by means of rules whose use are justified by the normative theory of coherent reasoning as developed by de Finetti. In my experience, expert assessors exposed to this theory understand and appreciate its normative force, and willingly accept the division of duties that assigns to them the task of quantifying their expert clinical and theoretical knowledge and leaves to the rules of probability the task of synthesis and inference.

In addition to these general advantages, the probabilistic procedure provides simple, canonical solutions to a number of difficult problems in the study of adverse drug effects that resist effective treatment by the alternative methods. I will briefly discuss two of these: the problems of multiple drugs and of missing information:

(1) Multiple drugs: The problem of multiple drug causal candidates is one of the most difficult aspects of causality assessment. None of the SAM deals with this problem in a satisfactory way. First, there is the conceptual difficulty of interpreting just what the result of a RCA means: the SAM give as output a qualitative probability rating, but their authors do not define these ratings (or say what to do with them!), so, for example, most of the SAM make it possible to report many "probable" drug causes of a single event E. For the probabilistic procedure, this issue is clarified by the notion of a probability distribution over the cause list, defined as a set of mutually exclusive, exhaustive propositions.

In addition, there is a kind of psychological difficulty that seems to affect most assessors, especially when they try to solve a RCA problem with global introspection: they have a hard time evaluating the evidence when there is more than one drug causal candidate around to think about. This problem is easily handled probabilistically, by means of the "one-drug-at-a-time strategy". This strategy works as follows. Suppose $D_1, \dots, D_n$ are the drug hypotheses on the cause list, and N is the union of all nondrug causes. Let A_i represent " D_i or N" (that is, A_i is the hypothesis that if E was caused by a drug, the drug cause was D_i). Also, let $PO(D_i)$ represent the posterior odds in favor of cause D_i , and $PO(D_i|A_i)$ the posterior odds in favor of cause D_i , given A_i . That is, $PO(D_i|A_i)$ solves the causality assessment problem for a case otherwise identical to the one under consideration, except that there is only a single drug causal candidate, D_i .

The following formula, which is a consequence of the rules of coherence, gives $PO(D_i)$ in terms of the conditional posterior odds:

$$(4) \quad PO(D_i) = \frac{PO(D_i|A_i)}{1 + \sum_{j \neq i} PO(D_j|A_i)}$$

Thus, when faced with a problem in which more than one drug candidate appears on the cause list, assessors can solve the problem by first addressing a series of problems each of which has only a single drug candidate, and then amalgamating the solutions to these problems using equation (4). While this strategy is not logically necessary, experience shows that its use does reduce assessors' confusion and hence incoherence considerably when they confront problems involving more than one possible drug cause.

(2) Missing Information: Since they are designed to be used on information derived from "third party" reports, causality assessment methods should be able to take into account what information is *not* available, as well as that which is. Neither global introspection nor SAM do well in this regard. Experts who are asked to carry out causality assessments based on spontaneous reports frequently say they *cannot* do so, without the results of this test or the exact time at which that occurred; while others may base their assessments entirely on plausible but highly improbable scenarios that posit particular values to missing quantities and ignore the possibility that these assumptions are in error. SAM tend either to *exclude* missing information from consideration or to count it *against* the hypothesis of drug causation. How missing information can be handled probabilistically will now be demonstrated, with respect to two important and common problems.

Noncompliance: A drug cannot cause an event if the patient did not take the drug. We generally know that a patient was *prescribed* a drug; our knowledge that he actually *took* the drug as prescribed is not usually so certain. Here is the probabilistic solution to the problem of noncompliance, when only one particular administration of the drug is under consideration.

Let Com be the proposition that the patient was compliant: he took the drug D as prescribed. Let B be background information as defined previously; so B includes the statement that the patient took D (that is, Com) and subsequently experienced an event of type E_t . Now let B_0 consist of the noncase information in B, plus the information that a patient with condition M who was *prescribed* (but did not necessarily *take*) D experienced an event of type E_t . That is, B is the intersection of B_0 and Com.

The following can then be derived from the rules of coherence:

$$(5) \quad \frac{P(D \rightarrow E \mid B_0, C)}{P(D \nrightarrow E \mid B_0, C)} = P(\text{Com} \mid M) \cdot \frac{P(D \rightarrow E \mid B, C)}{P(D \nrightarrow E \mid B, C)}$$

In this expression, $P(\text{Com} \mid M)$ is the probability that a "generic patient" with condition M prescribed D will comply with his prescription; it is *not* calculated conditionally on the patient experiencing an event of type E_t .

Thus, equation (5) can be read as follows. First, evaluate the chance that a patient with condition M who is prescribed D would comply with this prescription. Next, evaluate the posterior odds in favor of D -causation for the particular case under consideration, assuming that the patient actually took D as prescribed. Finally, to calculate *coherently* the posterior odds that D caused E *without* assuming compliance, multiply the results of the previous two evaluations together.

Missing 'Vital' Case Information: When assessors must rely on third party case information, they may think of some item of missing information that, for them, would play a decisive role in their causality assessment *were it available* (for example: whether the woman in the example case were alcoholic; or whether or not her bilirubin level was elevated on day 67). Suppose A is a case information proposition which, if known, would have differential diagnostic significance for the causality assessment. How should the fact that neither A nor A^c is specified in the given data affect the resulting assessment? Let A^* denote the information that neither A nor A^c is specified, and $LR(A^*)$ the contribution of A^* to the posterior odds. The problem, then, is to evaluate the quantity $LR(A^*)$.

At first sight, it might seem that "negative information" like A^* should have no effect on the causality assessment. However, the assessment would have a very different result if A were known to be true than if it were known to be false, and A^* should affect the assessors' opinions about how *likely A is to be true*. For suppose that whoever submits the report has some chance of appreciating the significance of the truth of A and may therefore have sought to ascertain it: then the chance that the report is submitted at all and, given that it is, that neither A nor A^c is mentioned in it, depends to some extent on whether A or A^c is true. Thus, A^* should affect the overall causality assessment.

The quantitative effect of A^* is given by the following formula, determined by the rules of coherence:

$$(6) \quad LR(A^*) = p + ((1-p) \cdot LR(A^c))$$

where

$$p = \frac{a}{a + bc} ,$$

$$a = P(A^* | A), \quad b = P(A^* | A^c) - P(A^* | A), \text{ and}$$

$$c = P(A^c | B, D \nrightarrow E, S).$$

Here, S is the case information chronologically preceding A. Now c is just the denominator of $LR(A^c)$, and hence would be evaluated in the course of assessing this likelihood ratio factor. The quantities a and b measure the assessors' opinion of the process *whereby the case information came to them*: that is, for a, they have to decide how likely they think it is that the case report would not mention A, if A were actually true (and also, for b, if A were false).

Frequently, A gives strong information in favor of one of the causal candidates (usually nondrug), while A^c gives weak evidence (by default) for the other candidates. In such cases, it is reasonable to believe that the author of the case report is more likely not to mention anything about A when A is false than when it is true. From equation (6), it then follows that $LR(A^*)$ can be regarded as an average of two components: 1, which is the appropriate likelihood ratio factor for information that does not discriminate at all between drug and nondrug causation; and $LR(A^c)$, the likelihood ratio factor that would be appropriate if A were known to be false. The more likely there is to be no mention of A when A is false than when it is true, the more $LR(A^*)$ will resemble $LR(A^c)$. Note that if A is *always* mentioned when it is true, then $LR(A^*)$ equals $LR(A^c)$ (since A^* then implies A^c); while if there is no mention of A just as frequently when A is true as when it is false (so $b = 0$), then $LR(A^*)$ equals 1 (that is, there is no information in A^*).

In summary, to evaluate the likelihood ratio corresponding to the fact that a potentially vital piece of case information, A, is simply not mentioned, do the following: evaluate $LR(A^c)$ and its denominator, $P(A^c | B, D \nrightarrow E, S)$; evaluate $P(A^* | A)$ and $P(A^* | A^c)$; now calculate $LR(A^*)$ from equation (6).

5. CURRENT DEVELOPMENT

Currently, there are two lines of development in the project of creating and implementing a probabilistic causality assessment procedure. One is concerned with applications of the procedure. These applications include such projects as:

(1) structured review of spontaneously generated case reports (for example, together with a neurologist at Vanderbilt University, Dr. Gerald Fenichel, I have

developed and implemented a procedure for guiding the deliberations of a panel of pediatric neurologists assembled by the Centers for Disease Control to review case reports received by the Centers on suspected serious neurological adverse effects of pertussis vaccinations);

(2) analysis of drug safety material from premarketing studies (together with Dr. Claudio Naranjo, a clinical pharmacologist at the University of Toronto, I developed for the manufacturers of a new drug for the treatment of congestive heart failure methodology for carrying out RCA analyses for 10 cases of agranulocytosis among users of the drug in a large premarketing observational study -- most of whom were taking a whole range of other medications -- and then updating on the basis of these RCA a predictive distribution for the incidence of agranulocytosis in future users of the drug); and

(3) incidence estimation from spontaneous reports (together with a team of pharmacologists from the University of Bordeaux and the Centre de Pharmacovigilance Bordeaux-Aquitaine, I am developing a model based upon the probabilistic decomposition of RCA that can be used to construct uncertainty distributions for the incidence of adverse effects as a function of drug and event type, drug sales history, the number of spontaneous reports received by the Centres, and a few other observable quantities).

The other direction of work on probabilistic RCA involves improving the procedure conceptually and technically. The most important conceptual problem has to do with the expression of "uncertainty about uncertainty", and the approach we are taking to this problem is to assess probability distributions that describe the panel's collective uncertainty about a given probability assessment. Because almost all of the probabilities assessed in the procedure bear frequency interpretations, it is possible to interpret these distributions in terms of uncertainty about these frequencies, and we have not found that the experts feel quite comfortable assessing such uncertainty distributions. We have not developed adequate methods yet for combining these distributions (unless we assume independence or some other simple specification for their joint distributions) to induce an overall uncertainty distribution on the posterior odds, but we are currently working on this problem.

The technical difficulties involve a number of modelling, elicitation, statistical and computational problems that must be solved before the procedure (and subjective probability) can reasonably be expected to find widespread use in the pharmaceutical industry, regulatory agencies and the medical community in general. Many of these problems are related to the speed with which the procedure can be carried out -- at

present, it is in general slow work, as shown by the example case, which occupied the panel for three sessions over a four week period, as well as many hours of the research assistant's time. Of course, it must be realized that this case was analyzed as part of an on-going project whose purpose it is to explore and develop the probabilistic procedure; the cases that are analyzed in this project were especially selected for their difficulty! Currently, five panels, consisting of experts from academia, industry and two national monitoring and regulatory agencies, are cooperating in this project.

6. BIBLIOGRAPHY

- Auriche M (1985). Approche Bayesienne de l'imputabilite des phenomenes indésirables aux médicaments. *Therapie*, 40, 301-306.
- Crocchiolo P, Caredda F, Monforte A, Lencioni R, Ragni M, Cenquales S, Farce P, Lavarini C, Latif A (1984). The aetiology of acute hepatitis in Zimbabwe. *Trans Royal Society of Tropical Med and Hygiene*, 78, 514-518.
- Dangomau J, Begaud B, Boisseau A, Albin H (1980). Les effets indésirables des médicaments. *La Nouvelle Presse Medicale*, 9, 1607-1609.
- de Finetti B (1972). *Probability, Induction and Statistics*. New York, John Wiley.
- Dienstag J (1983). Non-A, non-B hepatitis. I. Recognition, epidemiology, and clinical features. *Gastroenterology*, 85, 439-462.
- Farrow L, Stewart J, Stern H, Clifford R, Smith H, Zuckerman A (1981). Non-A, non-B hepatitis in West London. *Lancet*, 1981, i, 982-984.
- Fisher R (1973). *Statistical methods and scientific inference*. Third edition. Hafner, New York.
- Gimson A, White U, Eddleston A, Williams R (1983). Clinical and prognostic differences in fulminant hepatitis type A, B and non-A non-B. *Gut*, 24, 1194-1198.
- Greenfield C, Karayiannis P, Wankya B, Shah M, Tukei P, Galpin S, Jowett T, Thomas H (1984). Aetiology of acute sporadic hepatitis in adults in Kenya. *Journal of Medical Virology*, 14, 357-362.
- Herman R, ed. (1984). *Drug-event associations: perspectives, methods, and uses*, Proceedings of the Drug Information Association Workshop, Arlington VA, Oct. 30 - Nov. 2, 1983. *Drug Info J*, Vol. 18, Numbers 3 and 4.

- Jones J and Herman R, eds. (1986). The future of adverse drug reaction diagnosis: computers, clinical judgement and the logic of uncertainty, Proceedings of the Drug Information Association Workshop, Arlington VA, Feb. 1-4, 1986. Drug Info J, Vol. 20, Number 4.
- Kahneman D, Slovic P, Tversky A (1982). Judgement under uncertainty: heuristics and biases. Cambridge University Press, Cambridge.
- Karch F, Smith C, Kernzer B, Mazullo J, Weintraub M, Lasagna L (1976). Adverse drug reactions -- a matter of opinion. Clin Pharmacol Ther, 19, 489-492.
- Kendricks J (1974). Viral hepatitis in American missionaries abroad. Journal of Infectious Diseases, 129, 227-229.
- Koch-Weser J, Sellers E, Zacest R (1977). The ambiguity of adverse drug reactions. European J Clin Pharmacol, 11, 75-78.
- Kramer M (1986). Assessing causality of adverse drug reactions: global introspection and its limitations. Drug Info J, 20, 433-438.
- Kryger P, Aldershvile J, Christoffersen P, Hardt F, Juhl E, Mathiesen L, Nielsen J, Poulsen H, the Copenhagen hepatitis acuta programme (1980). Acute non-A, non-B hepatitis -- clinical, epidemiological and histological characteristics. Scand J Infect Dis, 12, 165-169.
- Lane D (1981). Coherence and prediction. Bull Int. Statist. Inst., 49 (1), 81-96.
- Lane D (1984). A probabilist's view of causality assessment. Drug Info J, 18, 323-330.
- Lane D (1987). Causality assessment for adverse drug reactions: an application of subjective probability to medical decision making. To appear in Berger J, Gupta S, eds, Proceedings of the Fourth Purdue Symposium on statistical decision theory and related topics..
- Larrey D, Castot A, Pessayre D, Merigot P, Machayekhy J-P, Feldmann G, Lenoir A, Rueff B, Benhamou J-P (1986). Amodiaquine-induced hepatitis: a report of seven cases. Annals of Internal Medicine, 104, 801-803.
- Lenoir A (1987). Personal communication.
- Mathiesen L, Skinoj P, Nielsen J, Purcell R, Wong D, Ranek L (1980). Hepatitis type A, B, and non-A non-B in fulminant hepatitis. Gut, 21, 72-77.

- McNeil M, Hoy J, Richards M, Lehmann N, Dimitrikakis M, Gust I, Lucas C (1984). Aetiology of fatal viral hepatitis in Melbourne: a retrospective study. *Med J Aust*, 141, 637-640.
- Neftel K, Woodtly W, Schmid M, Frick P, Fehr J (1986). Amodiaquine induced agranulocytosis and liver damage. *British Medical Journal*, 292, 721-723.
- Stephens M (1985). *The detection of new adverse drug reactions*. London, Plenum.
- Stephens M (1987). The diagnosis of adverse medical events associated with drug treatment. To appear, *Journal of Adverse Drug Reactions and Acute Poisoning Reviews*
- Venulet J, Berneker GD, Ciucci AG, eds. (1982). *Assessing causes of adverse drug reactions*. London, Academic Press.

FIGURE 1

Categories of Case Information Defined by Chronology

M -- Underlying clinical condition
D -- Suspected drug
E -- Adverse event

Hi -- Patient's history
Ti -- Timing of onset of adverse event
Ch -- Characteristics of adverse event
De -- Dechallenge
Re -- Rechallenge

CASE PARAMETERS

Clinical Condition (M):

- middle-aged French woman
- resident in Gabon for 3 years
- assumed healthy

Adverse Event Type (E_t):

- hepatitis with jaundice (def: enzymes > 5 x normal, bilirubin > 3 x normal)

Time Horizon:

- 4 months (D = amodiaquine)

Possible causes:

- amodiaquine
- chloroquine
- NANB
- alcohol-induced hepatitis
- hepatitis A
- hepatitis B

RELEVANT CASE INFORMATION

Historical Information:

Datum: 1st prodrome tiredness, vomiting (T,V)

Timing:

Datum: T,V appear day 36

Characteristics:

Data: a) onset of jaundice 12 days after T,V

b) symptomatology (days 47-49):

i) no fever

ii) no arthralgia

iii) liver function tests: enzymes 15-20 x normal
bilirubin 5 x normal

Dechallenge: (Day 50)

Data: a) subjective improvement by day 60

b) jaundice decreasing by day 60

Rechallenge: (Day 67)

Data: a) jaundice, pruritis evident day 70

b) liver function tests day 70:

enzymes 20-25 x normal

bilirubin 18 x normal

c) hepatic encephalopathy day 83

d) death with nearly total hepatic necrosis

RELEVANT CASE INFORMATION

Historical Information:

Timing:

Datum: T,V appear day 36

Characteristics:

Dechallenge: (Day 50)

Data: a) subjective improvement by day 60

b) jaundice decreasing by day 60

Rechallenge: (Day 67)

Data: a) jaundice, pruritis evident day 70

b) liver function tests day 70:
enzymes 20-25 x normal
bilirubin 18 x normal

c) hepatic encephalopathy day 83

d) death with nearly total hepatic necrosis

FIGURE 5:

FIGURE 6: