

**Causality Assessment for Adverse Drug Reactions:
An Application of Subjective Probability to
Medical Decision Making**

by

David A. Lane
University of Minnesota
Technical Report No. 492
July 1987

Text of a talk presented at the 4th Purdue Symposium on Statistical Decision Theory and Related Topics.

CAUSALITY ASSESSMENT FOR ADVERSE DRUG REACTIONS: AN APPLICATION OF SUBJECTIVE PROBABILITY TO MEDICAL DECISION MAKING

DAVID A. LANE

School of Statistics
University of Minnesota
Minneapolis, Minnesota 55455, U.S.A.

This paper describes an ongoing project on the causality assessment of adverse drug reactions. My collaborators in this project are physicians who are experts in various aspects of adverse reactions: Drs. Tom Hutchinson and Michael Kramer (clinical epidemiologists at McGill University), Claudio Naranjo (Department of Pharmacology, University of Toronto), and Judith Jones (Department of Medicine, Georgetown University). The project is supported by grants from the Drug Information Association, the American Medical Association's Education and Research Foundation, and several pharmaceutical companies.

The outline of the paper is as follows. Section 1 describes the problem of causality assessment. Section 2 outlines the approach my group is taking to solve the problem, based on coherent probability assessment. Section 3 raises and answers an important criticism of our approach, and Section 4 illustrates why it is important to supplement clinical judgement with probabilistic reasoning.

1. INTRODUCTION: CAUSALITY ASSESSMENT FOR ADVERSE DRUG REACTIONS

The problem of causality assessment concerns the following question: a patient is administered a drug and subsequently experiences an adverse clinical event; did the drug cause the event to happen? A number of important decisions depend in part on the answer to this question. For example:

- Should a clinician discontinue the use of a drug that is effective for his patient, but may be causing an adverse reaction?
- How should a pharmaceutical company react to the fact that two patients in a clinical trial of a new anti-ulcer drug developed liver disease?
- Should the editor of a medical journal publish a case report of a suspected new adverse drug reaction – and, if so, what standards of documentation should be required?
- How should an epidemiologist studying the incidence of in-hospital iatrogenic disease decide whether a particular case of renal failure is drug-induced?

In addition, many pharmaceutical companies and national regulatory agencies engaged in postmarketing surveillance routinely assess causality for case reports of suspected adverse reactions submitted to them by clinicians (these are called "spontaneous reports"). And of course, causality assessment has obvious implications in liability law.

Several factors make causality assessment a difficult medical problem (see, for example, Irey (1976) and Kramer (1981)). In particular, almost all drug-associated adverse events have alternative drug and nondrug etiologies, often including the disease that the drug is being used to treat, and there are rarely tests that conclusively differentiate between these etiological candidates. Polypharmacy poses another difficulty: many patients receive more than one drug during the same time period, so that even if it seems clear that an adverse event is drug-induced, the problem of *which* drug is responsible must still be resolved.

Here is an example of a case, paraphrased from a spontaneous report received by the manufacturer of one of the drugs (this case is discussed in Kramer (1986)):

CASE SUMMARY

A 42-year old woman was found dead in her home after a mid-day visit to the dentist. The reconstructed history from friends indicated that she was probably in good health but had been taking propranolol for high blood pressure. The dentist related that she had given a history of a heart murmur on her previous visit and that he had instructed her to take two penicillin V tablets one hour prior to her appointment for extraction of a wisdom tooth. She related no history of drug allergy. Prior to the oral surgery, she had an injection (infiltration of the area) with xylocaine. The extraction procedure lasted 40 minutes. She was instructed to go home and rest and was given a prescription for 5 zomepirac tablets, 100 mg every 4 to 6 hours for pain (p.r.n.). She was also instructed to continue the penicillin for one day.

The woman's death was discovered that evening when dinner guests arrived. Prescriptions for zomepirac, penicillin V, and propranolol were found in her home. Autopsy revealed pulmonary congestion and some hyperinflation, along with evidence of laryngeal edema and swelling of the extremities. An analysis of the stomach contents revealed no drugs associated with abuse, such as narcotics, amphetamines, or barbituates. The coroner estimated the time of death as 4:00 P.M., plus or minus 2 hours.

Notice that four drugs are mentioned in this report; at least three of them can cause sudden death from anaphylactic shock, and the fourth, propranolol, is believed to increase the chance that other drugs will produce immunologic reactions. Also, note how much the report does *not* say: for example, did the patient actually take zomepirac or penicillin (and did she just take penicillin before the procedure, or did she also take one after the procedure as recommended?) We do not know, nor do we know exactly when she died.

The usual approach to a causality assessment problem is to refer it to an expert, who solves it by an act of global introspection: that is, the expert 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. Because of the complexity and uncertainty inherent in these problems, global introspection does not work well. About ten years ago, several groups of clinical pharmacologists — including Karch et al. (1976) and Koch-Weser et al. (1977) — showed how unreliable expert global introspection is, by comparing their evaluations of a series of suspected adverse drug reactions and documenting

the extent of their disagreement. In addition, global introspection is uncalibrated: it is usual for experts to express their judgements in terms of a qualitative probability scale (that is: "doubtful", "possible", "probable", and "definite"), but your "possible" might be equivalent to my "probable". Other shortcomings of global introspection are discussed in Lane (1984).

Because of these difficulties with global introspection, during the past decade there has been much effort devoted to developing decision aids for this problem. There are now more than a dozen published algorithms or standardized assessment schemes for the causal analysis of adverse drug reactions, ranging from simple flow charts posing ten or fewer questions to lengthy questionnaires containing up to 84 items. These schemes have been constructed by physicians from industry, regulatory agencies and academia.

While the schemes differ in many particulars, they share a common basic structure. They divide the considerations that bear on causality assessment into a number of factors or axes: for example, the *timing* of the adverse event in relation to administration of the drug; *alternative etiological candidates*; *previous recognition* of the event as a possible adverse reaction to the drug; the response when the drug is discontinued (*dechallenge*) and when subsequently the drug is readministered (*rechallenge*). Information relevant to each factor is elicited by a series of questions, the answers to which are restricted to yes/no/(and occasionally) don't know. The answers to these 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.

These schemes are not free from criticism. Some experts in adverse reactions complain that the schemes, with their preset series of questions, are too inflexible (in the words of Dukes (1984), the schemes eliminate consideration of those "entirely unexpected but vital imponderables which play a role in determining the value of a particular piece of evidence"). In particular, the schemes cannot handle uncertainty about the details of a case, like whether or not the woman in the example quoted above actually took zomepirac and penicillin. Moreover, even adherents of the schemes agree that their procedures for converting answers into factor scores and factor scores to qualitative probability ratings are arbitrary.

A further defect of the schemes, evident to anyone who thinks about epistemic probability, is that they are incoherent. In particular, in the presence of the kind of "hard" frequency information that everyone could agree would justify a quantitative treatment of causality assessment, the schemes do not perform in accordance with the rules of probability. For example, suppose that when a particular analgesic is taken to relieve the pain associated with flu, 1 in 8 patients become nauseous; while 1 in 10 untreated flu patients suffer from nausea. On the other hand, drug-induced nausea is overwhelmingly likely to occur within one hour of administration of the drug, while nausea caused by the flu can occur at any time within the first two days following the onset of the flu. If in a particular case, the patient becomes nauseous 45 minutes after taking the analgesic, the odds in favor of drug causation is 12, corresponding to a probability of 0.92. Yet when this artificial case is run through current standardized schemes, the resulting causality ratings range from "doubtful" to "possible"!

There are three reasons for these too conservative assessments. First, the schemes do not allow sharp information about a single factor (in this case, timing) to override neutral or weakly negative data relating to other factors (for example, the absence of dechallenge and rechallenge information or the existence of an alternative etiologic candidate). Second, the schemes treat the *existence* of alternative etiologies as a separate factor and just examine the data about other factors for their consistency with the hypothesis of drug causation, rather than *comparing* how consistent the observed data are with each alternative etiology. Third, the schemes do not process quantitative information directly, so that the *strength* of the timing information in the example cannot be adequately assessed, nor can the information about incidences and timing be directly combined. For a detailed discussion of the problems with existing causality assessment methods, see Hutchinson and Lane (1986).

2. A BAYESIAN APPROACH TO CAUSALITY ASSESSMENT

Motivated by dissatisfaction with existing methods, my collaborators – all of whom had previously constructed published standardized schemes – and I decided to join forces to create a coherent, quantitative approach to the problem. Our project has two versions of its goal. The weak version is to develop a causality assessment method that does better than existing methods, and we have, I believe, already achieved that (see Hutchinson and Lane (1986) for a discussion of what it means to “do better”). The strong version is harder. It is to develop methods that medically-trained assessors can use and that satisfy Fisher’s fundamental dictum: never throw information away. That is, the methods should allow assessors to incorporate information about *all the factors* they regard as relevant in a particular case, from *all the sources* available to them. We have a good deal of work ahead of us before this version of our goal is achieved.

In the remainder of this section, I will summarize the ideas underlying our approach to causality assessment. See Lane et al. (1986) for details of the approach and Jones and Herman (1986) for examples of its use.

As we have formulated the problem, the goal of causality assessment is to calculate the *posterior odds in favor of drug causation*:

$$P(D \rightarrow E | B, C) \div P(D \nrightarrow E | B, C). \quad (1)$$

Here, D is a drug and E is an adverse event suffered by a particular patient. The proposition “ $D \rightarrow E$ ” (D caused E) is taken to mean that E would not have happened *as* and *when* it did, had D not been administered to the patient; the proposition “ $D \nrightarrow E$ ” denies “ $D \rightarrow E$ ”. B represents *background information*, including everything the assessor can bring to bear to analyze any case involving a connection between the drug D and events similar to E (for example: his clinical experience, the results of epidemiological studies, facts and theories from basic sciences such as pharmacology). The only case-specific information in B is the proposition that a patient with *clinical condition* M who has been administered *drug* D in a specified way subsequently develops an *adverse event of type* E_t . (M and E_t are generic characterizations. There is a good deal of freedom in defining them for any

particular assessment — see Lane et al. (1986) for some guidelines. The critical consideration is that they be explicitly defined and remain fixed throughout the course of any particular assessment problem). C represents *case information*, all the details referring to the particular patient at hand and his adverse event E (of type E_t , of course).

There are three aspects of the causality assessment problem, formulated by expression (1), that make it interesting from a statistical point of view. First, unlike some diagnostic problems in which additional confirming information will eventually result in a certain diagnosis, whether or not D caused E in any particular case is inherently and permanently unobservable. Thus, standard statistical fitting techniques for estimating the value of a probability cannot be applied to the causality assessment probability $P(D \rightarrow E | B, C)$, since there are no data on the value of the response for cases constituting a training sample. Second, there is very little “hard” data on adverse reactions, like frequencies obtained from clinical trials, since serious adverse events are quite rare, with incidences in the range 1/1000 to 1/100,000 and beyond. Third, there is an abundance of potentially useful “soft” data: clinical lore and experience, the information contained in spontaneous case reports, and facts and theories from the basic sciences. The challenge is to access and quantify this “soft” data, in such a way that it can be combined with whatever “hard” data is available to assess the posterior odds given in (1) above.

Directly assessing the posterior odds itself is possible, but very unreliable, since this is just “global introspection” made quantitative. The strategy we employ to evaluate the posterior odds is *coherent decomposition*. That is, we use the coherent linkages of probability theory (especially Bayes’ Theorem and the Law of Total Probability) to decompose the posterior odds into components that are sufficiently “local” that they are directly *accessible* to the assessor’s knowledge and experience and are *predictive*.

A probability assessment problem is *predictive* if the proposition that is the subject of the assessment can be interpreted in terms of future observables. If a suitable exchangeable reference population can be obtained, then predictive probabilities can be interpreted in terms of frequencies of future observations in the reference population, and so the methods that are used to obtain these probabilities can be subjected to predictive validation tests.

It is more difficult to know whether a probability assessment problem is in fact *accessible*. Accessibility is highly subject-matter dependent, although I am sure that there are some meta-principles that might characterize accessibility across fields. At this stage in our work, we are only trying to discover what assessment tasks are accessible to experts in adverse drug reactions. We use two empirical checks to help determine when we have an assessment problem that is accessible. First, do experts “feel comfortable” making the required assessment, or do they complain that they cannot quantitate their uncertainty about the proposition in question — they just “don’t know”? Second, when they carry out the assessment, do the experts agree with one another? If the propositions are sufficiently accessible to the experts’ experiences and those experiences adequately mirror what actually happens in the world, then it is reasonable to expect a fair amount of agreement. And, much to our surprise, we are finding that we can obtain it among the experts with whom we work. (We

are planning some experiments to document this phenomenon of “converge by localization” in the relatively near future). I believe that the ability to determine which probability assessment problems are accessible is the key to the successful application of subjective Bayesian methods in areas that share with causality assessment a preponderance of “soft” to “hard” data.

The first step in our decomposition of the posterior odds is designed to increase accessibility. Assessors find it difficult working through causality assessment problems that, like the case presented above, involve more than one drug etiological candidate at a time. We deal with this problem by the “one-drug-at-a-time” strategy, which works as follows. Suppose there are n possible drug causes, D_1, \dots, D_n . Let A_i be the proposition that if there is a drug cause of E , then that cause is D_i , and let $PO(D_i | A_i)$ be the posterior odds in favor of D_i , given A_i . That is, $PO(D_i | A_i)$ is the appropriate posterior odds for a problem identical in every respect to the original problem, except that only a single possible drug cause, D_i , is entertained by the assessor. Then by Bayes’ Theorem,

$$PO(D_i) = PO(D_i | A_i) \div (1 + \sum_{j \neq i} PO(D_j | A_j)). \quad (2)$$

By separately computing $PO(D_i | A_i)$ for each drug candidate D_i in a given problem and then using equation (2) to solve the original problem, assessors need only consider problems that have only one drug etiological candidate, which I shall henceforth refer to as D .

The second step in the decomposition illustrates how probabilities about unobservable propositions can be coherently linked to probabilities that are (at least in principle) predictive. By Bayes’ Theorem,

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

posterior odds prior odds likelihood ratio

(3)

Note the difference between the posterior and prior odds terms. Since both are evaluated conditionally on B , they both refer to a patient with clinical condition M who has been administered D and experienced an event of type E_t . But the identity of the patient differs: in the posterior odds, it is the particular patient for whom the causality assessment is being carried out, while in the prior odds, it is a “generic” patient (for example, the “next” patient) with the three defining properties M , D , and E_t . Hence the substitution of the symbol E_t for E in the prior odds.

To see how equation (3) can convert assessment problems involving unobservable propositions to problems in which all the probabilities generate predictions, consider the following special case. Imagine that records are available for a large group of patients with M , and that the subgroup of patients who receive D is otherwise prognostically equivalent to the subgroup who do not, with respect to their chances for suffering an event of type E_t . In this case, it is easy to see that

$$\frac{P(D \rightarrow E_t | B)}{P(D \nrightarrow E_t | B)} = \frac{P(E_t | D) - P(E_t | D^c)}{P(E_t | D^c)}, \quad (4)$$

where $P(E_t | D)$ is the probability of an event of type E_t among the patients who receive D , and $P(E_t | D^c)$ is the probability of such an event among the patients who do not receive D . Both

$P(E_t | D)$ and $P(E_t | D^c)$ are then predictive probabilities, and hence in this case the prior odds, without further decomposition, is a function of probabilities subject to predictive validation.

The probabilities that appear in the likelihood ratio term in (3) are not predictive, since they are conditioned on an unobservable cause. On the other hand, they can be mixed with respect to the incidence probabilities that appear in the right-hand side of (4), to yield $P(C | B)$, which generates predictions that can be verified in the class of future patients with M who experience an event of type E_t after taking drug D .

The third step in the decomposition of the posterior odds involves decomposing the likelihood ratio term in (3) into a series of factors determined by chronology. Figure 1 indicates the chronological categories of case information for a typical case: H_i (patient's history antedating the onset of E), T_i (the timing of the onset of E in relation to the administration of D), Ch (the characteristics of the event from time of onset to time of dechallenge), De (response to dechallenge), and Re (response to rechallenge). There are, of course, other ways to structure case information than chronologically, but assessors find working through the case sequentially in time easier than other patternings we have tried. In addition, the data in the various chronological categories are often mutually independent, given the etiology of the event, which simplifies the assessments of the likelihood ratio factors. Figure 2 presents some questions that elicit relevant case information in each of the five chronological categories.

In summary, the posterior odds is decomposed into six factors: the prior odds and five likelihood ratio factors. In turn, each of these must be subjected to secondary decompositions that depend on the kind of information that is available. For example, to evaluate the likelihood ratio for timing in the sudden death case, the distributions for the following quantities had to be assessed: actual time of death, given the interval provided by the coroner; the time the drugs were taken (including, for zomepirac and the two doses of penicillin, the probability they were taken at all); and, for each drug D , given D causes sudden death, the time from administration of D until death.

At this point in our work, we carry out these assessment tasks as a group activity. First, each expert in the group separately assesses the distribution under consideration. When the experts disagree with one another, we do not try to merge their probability assessments to achieve some sort of group distribution. Rather, we try to determine the informational (and conjectural) background that underlie the differences. Typically, we can identify the sources of disagreement in terms of data or possible mechanisms taken into account by some but not all of the experts. We then try to resolve the disagreements by formally incorporating the new pooled informational background by conditioning, expanding out the original probability assessment problem through the Law of Total Probabilities. In this way, we learn what kinds of considerations turn out to be relevant to each of the assessment tasks we face, and we come closer to canonical, accessible decompositions for each of these tasks.

We are now in the process of constructing models for each of these assessment tasks to replace the rather arduous *ad hoc* subjective assessment process I have described above. For example, T , the time to onset of a dose-dependent adverse reaction, can be modeled as the sum of three random variables

that are typically independent of one another: T_1 , the time to toxic accumulation of the relevant drug metabolite at the target organ (whose distribution depends on pharmacologic properties of D and on the target organ); T_2 , the time from T_1 to clinical detectability (whose distribution depends on E_t); and T_3 , the time from T_2 to detection (whose distribution depends on E_t and characteristics of the patient). We can model each of these in such a way that the relevant distribution for T can be derived in response to the answers that the assessor gives to a few factual questions, so that, if the conditions for the validity of this model are satisfied in a particular case, the assessor need do no direct probability appraisal to determine the numerator of the likelihood ratio for timing. The advantages of such models are clear: they give more reproducible assessments in an efficient way; they are relatively automatic and so can be implemented for interactive computing; and, because they can be used in many different situations, they can generate a lot of predictions to validate.

The prior odds term poses the most difficult assessment problem, because there is little in the experience of any individual clinician that helps him estimate the magnitudes of the incidence probabilities of equation (4). On the other hand, techniques for assessing these incidence probabilities (and hence the prior odds) are particularly important even outside the causality assessment context, since these probabilities play a key role in risk-benefit analyses designed to answer the question: should drug D be used in a particular clinical situation or not? There are three primary sources of quantitative information relevant to assessing the prior odds:

1) *Direct Frequencies*: Some large databases record data about exposures to drugs and subsequent experience with adverse events for particular classes of patients, drugs and events (see, for example, Jones et al. (1984) and Wilholm and Westerholm (1984)). Generally, a great deal of care must be taken when interpreting the information in these databases, particularly with regard to diagnostic classification. In addition, these databases typically provide no case details about the events they record.

2) *Indirect Frequencies*: Spontaneous report registries maintained by pharmaceutical companies, national monitoring and regulatory agencies and the medical literature contain records of suspected adverse reactions. To use the information in these registries, models need to be constructed to account for the fraction of all events that get into the registries, which presumably depends on such factors as the time on the market of the drug D , the seriousness of the event type E_t and so forth. In addition, so-called "denominator" data documenting number of exposures to the drug need to be obtained elsewhere (this can be done for many drugs fairly accurately).

3) *Related Frequencies*: Many drugs are chemically or pharmacologically related to other drugs, for which the incidence of adverse reactions may be better known. It should be possible to use information about related drugs to generate predictive distributions for new drugs about which less is known, and we are beginning to work on this problem.

Our aim is to develop models to estimate the incidence probabilities of equation (4) that directly assimilate the three kinds of data described above. We have made some progress in this direction; what we do now in practice is to use "plug-in" estimates based on subjective assessments of the terms that will eventually be modeled. For example, to assess $P(E_t | \text{zomepirac})$ in the sudden

death case, we used an indirect frequency argument, based on the number of reports that the F.D.A. had received at the time this case occurred, an estimate of number of exposures based on data from I.M.S., America, (Ambler, PA), and an estimate of reporting frequency based on the experience of one of our group (Dr. J. Jones), who had previously served as head of the F.D.A.'s spontaneous report system. We also typically now use a "concurrent validation" approach to estimating the prior odds, by generating separate estimates from different information sources, which we pool with subjective weights if they do not differ substantially and attempt to gather further information and probe for sources of the discrepancy if they are widely disparate.

At this stage in our work, our causality assessments are fairly complicated. For example, the sudden death case required the subjective assessment of about 30 probability distributions. The conclusion that followed from all these assessments was that the woman's death was nearly certainly caused by a drug, with the leading two drug candidates zomepirac (posterior probability of causation = 0.92) and the dose of penicillin taken before the dental surgery (posterior probability 0.04). This conclusion was quite robust under reasonable perturbations of the assessed values for the component probabilities (see Kramer (1986)).

3. A CRITICISM AND ITS REBUTTAL

To some, the kind of analysis describe above may seem like quantification run amuck. Thirty subjective probability assessments, each necessarily carried out with at least a little imprecision, may seem a lot more likely to produce mere *ramified imprecision* than truth.

There are several answers to this criticism:

1) *The Role of Decision*: The purpose of causality assessments should be to guide decisions. Frequently, when there are a limited number of available actions, like whether to continue or stop administering a drug suspected of causing an adverse reaction, the exact value of a probability is not the issue. The key question is whether or not the probability exceeds a certain threshold value at which the optimal action changes. If, when the values obtained for the component probability assessments are subjected to reasonable sensitivity perturbations, the probability of *D*-causation stays on the same side of the threshold, then the issue of imprecision is irrelevant. When it does not, it may be possible to identify the assessments whose imprecision is responsible for the change and to develop a coherent decomposition to convert these problems into ones that are more accessible and whose assessments are correspondingly more precise.

In addition, if decisions really hinge on the outcome of a causality assessment, what is the alternative to the approach described here? Uncertainty is a fact of life; good decision-making requires that uncertainty be measured, not ignored. And surely it is easier and more meaningful to measure uncertainty about propositions that are accessible to the assessor's knowledge and experience than to propositions that are not.

2) *The Extent of Imprecision:* The more successful a coherent decomposition is in producing a set of assessment problems that are accessible and predictive, the less imprecision there should be in the component probabilities and, I conjecture, in the overall probability. This is a testable conjecture, and testing it will provide one focus of my group's research activity in the next year.

3) *Compensating Advantages:* The probabilistic approach to causality assessment has some advantages over other approaches that directly compensate and help alleviate the disadvantage of quantitative imprecision. In particular, the approach is *explicit* (that is, it puts out in the open all the information and the assumptions that an assessment is based upon); it is *transparent* (that is, it reveals how each element in the overall assessment fits together); and it is *logical* (since it synthesizes the evidence coherently). As a consequence of these properties, the approach focuses disagreement either on the assessor's structuring of the problem (that is, whether he took into account all possible causes and sources of evidence) or on the *local* level of the component assessments, where resolution is most easily achieved.

4. PROBABILITY AND CLINICAL JUDGEMENT

Another possible approach to the causality assessment problem would be to try to construct a rule-based expert system, similar to other diagnostic expert systems like Mycin (Shortliffe and Buchanan (1984)). Such systems start with the assumption that the goal is to capture the "knowledge base" that experts use when they carry out the task that the expert system is designed to perform. In the rules that comprise this "knowledge base", information about the factors that can affect the diagnosis and opinions about the evidentiary significance of this information are inextricably intertwined. One of the main conclusions that has come out of our experience with the probabilistic analysis of causality assessment is that the experts' opinions about evidentiary significance are not necessarily trustworthy, and that to model the way experts weigh and integrate evidence may give better insight into psychology than into adverse drug reactions. I contend that the rules of coherence provide a better structure for inference than the experts' heuristic rules. I will offer in support of this contention some examples comparing our results to expert global introspection and standardized assessment schemes (which can be thought of as primitive expert systems).

Consider the question of whether or not the woman who died suddenly in the example case actually took zomepirac. Some physicians who reviewed this case at a recent conference on causality assessment reasoned as follows: since we think it is at best only "possible" that the woman took zomepirac, we cannot consider the drug more than a "possible" cause of her death, in contrast to the conclusion reached by probabilistic analysis in Kramer (1986). The fallacy in this argument is easy to find probabilistically, but apparently not so obvious through standard clinical reasoning; it mistakes a prior for a posterior probability. (Many patients do not take an analgesic after dental surgery; but most patients do not die suddenly! As Sherlock Holmes used to say, once you have eliminated the obvious, only the unlikely remains.) In fact, the analysis in Kramer (1986) assigned

probability 0.8 to the event that a patient would take a prescribed analgesic for pain after extraction of a wisdom tooth; even a much lower assignment would not have changed the overwhelmingly high posterior probability that *this* woman took zomepirac (and died as a result).

A more general example of inadequate heuristic reasoning concerns the role of the prior odds. Psychologists have shown how people seem to ignore background incidences when reasoning under uncertainty (see Kahneman et al. (1982)), and the extent to which this is true in the causality assessment context is evident from a perusal of all the standardized assessment schemes. Yet frequently our probabilistic analyses find the prior odds as the driving factor in determining drug causation in particular cases. An interesting example concerns some work that one of my collaborators (Dr. C. Naranjo) has been carrying out with respect to the connection between the anti-depressant drug zimelidine and Guillain-Barré syndrome (GBS). GBS can develop as a sequela to a viral infection, which manifests itself as general malaise, perhaps accompanied by nausea and vomiting. On the other hand, these are also the first symptoms of GBS, regardless of cause. Thus, as long as zimelidine is withdrawn quickly after onset of the syndrome, the likelihood ratio terms for any case of GBS will favor the viral cause over drug cause (since the onset of GBS follows predictably and rapidly after the flu-like symptoms and the characteristics of the syndrome do not seem to depend on the cause). So all the evidence favoring drug causation is in the prior odds, which for this problem can be estimated very precisely and are overwhelmingly in favor of zimelidine as the cause in each of the cases in the series under analysis. It is hard to see how this result could be reproduced in the expert systems context.

Again, a reading of any of the standardized assessment schemes will show that experts in adverse drug reactions seem to believe that positive responses to dechallenge and especially rechallenge provide strong and even conclusive evidence of drug causation. We have encountered examples in our assessments in which these beliefs turn out to be false; it is possible for a seemingly "positive" response to dechallenge to tell against drug causation and for a positive response to rechallenge to be essentially neutral. These examples involve delicate quantitative comparisons, and although it might be possible to construct rules that cover each case in which we have so far found exceptions to the prevailing heuristics, I am not confident that an expert who did not think probabilistically would be likely to do so.

As a final example, we find that our probabilistic assessments tend to be much more definite (one way or the other) than are the results obtained by global introspection and (particularly) standardized assessment schemes. I believe that the explanation for this phenomenon is that the heuristic rules for weighing and combining evidence that drive these alternative methods are incapable of extracting the full measure of evidence from incomplete data and uncertain beliefs. That is the job of subjective probability assessment.

BIBLIOGRAPHY

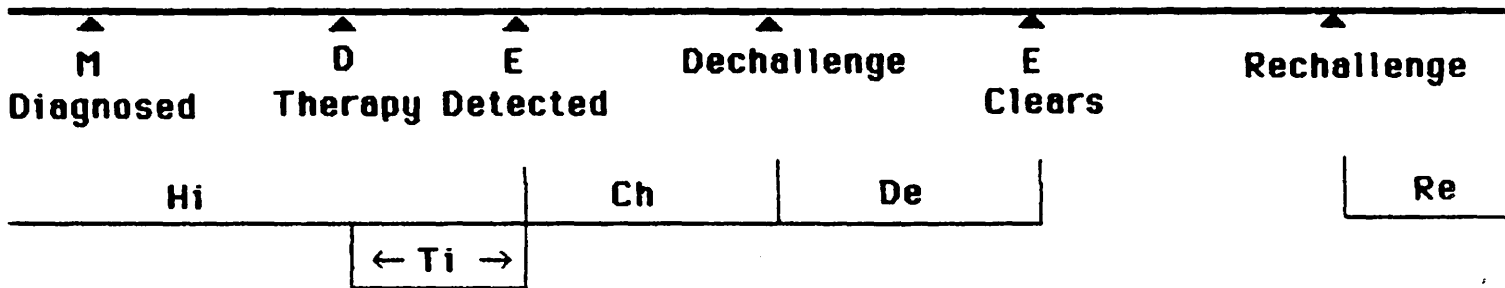
- Begaud, B. (1984), Standardized assessment of adverse drug reactions: the method used in France. *Drug Info J*, 18, 275-282.
- Blanc, S., Leuenberger, P. et al. (1979), Judgements of trained observers on adverse drug reactions. *Clin Pharmacol Ther*, 25, 493-498.
- Dangoumau, J., Evreux, J.C., and Jouglard, J. (1978), Methode d'imputabilite des effets adverses des medicaments. *Therapie*, 33, 373-381.
- Dukes, M. (1984), Adverse reactions: a changing challenge. *Proceedings of the Second World Conference on Clinical Pharmacology and Therapeutics*, Bethesda, 223-232.
- Emanuelli, A. and Sacchetti, G. (1980), An algorithm for the classification of untoward events in large scale clinical trials. *Agents and Actions*, 7, 318-322.
- Hutchinson, T. and Lane, D. (1986), *Assessing causality assessment methods*. University of Minnesota School of Statistics, Technical Report #460.
- Irey, N. (1976), Tissue reactions to drugs. *Am J Pathol*, 82, 617-647.
- Jones, J. (1982), Adverse drug reactions in the community health setting. *Family and Community Health*, 5, 58-67.
- Jones, J., Van de Carr, S., et al. (1984), Medicaid drug-event data: an emerging tool for evaluation of drug risk. *Drug Utilization Studies: Implications for medical care* (Uppsala), 127-134.
- Jones, J. and Herman, R. (Eds.) (1986). The future of adverse drug reaction drug diagnosis: computes clinical judgement and the logic of uncertainty. Proceedings of the Drug Information Association Workshop, in the *Drug Information Journal*, Vol. 20, 383-575.
- Kahneman, D., Slovic, P., and Tversky, A. (1982), *Judgement Under Uncertainty: Heuistics and Biases*. Cambridge, Cambridge University Press.
- Karch, F., Smith, C. et al. (1976), Adverse drug reactions - a matter of opinion. *Clin Pharmacol Ther*, 19, 489-492.
- Karch, F. and Lasagna, L. (1977), Toward the operational identification of adverse drug reaction. *Clin Pharmacol Ther*, 21, 247-254.
- Kramer, M. (1981), Difficulties in assessing the adverse effects of drugs. *Br J Clin Pharm*, 11, 108S-110S.
- Kramer, M. (1986), Evaluation of a case of fatal anaphylaxis. To appear in *Drug Info J*, 20, Number 4.
- Kramer, M., Leventhal, J., Hutchinson, T., and Feinstein, A. (1979), An algorithm for the operational assessment of adverse drug reactions: I. *JAMA*, 242, 623-633.
- Koch-Weser, J., Sellers, E.M., and Zacest, R. (1977), The ambiguity of adverse drug reactions. *Europ J Clin Pharmacol*, 11, 75-78.
- Lagier, G., Vincent, M. et al. (1982), Diagnostic differential des leucopenies medicamenteuses. *Therapie*, 37, 377-384.
- Lane, D. (1984), A probabilist's view of causality assessment, *Drug Info J*, 18, 323-330.
- Lane, D., Hutchinson, T., Jones, J., Kramer, M. and Naranjo, C. (1986), *A Bayesian approach to causality assessment*. University of Minnesota School of Statistics Technical Report #472.
- Naranjo, C., Busto, U. et al. (1981), A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*, 30, 239-245.
- Proceedings of a Drug Information Association Workshop on Drug-Event Associations: Perspectives, Methods, and Uses, Arlington, VA, Oct. 30-Nov. 2, 1983. *Drug Info J*, Vol. 18, Numbers 3 and 4 (1984).

- Shortliffe, E. and Buchanan, B. (1984). *Rule-Based Expert Systems*. Wiley, N.Y.
- Stephens, M. (1985), *The detection of new adverse drug reactions*. London, Plenum.
- Venulet, J., Ciucci, A., and Berneker, G.C. (1980), Standardized assessment of drug-adverse reaction associations. *Int J Clin Pharmacol Ther Toxicol*, 18, 381-388.
- Venulet, J., Berneker, G.D. and Ciucci, A.G., eds. (1982), *Assessing causes of adverse drug reactions*. London, Academic Press.
- Wilholm, B. and Westerholm, B. (1984), Drug utilization and morbidity statistics for the evaluation of drug safety in Sweden. *Drug Utilization Studies: Implications for medical care* (Uppsala), 107-118.

FIGURE 1

Categories of Case Information Defined by Chronology

Time →



- 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**

FIGURE 2

Eliciting Case Information

The assessor should answer each of the following questions. If he is unsure of the correct answer, he should state the grounds and extent of his uncertainty (in probabilistic terms). While the answers to these questions will provide all of the relevant case information for many cases, any additional case information that can help differentiate between drug and nondrug etiological candidates in a particular case under review should also be noted in the appropriate chronological period.

1. Hi:

- a. Has the patient taken *D* or similar drugs before? How frequently? On how many of these occasions did he experience an event of type E_t or another possible adverse reaction? Describe, if different from E_t .
- b. How frequently has the patient previously experienced events of type E_t without exposure to *D* or related drugs?
- c. Are there any attributes of the patient that place him at special risk to event of type E_t from any cause? If so, what are they, and from which cases is he at special risk?

2. Ti: When in relation to the course of *D*-therapy did the patient experience the event *E*? If available, give the time-course of all prodromal events.

3. Ch:

- a. Are there any data about levels of *D* in tissues or body fluids during the time the patient experienced *E*? If so, what are they?
- b. Are there any distinctive details in clinical presentation, laboratory results, pathological findings, or duration that can help differentially diagnose the cause of *E*? If so, what are they?
- c. Did the symptoms of *E* abate before dechallenge occurred? If so, how long after the time of onset of *E*?

4. De:

- a. Was *D* discontinued or its dosage reduced after the onset of *E*? If so, describe how and when.
- b. If dechallenge occurred, did the manifestations of *E* abate? If so, to what extent and when?
- c. Were the manifestations of *E* treated directly? Was a specific antagonist to *D* administered? What was the result?

5. Re:

- a. If dechallenge occurred, was the patient subsequently rechallenged with *D*? If so, when and in what dosage?
- b. If rechallenge occurred, did the manifestations of *E* recur? If so, to what extent and when?