

directed acyclic graphs. In *Proc. 13th Conf. Uncertainty in Artificial Intelligence* (eds D. Geiger and P. Shenoy), pp. 409–442. San Francisco: Morgan Kaufmann.
 Rubin, D. B. (1978) Bayesian inference for causal effects: the role of randomization. *Ann. Statist.*, **6**, 34–58.
 ——— (1986) Which ifs have causal answers. *J. Am. Statist. Ass.*, **81**, 961–962.
 Shachter, R. D. (1986) Evaluating influence diagrams. *Ops Res.*, **34**, 871–882.

Discussion on the paper by Murphy

Elja Arjas (*University of Helsinki*)

If I am right the burning issue in social and educational programmes, of which the Fast Track study is an example, is not how to determine an individually optimal dynamic treatment regime but, rather, how a limited total resource should be optimally shared between those who need help. Treatments cannot then be assigned purely on the basis of individual status and treatment histories. The situation is different in clinical trials, where the availability of drugs is not a problem but where they often have unwanted side-effects. The simple logic ‘more is better’ does not then necessarily apply.

My second comment deals with the concept of *potential outcomes* and the *no unmeasured confounders* postulate. I understand that it is tempting to postulate the existence of individual potential outcomes that are indexed by a list of treatments received, the main advantage being that the appealing concept of an ‘individual causal effect’ associated with one regime *versus* another can then be defined directly as the contrast between the corresponding potential outcomes. But I have great difficulty in understanding what the postulate of no unmeasured confounders, as formulated in the paper, would mean in a concrete study.

To illustrate this point, consider again the Fast Track study. It is mentioned in Section 5.1 that

‘staff may have used information from detailed summer interviews to assign treatment; however, in future, summer interviews may not be available’.

It is obvious that decision rules which are to be used later cannot be based on information which then will not be available. But if the summer interviews were actually determinants of how the treatments were assigned in the original study, but are no longer available when the data are analysed, the resulting statistical inference can be seriously confounded. How does this problem relate to potential outcomes? I think that it would be more natural to formulate the no unmeasured confounders assumption by referring, instead of to potential outcomes, to *potential confounders*. Somewhat more formally, we would say that $\{U_1, U_2, \dots, U_j\}$ are potential confounders at time j if the prediction of the response Y , given the observed past $\{S_1, A_1, S_2, A_2, \dots, S_j\}$, would change if the conditioning would also involve known values of $\{U_1, U_2, \dots, U_j\}$. A natural way to formulate the no unmeasured confounders assumption is now to require that for each j , given the observed past $\{S_1, A_1, S_2, A_2, \dots, S_j\}$, A_j is chosen independently of all such potential confounders $\{U_1, U_2, \dots, U_j\}$. Of course, such an assumption can never be verified from the data if the potential confounders have not been measured. But at least this alternative formulation would lead the analyst to contemplate the possible existence of factors whose values are unknown but that nevertheless might have influenced the treatment assignments that were made when the original study was carried out.

My third comment concerns the methods of statistical inference. Frankly, the many estimation methods, ranging from maximum likelihood to least squares based on nonparametric frequency estimates, left me in a state of considerable confusion. Would it not be more logical, and simpler, to start from the likelihood expression (cf. the formula below equation (12) in the paper)

$$\prod_{j=1}^K f_j(S_j | \bar{S}_{j-1}, \bar{A}_{j-1}) \prod_{j=1}^K p_j(A_j | \bar{S}_j, \bar{A}_{j-1}) g(Y | \bar{S}_K, \bar{A}_K).$$

Here we can see likelihood contributions coming, in an alternating fashion, for each j , first from observing a new value for the status variable S_j and then from recording the corresponding treatment assignment A_j , and ending after K steps with the contribution of the observed response Y given the entire status and treatment history. Under the above-mentioned version of the no unobserved confounders postulate the middle term in this likelihood does not depend on the potential confounder variables U_1, U_2, \dots, U_K . Using statistical terminology in a somewhat liberal manner we include here parameters involved in the definition of the functions f_j and g among such potential confounders. But then, in likelihood inference (including Bayesian), the middle term will only have the role of a proportionality constant, and therefore the inference regarding the functions f_j and g is unaffected by what particular distributions p_j were used

when the treatments were assigned in the original study. This reasoning is analogous to considering the likelihood expression arising from right-censored survival data: as long as the censoring mechanism is non-informative, the precise form of the censoring time distribution can be ignored in likelihood inference. There is also a close correspondence with the *set* or *do* conditioning introduced by Pearl, e.g. Pearl (2000). Murphy, at the beginning of Section 5.1, makes some brief comments in this direction but then seems to reject this option because ‘we must [then] model the distribution of each status S_j as outcomes of the past statuses and past treatment’. Yes, but I do not think that this is any more difficult than modelling the regret function μ_j . In fact, it should be easier, because we will only have to consider status variables one step ahead in time.

But if this recipe is followed and the inferential problem is considered without reference to optimization, how should the latter problem be solved? I believe that the key to explicit numerical solution here is simulation. After the distributions f_j and g have been estimated from the data, we can ‘plug in’ any regime d_1, d_2, \dots, d_k into the likelihood expression above, thereby replacing the p_j distributions by (usually deterministic) rules for treatment assignment, and then simulate values from the resulting predictive distribution of Y . If maximum likelihood has been used in estimating the functions f_j and g then the ‘randomness’ in these simulations corresponds to sequentially drawing the variables S_1, S_2, \dots, S_K and Y from the corresponding conditional distributions which are determined by the earlier values and the chosen regime. If, however, the inference has been Bayesian, the parameters of f_j and g also are considered to be random and then the simulation from the predictive distribution involves also sampling from the corresponding joint posterior. Such simulation procedures are very fast to carry out in practice and the corresponding predictive expectations can therefore be computed numerically to a high degree of accuracy by using Monte Carlo averages. If this is done, and unless the collection of all possible K -vectors of treatment decisions is very large, the optimization problem can be solved numerically by brute force: by simple ordering of such Monte Carlo estimates. Note also that, unlike in Section 5, the simulations are only a means of computing numerical values of predictive expectations, and not a way of generating alternative data sets.

A more thoughtful solution, although perhaps not ‘optimal’ in a narrow technical sense and not necessarily leading to complete ordering, would be just to consider a few interesting and reasonable looking treatment regimes and to compare how they perform. Given all the uncertainties involved in model specification and parameter estimation, this may actually be enough for answering the practical question of what treatment regime should be used. Moreover, in such a more restricted comparison of only few regimes it is also feasible to consider, not just Monte Carlo averages of the simulated values of the response variable Y , but entire predictive distributions of Y generated by the same Monte Carlo scheme.

Susan Murphy considers problems which are genuinely difficult, attempting simultaneously to carry out parameter estimation and to solve a dynamic optimization problem. In this she is genuinely testing the limits of what statistics can do. It is a pleasure to propose a vote of thanks to Susan Murphy for her interesting and thought-provoking paper.

C. Jennison (*University of Bath*)

The idea of tailoring treatment to the individual is a natural one and the prospect of further adaptation of a continuing treatment to the individual’s progress is exciting and challenging. This paper shows how problems arising in this process can be addressed within a general framework. The methodology starts at the point of model formulation, defining a probabilistic model which, when fitted, presents the optimal dynamic treatment strategy directly as an element of the model.

The paper explains that the more traditional approach of developing a model without anticipating the need to find an optimal treatment strategy can lead to difficulties for some classes of models and certain types of data. However, it may be premature to rule out altogether this ‘model first, optimize later’ approach. For instance, although the example of Section 5 is described in terms of the methods proposed, the stochastic process defined has many familiar features: the S_j -sequence is an autoregressive process controlled through the interventions A_j and, from ϕ_j , we see that the final pay-off Y is reduced by cumulative contributions from high values of the state variable S_j . What is difficult to comprehend is the full distribution of Y that is implied by the model in this format.

Careful distinction is made in Section 2 between the response of a new subject treated according to actions $a_j = d_j(\bar{S}_j, \bar{a}_{j-1})$ and the responses seen in the data available for modelling. The assumption of no unmeasured confounders allows us to disregard this distinction since the training data are representative of what future subjects will experience. Not surprisingly, the effectiveness of an action can only be assessed on states to which it is applied in the training data. However, for the assumption of no unmeasured confounders to be valid, the vector S_j must include all important covariates: each additional variable fur-

Table 3. Minimum average sample sizes

<i>Number of analyses, K</i>	<i>Non-adaptive test with optimized group sizes (%)</i>	<i>Optimal adaptive group sequential design (%)</i>
2	72.1	71.9
3	64.4	63.9
4	61.1	60.2
6	58.2	56.9
8	56.8	55.5
10	56.1	54.8

ther refines the state space and increases the sparsity of specific pairs (S_j, a_j). Now, when working with observational data, the likely reasons for subjects in the same state being exposed to different actions are variations in time and place or in the backgrounds of different practitioners. These factors are potential confounders and should, arguably, be added to the state variables S_j —removing the instance of different actions applied in a common state S_j . The underlying problem here is, of course, the familiar difficulty of drawing sound inferences from observational data.

Even in the more controlled environment of a randomized experiment, the need to compare actions applied to a subset of patients with similar states has serious sample size implications. Many studies struggle to recruit enough subjects to make an effective comparison of a major end point over two competing treatments. Identification of an optimal adaptive strategy requires comparisons within each of many subgroups of patients. Thus, severe parsimony in model definition appears to be necessary if sample size requirements are not to become completely prohibitive.

As the author points out, statisticians are accustomed to applying dynamic programming to solve sequential decision problems. This approach has proved effective in deriving optimal sequential tests (see Lai (1973), Eales and Jennison (1992) and Barber and Jennison (2002)). I have recently worked on extensions of these methods to investigate the possible benefits of *adaptive* group sequential designs where the maximum number of analyses is fixed but group sizes can be chosen in the light of observed responses. As a simple illustrative example, suppose that observations $X_i \sim N(\theta, \sigma^2)$ are available and it is desired to test the null hypothesis $H_0: \theta = 0$ against the one-sided alternative $\theta > 0$ with type I error probability $\alpha = 0.05$ and power $1 - \beta = 0.95$ at $\theta = \delta$. We set the maximum sample size at 1.2 times the necessary fixed sample size and search for group sequential tests with at most K analyses which minimize expected sample size averaged over the two cases $\theta = 0$ and $\theta = \delta$. Minimum average sample sizes are shown in Table 3, expressed as a percentage of the fixed sample size; for the optimal non-adaptive test, group sizes were optimized subject to a final sample size of 1.2 times the fixed sample size. The results show that, in fact, there is little to be gained here from allowing group sizes to be modified adaptively.

At least these unimpressive results for adaptive test show that there is no harm in keeping to simply implemented, standard group sequential designs. It would, however, be disappointing to conduct a large scale medical trial only to discover that the optimal adaptive treatment strategy offered such minor benefits. My question for the author is how great might we expect the benefits to be in the real life examples that motivated her methods? Also, might there be simple rules with, say, treatment fixed after an initial patient assessment that are close to optimal?

This last question raises an interesting point. Suppose that, first, a model is fitted to a set of data in the manner proposed. Then the procedure is repeated using a reduced set of actions to give a second model defined, as in Section 3, through the optimal strategy for this smaller action space. There is no reason to expect that the second model should be the same as the first model restricted to the reduced set of actions. Is this not a parallel to the inconsistency issues raised in Section 3.1 for ‘standard’ models?

I am grateful to the author for this stimulating paper and it gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

A. P. Dawid (*University College London*)

I want to describe the approach to thinking about ‘no unmeasured confounders’—or, much better,

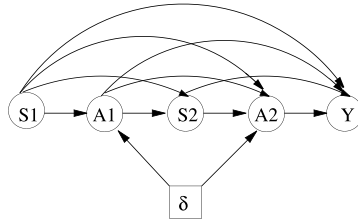


Fig. 2. Influence diagram

‘sequential ignorability’—developed by Dawid *et al.* (2001): a simpler alternative to that described in Section 2 of Susan Murphy’s paper. This uses the well-established methodology of influence diagrams, which extend directed acyclic graph representations of probability distributions by incorporating *decision nodes*, describing external interventions, in addition to the usual *random nodes* for nature’s choices. An introduction to the use of influence diagrams for causal inference can be found in Dawid (2002); see also Dawid (2003).

Consider the influence diagram of Fig. 2. Ignoring node δ , this is just a directed acyclic graph model for an alternating sequence S_1, A_1, S_2, A_2, Y of status and action variables, where each variable can depend probabilistically on all earlier variables. With a suitable specification of the conditional distribution of each variable given its past, this could represent any joint distribution of the whole system, e.g.

- (a) that under some (possibly) randomized experimental regime or, alternatively,
- (b) that under some ‘natural’ observational regime.

To support Susan Murphy’s enterprise, we need to *relate* the different circumstances of (a) and (b) above: only then can we possibly learn from (b) what we need to know about (a). The additional decision variable δ , which indexes which regime is operating, accomplishes this. The possible values of δ range over $\Delta^* := \Delta \cup \{N\}$, where Δ is the collection of experimental regimes that we care about—e.g. we wish to optimize over Δ —and N denotes natural observation. The conditional distribution $p(A_i | \bar{S}_i, \bar{A}_{i-1}; \delta)$ under any $\delta \in \Delta$ (though not necessarily under N) is supposed known, since it describes the specified, possibly randomized, experimental assignment (this distribution is degenerate in the case of deterministic experimental regimes, as considered in the paper).

The significant property of Fig. 2 is that there are no arrows out of δ into any status variable S_i (including $Y \equiv S_3$). This is equivalent to the conditional independence properties (Dawid, 1979): $S_i \perp\!\!\!\perp \delta | (\bar{S}_{i-1}, \bar{A}_{i-1}), i = 1, 2, 3$, i.e. the conditional distribution of S_i , given the past and δ , does not depend on δ —in particular, it is the same for $\delta = N$ as for any $\delta \in \Delta$. In other words, the probabilistic effect of all previous variables (both status and action) on a status variable is supposed to be a ‘stable feature’, the same across all regimes considered. This is our simple interpretation of the sequential ignorability property. When it can be assumed, it permits (under a positivity condition) estimation of $p(S_i | \bar{S}_{i-1}, \bar{A}_{i-1}; \delta)$, for an experimental regime $\delta \in \Delta$, from observational data collected under $\delta = N$.

D. R. Cox (*Nuffield College, Oxford*)

Professor Murphy’s paper addresses an important and challenging problem to the study of which she has made an impressive contribution. Nevertheless the paper left me rather puzzled. I think that it would have helped to have set out explicitly the very simplest special case that illustrates the issues involved. Such an example would have two or at the very most three periods and other simplifying features. What precisely is being assumed in any specific application? Would it be possible in the limited space available for a reply to give such a special case?

The emphasis is on optimization of individual performance. At the end of a period of applying such a procedure to a number of individuals what could be learned from an analysis of the results? Are the procedures, appropriate for optimization, good or bad for developing understanding?

The emphasis is on individual optimization assuming that all the relevant information characterizing an individual is contained in the associated scores. What would happen in extreme cases, such as application to a population containing a mixture of individuals who respond only to an initial period of special treatment and those who respond only to a uniform treatment over the study period?

In assessing the conclusions from the simulations it would help to have some simple way of interpreting

the differences found: how does one discuss whether a difference of 1 unit in mean outcome is likely to be important in the context studied?

Stephen Senn (*University College London*)

I compliment the author on having produced a paper that worries me greatly, principally because of the new and difficult theory that it implies I shall be required to learn to continue to advise on the analysis of the effects of treatments. In the past I have assumed that trying to identify individual responses was frequently impractical and that going one step further and adjusting therapy to respond to an individual response that evolved over time was usually impossible. The one exception to this rule, I had assumed, was that of dose adjustment using pharmacokinetic monitoring (here some fairly solid scientific theory, Bayes theorem and a random-effects model using data obtained from many patients helped the implementation) and I was a little surprised not to see any of the literature in this area referenced. See, for example, BagarryLiegay *et al.* (1996), Holford (1999), Racine and Dubois (1989), Rousseau *et al.* (2000), Sanathanan and Peck (1991) and Thomson and Brodie (1992). It is a general assumption that much of the variation that we see in clinical trials is patient-by-treatment interaction but there are good reasons for believing that physicians frequently overestimate this component (Senn, 2001).

I worry that there is a danger of overcontrol introducing noise into the process. In principle Professor Murphy's mathematics should deal with this. She will not, it is presumed, fall into the trap to which the late W. Edwards Deming used to draw attention (Deming, 1982) using the funnel experiment of Lloyd Nelson. One has to adjust the position of a funnel through which a marble is dropped so as to hit a cross on a table. Using only the result of the previous drop without any other form of prior knowledge is worse than making no adjustment at all (Leach, 2001). What conditions does Professor Murphy require for dynamic treatment using only past data from a subject to be superior to a more naïve approach using many results from others?

The emphasis on decision analysis in the paper is welcome. This is underutilized in medical statistics. Nevertheless in recent years there have been developments in cancer for using this to find the best dose for the next patient (O'Quigley *et al.*, 1990), in clinical trials for finding the optimal allocation approach for a sequence of patients to maximize a (time-discounted) horizon of cures (Berry and Eick, 1995), and even for choosing drugs to develop (Burman and Senn, 2003; Senn, 1996). It would be interesting to see whether there are any connections between this work and the current paper.

I congratulate the author on a fine paper and look forward with trepidation to future developments.

The following contributions were received in writing after the meeting.

Robert G. Cowell (*City University, London*)

I congratulate the author on an interesting paper that deals with a difficult problem. I would like to ask her whether she believes that her approach could be developed to analyse the types of multistage decision problems called limited memory influence diagrams introduced by Lauritzen and Nilsson (2001). In such problems a sequence of decisions is to be taken, but when making a decision the decision maker only uses some of the information from the past, not all of the information. If the author's approach can be extended to such problems, then I have two further questions relating to the regret functions.

- (a) Would their form be simplified by using only the same information that was available to the decision maker in the limited memory influence diagram or would they be parameterized by the whole past?
- (b) Will the missing information (from the decision maker's viewpoint) increase the complexity of estimating the regrets, for example, by having to resort to a method such as the EM algorithm?

V. Didelez (*University College London*)

Often we shall not find it easy to think directly about the acceptability of the stability property that Professor Dawid describes. Instead, we may be able to build an acceptable influence diagram model of an extended situation involving additional variables. We can then use graphical manipulations, e.g. the 'moralization property' (Cowell *et al.* (1999), section 5.3), to investigate whether the desired conclusions $S_i \perp\!\!\!\perp \delta \mid (\bar{S}_{i-1}, \bar{A}_{i-1})$ follow. It might sometimes, but not always, be appropriate to describe such additional variables as 'potential confounders', but in any event their existence is not fundamental, either to our definition or to its application.

Robins's G -computation formula, which can appear mysterious, becomes a triviality in our approach. We are interested in discovering the distribution $p(Y | \delta_0)$ of the response Y that would result from applying some experimental regime δ_0 . Simple probability theory gives

$$P(Y|\delta_0) = \int p(Y|S_1, A_1, S_2, A_2; \delta_0) p(A_2|S_1, A_1, S_2; \delta_0) \\ \times p(S_2|S_1, A_1; \delta_0) p(A_1|S_1; \delta_0) p(S_1|\delta_0) dA_2 dS_2 dA_1 dS_1.$$

Now, because δ_0 is a well-specified experimental regime, we know the (perhaps degenerate) randomization distributions $p(A_2|S_1, A_1, S_2; \delta_0)$ and $p(A_1|S_1; \delta_0)$. Also, the sequential ignorability property $S_i \perp\!\!\!\perp \delta_i | (\bar{S}_{i-1}, \bar{A}_{i-1})$ ensures that we can replace δ_0 by N in $p(Y|S_1, A_1, S_2, A_2; \delta_0)$, $p(S_2|S_1, A_1; \delta_0)$ and $p(S_1|\delta_0)$. What results is precisely the G -formula.

The G -formula in fact holds under weaker conditions than sequential ignorability (Pearl and Robins, 1995); again, these can be naturally expressed by using influence diagrams.

As indicated in Susan Murphy's paper, the usual approach to these issues is based on considerations of potential outcome variables. We find the ingredients and assumptions that are required for this rather inscrutable. The purely probabilistic approach that we have outlined is more straightforward, both mathematically and conceptually.

Richard D. Gill (*University of Utrecht*)

This is a beautiful and original paper, but I would like to be pedantic about lemma 1 in Appendix A. The statement of the lemma is meaningless so one wonders what the author is really trying to say. Conditional expectations are random variables. So they can be almost surely equal on a probability space, or they can be measurable with respect to a σ -algebra, but not 'almost surely equal on a σ -algebra'. They are essentially uniquely defined anyway.

This is not merely a technical issue; it is important to demystify conditioning. The measure theoretic conditional expectation is a random variable, an essentially unique function of the conditioning variable, and it can be computed as an ordinary expectation with respect to the essentially unique and always existing conditional law (Pollard, 2001). When formulae from elementary probability (assuming continuous or discrete densities) are applicable, they give the right answer.

Recall that we are given three random variables Z_1, Z_2 and Z_3 where Z_2 is discrete and Z_3 is continuous. We are given a 'discretizing' function d such that the probability that $d(Z_3)$ equals Z_2 is positive; in fact, also conditional on $Z_2 = z_2$ it is positive (with probability 1 over the values z_2 of Z_2). Thus the event $\{Z_2 = d(Z_3)\}$ has positive probability and hence naïve probabilistic conditioning given this event is meaningful. It also has positive probability given Z_3 . We can compute the expression $E[Z_1|Z_3, Z_2 = d(Z_3)]$ in various ways, which by the general theory all give the same answer. For instance, take the measure theoretic conditional expectation of the random variable Z_1 given the random vector (Z_3, Z_4) where $Z_4 = Z_2 - d(Z_3)$. The result is a function of Z_3 and Z_4 which I can partially evaluate at $Z_4 = 0$. I can show that the result of this is an essentially unique function of Z_3 —is this the point?

Alternatively compute the conditional expectation given Z_2 and Z_3 ; then average with respect to the discrete conditional probability distribution of Z_2 given Z_3 and given that $Z_2 = g(Z_3)$. Where is the mystery?

J. B. Kadane (*Carnegie Mellon University, Pittsburgh*)

This is a most impressive paper, on a challenging problem.

Although the treatments in this paper are dynamic, the data are collected in a batch. Thus there is no possibility to use the experience of previously treated individuals to improve that of yet to be treated individuals. Extensions of the methods in this paper to sequentially observed data would be interesting.

Observational studies must be treated with great caution, whether the data are batch or dynamic. How do we know whether the treatments were assigned solely on the basis of the covariates? Even the people assigning the treatments may not know. I wonder how robust empirical conclusions using the methods suggested in the paper are to this kind of misspecification.

James M. Robins (*Harvard School of Public Health, Boston*)

I congratulate Susan Murphy on this seminal paper. Here I describe an alternative approach based on optimal double-regime structural nested mean models (DRSNMMs), developed after reading Susan's paper (Robins, 2003). First I show that Murphy's regret model is isomorphic as a counterfactual model to

an SNMM of Robins (1994). Thus, as conjectured by Murphy, my SNMM results immediately indicate how to extend her results to include

- (a) sensitivity analysis and instrumental variable methods for unmeasured confounding (Robins *et al.* (1999), section 2d.5),
- (b) continuous time treatments (Robins, 1998),
- (c) locally semiparametric efficient doubly robust (LSEDR) estimation (Robins, 2000) and
- (d) an asymptotic distribution-free test of the g -null hypothesis that the mean response is the same for all regimes (Robins, 1997) (provided that treatment probabilities are correctly modelled).

In many biomedical studies, the g -null hypothesis holds, so no treatment is necessary. Methods lacking (d) often result in active treatment being inappropriately recommended. Only Murphy's and my methods (Robins, 2003) can include (d).

Let $\underline{z}_k = (z_k, \dots, z_K)$ and $\bar{z}_k = (z_0, \dots, z_k)$. Given regimes $d = \bar{d} = (d_0, \dots, d_K)$ and $d^\dagger = \bar{d}^\dagger$ plus treatment \bar{a}_{k-1} , let $Y(\bar{a}_{k-1}, d_k^\dagger, \underline{d}_{k+1})$ be the response when \bar{a}_{k-1} is followed through $k - 1$, d^\dagger is followed at k and \bar{d} is followed from $k + 1$. The function

$$\begin{aligned} \gamma_m^{\bar{d}, \bar{d}^\dagger}(\bar{s}_m, \bar{a}_m) &\equiv \gamma_m^{d_{m+1}, d_m^\dagger}(\bar{s}_m, \bar{a}_m) \\ &= E[Y(\bar{a}_m, \underline{d}_{m+1}) - Y(\bar{a}_{m-1}, d_m^\dagger, \underline{d}_{m+1}) | \bar{S}_m = \bar{s}_m, \bar{A}_m = \bar{a}_m] \end{aligned}$$

represents the conditional mean causal effect of treatment a_m versus treatment d_m^\dagger (\bar{s}_m, \bar{a}_{m-1}) before following \bar{d} from $m + 1$. Henceforth assume no unmeasured confounders. Then,

$$\gamma_m^{d_{m+1}, d_m^\dagger}(\bar{s}_m, \bar{a}_m) = E[Y(\bar{a}_m, \underline{d}_{m+1}) - Y(\bar{a}_{m-1}, d_m^\dagger, \underline{d}_{m+1}) | \bar{S}_m(\bar{a}_{m-1}) = \bar{s}_m]$$

and $\gamma_m^{\bar{d}, \bar{d}^\dagger}(\bar{s}_m, \bar{a}_m) \equiv 0$ represents the g -null hypothesis (Robins, 2003). A $(\bar{d}, \bar{d}^\dagger)$ DRSNMM specifies

$$\gamma_m^{\bar{d}, \bar{d}^\dagger}(\bar{s}_m, \bar{a}_m) = \gamma_m^{\bar{d}^\dagger}(\bar{s}_m, \bar{a}_m; \beta^\dagger)$$

where $\gamma_m^{\bar{d}^\dagger}(\bar{s}_m, \bar{a}_m; \beta)$ is known, $\beta^\dagger \in R^p$ and $\gamma_m^{\bar{d}^\dagger}(\bar{s}_m, \bar{a}_m; \beta) = 0$ if $a_m = d_m^\dagger(\bar{s}_m, \bar{a}_{m-1})$ or $\beta = 0$. Robins's (1994) SNMMs are the case in which \bar{d} and \bar{d}^\dagger are both the 'zero' regime. By defining the zero level of A_m for subjects with history $(\bar{s}_m, \bar{a}_{m-1})$ to be $d_m(\bar{s}_m, \bar{a}_{m-1})$, this model applies whenever $\bar{d} = \bar{d}^\dagger$. (Note that the substantive meaning of $a_m = 0$ changes.) Murphy's regret model is the SNMM with $\bar{d} = \bar{d}^\dagger$ being the optimal regime \bar{d}^* . Like Robins (1999), page 125, she uses a parameterization where $\beta_{scale}^\dagger = (\beta_1^\dagger, \beta_4^\dagger, \beta_7^\dagger)$ being 0 in her equations (15)–(16) implies the g -null hypothesis and that the remaining components of β^\dagger are undefined.

Limitations of Murphy's methodology include

- (a) estimation of β^\dagger based on smooth function optimization methods necessarily requires (differentiable) approximations of indicator functions and
- (b) regrets are not effect measures about which scientists have clear substantive opinions that are amenable to easy modelling.

Optimal DRSNMMs overcome these limitations. They specify that \bar{d} is the optimal regime \bar{d}^* and that \bar{d}^\dagger is the zero regime, with zero again being substantively meaningful. Then $\gamma_m^{\bar{d}^\dagger}(\bar{s}_m, \bar{a}_m; \beta^\dagger)$ is simply the mean effect of a_m (versus zero) at m , before following \bar{d}^* . Further $d_m^*(\bar{s}_m, \bar{a}_{m-1}) = \arg \max_{a_m} \{\gamma_m^{\bar{d}^\dagger}(\bar{s}_m, \bar{a}_m; \beta^\dagger)\}$. Robins (2003) derives a LSEDR estimator of β^\dagger .

The author replied later, in writing, as follows.

I thank each of the discussants for their thoughtful comments; these comments have helped me to see this work from heretofore unappreciated angles. In the following I provide brief replies to some of the issues; certainly many deserve more detailed explanation.

I am particularly intrigued by Professor Robins's alternative approach to estimation of the optimal rules. For simplicity set the number of decisions to $K = 2$ and make the assumption of no unmeasured confounders. Recall that the regrets $\mu_2(\bar{s}_2, \bar{a}_2)$ and $\mu_1(s_1, a_1)$ can be expressed in terms of the potential outcomes as

$$\mu_2(\bar{S}_2, A_1, a_2) = E[Y(A_1, d_2^*) | \bar{S}_2, A_1] - E[Y(A_1, a_2) | \bar{S}_2, A_1]$$

and

$$\mu_1(S_1, a_1) = E[Y(d_1^*, d_2^*)|S_1] - E[Y(a_1, d_2^*)|S_1].$$

In this paper the regrets are modelled in terms of the optimal decision rules. Robins’s proposal is to decompose each regret into two terms of which only the first term need be modelled, i.e.

$$\begin{aligned} \mu_2(\bar{S}_2, A_1, a_2) &= \{E[Y(A_1, a_2)|\bar{S}_2, A_1] - E[Y(A_1, 0)|\bar{S}_2, A_1]\} \\ &+ \{E[Y(A_1, 0)|\bar{S}_2, A_1] - E[Y(A_1, d_2^*)|\bar{S}_2, A_1]\} \end{aligned}$$

and

$$\begin{aligned} \mu_1(S_1, a_1) &= \{E[Y(a_1, d_2^*)|S_1] - E[Y(0, d_2^*)|S_1]\} \\ &+ \{E[Y(0, d_2^*)|S_1] - E[Y(d_1^*, d_2^*)|S_1]\}. \end{aligned}$$

Robins’s optimal double-regime structural nested mean model is a model for the first term in each of these sums. This method does not provide an explicit parameterization of the optimal decision rules. It may be that, by modelling only the first term in the decomposition of the regrets, the computational difficulties that are inherent in the approach proposed here, that of modelling the regrets (the computational difficulties are due in part to the fact that the regrets must be non-negative functions) will be simplified. I look forward to Robins (2003).

Professor Robins also comments that the estimation of β requires differentiable approximations of indicator functions. One might form this impression if Section 3 on modelling the regrets and theorem 2 are skipped and only the simple simulation is read. Note that theorem 2 provides an objective function that is to be minimized to estimate the unknown parameters in the regret. This objective function need not be differentiable in the unknown parameters. Discrete optimization and optimization of non-smooth functions are old areas in optimization theory. When the decision values are binary, we might parameterize the regret by

$$\beta_1(A - I\{\beta_2^T S > \beta_3\})^2$$

where the β s are the unknown parameters constrained by $\beta_1 \geq 0$ and $\beta_3 \in \{-1, 1\}$.

Professor Robins’s last comment is that ‘regrets are not effect measures about which scientists have clear substantive opinions that are amenable to easy modelling’. This is a clear indication that Professor Robins and I participate in different scientific communities. Many substantive scientists are at present using their clinical experience, past experimental evidence and their scientific theories to formulate optimal decision rules and to compare the resulting dynamic treatment regime with treatment as usual or control conditions. See, for example, Brooner and Kidorf (2002), Sobell and Sobell (1999), Prochaska *et al.* (2001), Kreuter *et al.* (1999), Breslin *et al.* (1999), Conduct Problems Prevention Research Group (1999) and Cooperative Research Group (1988). These scientists have strong substantive opininons about the form of the optimal rules. Thus it makes sense to model the optimal rules explicitly rather than implicitly. It is certainly true that some scientists do not have a clear quantitative understanding of the assumptions that they make when they formulate optimal decision rules; it is our job to communicate such an understanding.

Professor Cowell asks whether the estimation method that is proposed here might be used when we want to find the best decision rules out of the class of decision rules that use only a subset of the past information (Lauritzen and Nilsson’s (2001) limited memory influence diagrams). Such multistage decision problems arise when it is too expensive to keep a full record of a subject’s past information. This question is much more subtle than I first thought. For example, suppose that $K = 2$; then maximizing $E[Y|S_2 = s_2, A_2 = a_2]$ over all a_2 will not necessarily result in the limited memory optimal rule. This is because the relationship between Y and S_2 may differ according to the distribution of treatment at time 1, A_1 . This is not a problem that is amenable to a straightforward dynamic programming argument. A brute force method would entail estimating not only the regrets but also the distribution of S_2 given S_1 , indexed by $a_1(f_2(s_2|s_1, a_1))$; we could then minimize

$$\int_{s_1} \mu_2\{\bar{s}_2, d_1(s_1), a_2\} f_2\{s_2|s_1, d_1(s_1)\} f_1(s_1) ds_1$$

over a_2 . I find this method displeasing as we do not directly model the limited memory optimal rule. At this time I do not have a more pleasing alternative.

I appreciate Professor Gill’s comments and he is of course, correct; we should always strive for precision!

Before addressing Professor Jennison's interesting points, I note that the methods proposed here do not assume that the underlying process $S_1, A_1, S_2, A_2, \dots$ follows an autoregressive model or satisfies Markovian-type properties. We can assume Markovian-type properties in modelling the regrets but this is a modelling decision and is not necessary. Additionally in practice Y is often a summary statistic; for example, in an aftercare programme for recovering alcoholics, Y might be the average number of days of heavy drinking over the entire study period.

Professor Jennison asks whether the benefits of an adaptive strategy outweigh the complexity of such a strategy in a real life setting and whether a simple matching strategy (i.e. match all future treatments to individual on the basis of an initial assessment) might be sufficient. These issues are currently being discussed and investigated in the treatment of alcohol addiction. To a large extent, a simple matching strategy has been abandoned; see the papers about 'Project match' (Project Match Research Group, 1997, 1999) for discussion. Perhaps as more biological information becomes available such matching strategies will be found to be useful. It is not yet known whether the benefits of an adaptive strategy (i.e. tailoring the type or level of treatment to measures of individual need over time) will outweigh a one size fits all treatment. However, the current thought is that adaptive treatment strategies are more likely to be successful in the treatment of chronic relapsing disorders, such as aftercare programmes that have the goal of preventing relapse to alcohol misuse. These treatment strategies are commonly called stepped care approaches (Sobell and Sobell, 1999). Some important questions are how long should we provide the initial prevention treatment before deciding that it is not working, what information should be used to decide whether a prevention treatment is not working, which therapy should be used as a secondary treatment and what information should be used to taper off treatment? As soon as we envision treating individuals for whom the initial prevention treatment is 'not working' differently from individuals for whom treatment appears to be preventing relapse we are envisioning an adaptive treatment strategy.

Professor Arjas comments that in many social and educational programmes a burning issue is how to allocate resources efficiently between individuals in need of help. This issue is very important in the translation of efficacious clinical treatments into effective community interventions. In this paper, the focus is on building an efficacious clinical treatment. Of course cost can play a role even here; the outcome Y can involve cost in addition to treatment response.

To outsiders, it often appears that, in social and behavioural programmes, more intensive treatment is always better. Unfortunately this is not the case; instead negative side-effects are more subtle than in medical trials. For example, trying to provide too much of one treatment component in a multicomponent treatment may result in reduced compliance with other treatment components. The result may be that the client does not receive the treatment component that is most likely to be beneficial. This occurs because the client views his own time as valuable. Furthermore behavioural therapies such as counselling sessions are time consuming and may occur during work hours, thus requiring that the client misses work. Also at first thought we might believe that social and behavioural programmes cannot have unwanted negative consequences. This is simply untrue. See for example Dishion *et al.* (1999) and Poulin *et al.* (2001).

Professor Cox poses the intriguing question about whether the proposed procedures, although appropriate for optimization, might be as good for developing understanding as other procedures. Statisticians need to develop a paradigm for evaluating the effect of time-varying treatments. For example, are we making statements about future treatments when we discuss the effect of treatment at time j ? Furthermore the effect of treatment at time j has no meaning except in contrast with something else. What is this something else? Robins's structural nested mean models compare the 'treatment a_j and control treatment thereafter' with 'control treatment at time j and control treatment thereafter'. The regrets compare 'treatment a_j and optimal treatment thereafter' with 'optimal treatment at time j and optimal treatment thereafter'. The analogy in regression is how we choose the meaning for the intercept term. These different ways of modelling help us better to understand effects of time-varying treatments. Note also that the method proposed requires only a model for parts (the regrets) of the conditional mean of Y , leaving the remaining parts and also the distribution of each status S_j given the past free. We can choose to model these other parts as well.

Professor Cox also requests a description of this method in a simple case to gain intuition and he poses the question about what would happen in the case in which the population is actually a mixture of two types of individuals. Here is a very simple case. Suppose that type 0 individuals respond only to the treatment labelled 0 and type 1 individuals respond only to the treatment labelled 1. Let U be the binary type. We do not observe U ; rather we observe S , a characteristic of the individual, A , the outcome of randomization to treatment 1 versus treatment 0 and lastly the response Y . Suppose that $E[Y|S, U, a] = 100aU + 100(1 - a)(1 - U)$; i.e. if a person is of type $U = 1$ and is treated with treatment

$a = 1$ then the mean response is 100; otherwise it is 0. Similarly if a person is of type $U = 0$ and is treated with treatment $a = 0$ then the mean response is 100; otherwise it is 0. To derive the regret at time 1 (there is only one decision), we observe that $E[Y|S, a] = 100a E[U|S] + 100(1 - a)(1 - E[U|S])$. Maximizing we see that the optimal rule is $d(S) = \mathbf{1}_{E[U|S] > 0.5}$ and thus the regret $E[Y|S, a = d(S)] - E[Y|S, a]$ is

$$100(\mathbf{1}_{E[U|S] > 0.5} - a)E[U|S] - 100(\mathbf{1}_{E[U|S] > 0.5} - a)(1 - E[U|S]) = 100|2 E[U|S] - 1|(a - \mathbf{1}_{E[U|S] > 0.5})^2.$$

So in this case the scale parameter $\eta(s)$ is $100|2 E[U|S = s] - 1|$. In an experimental setting (A is randomized) and using a quadratic link function, the consistency of the estimation method would depend on an appropriate model for the scale parameter $\eta(s)$. Of course this estimation method only uses a model for the regret; if I were willing to model the combination of all three terms in $E[Y|S, a] = -\{E[Y|S, a = d(S)] - E[Y|S, a]\} + \{E[Y|S, a = d(S)] - E[E[Y|S, a = d(S)]]\} + E[E[Y|S, a = d(S)]]$ then I would produce less variable estimators than the method proposed but at the price of potential bias as I would be modelling more of the distribution.

I appreciate Professor Senn's list of additional references; I have learned much by reading in this area. Professor Senn states that this paper worries him greatly; please blame my exposition not the method! I must stress that this paper is not about optimizing treatment effect on the present patient by using the last patient's response. Rather this paper is about using a sample of past patients to find a good strategy that can be used to make treatment decisions on a present patient, such as which information (and how that information) should be used to decide when to switch from an initial treatment to a secondary treatment and so on. In effect, information from the sample of past patients is pooled (via the estimated rules) with the data from the present patient (the S_j s) to produce treatment recommendations. In spirit, this is similar to individualizing doses by using pharmacokinetic models. There data on a sample of past patients is used to estimate a population model that is then subsequently used as a prior in a Bayesian forecasting model for the present patient. In both cases a next step is a confirmatory randomized control trial comparing the rules-based dynamic treatment regime with the standard treatment. It would be rather interesting to see whether methods from the dose individualization area are more widely useful, particularly the Bayesian approaches.

Comments by Professor Kadane, Professor Senn and Professor Jennison suggest or hint at the idea of sequentially updating a dynamic treatment regime. The beginnings of this can be seen in the work of Legedza and Ibrahim (2001), wherein they sought to match the maximum tolerated dose to the individual's background characteristics. They updated their rule for ascertaining the maximum tolerated dose sequentially from patient to patient by using the Bayes rule and past information. There is, however, only one decision per patient.

Professor Dawid and Dr Didelez's formulation of sequential ignorability is attractive for several reasons. First if my goal in analysing an observational study is to propose an experimental study then I need to compare all the ways in which responses from the observational study might differ from the experimental study; similarly if my goal in analysing an experimental study is to propose the communitywide implementation of the treatment then again I need to compare all the ways in which the response from the experimental study might differ from the response from the community. The quantification of this, expressed by Professor Dawid and Dr Didelez via δ , reminds me that I must be mindful of the setting in which I want to use my results. A second reason why I welcome their formulation of sequential ignorability is that I suspect that this approach may help us to eliminate much of the measure theoretic difficulties that occur in the formulation of Robins's G -computation theorem (see equation (6)) when we have a continuous treatment space. This would be nice indeed!

References in the discussion

- BagarryLieghey, D., Nicoara, A., Duffaud, F., Guillet, P., Pignon, T., Catalin, J., Durand, A. and Favre, R. (1996) Individual dosage adjustment of high-dose methotrexate in clinical practice. *Rev. Med. Intern.*, **17**, 689–698.
- Barber, S. and Jennison, C. (2002) Optimal asymmetric one-sided group sequential test. *Biometrika*, **89**, 49–60.
- Berry, D. A. and Eick, S. G. (1995) Adaptive assignment versus balanced randomization in clinical-trials—a decision-analysis. *Statist. Med.*, **41**, 231–246.
- Breslin, F., Sobell, M. B., Sobell, L. C., Cunningham, J. A., Sdao-Jarvie, K. and Borsoi, D. (1999) Problem drinkers: evaluation of a stepped-care approach. *J. Subst. Abuse*, **10**, 217–232.
- Brooner, R. K. and Kidorf, M. (2002) Using behavioral reinforcement to improve methadone treatment participation *Sci. Pract. Perspect.*, **1**, 38–46.

- Burman, C.-F. and Senn, S. J. (2003) Examples of option value in drug development. *Pharmaceut. Statist.*, **2**, in the press.
- Conduct Problems Prevention Research Group (1999) Initial impact of the Fast Track prevention trial for conduct problems: I, the high-risk sample. *J. Consult. Clin. Psychol.*, **67**, 631–647.
- Cooperative Research Group (1988) Rationale and design of a randomized clinical trial on prevention of stroke in isolated systolic hypertension. *J. Clin. Epidem.*, **41**, 1197–1208.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. and Spiegelhalter, D. J. (1999) *Probabilistic Networks and Expert Systems*. New York: Springer.
- Dawid, A. P. (1979) Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc. B*, **41**, 1–31.
- Dawid, A. P. (2002) Influence diagrams for causal modelling and inference. *Int. Statist. Rev.*, **70**, 161–189.
- Dawid, A. P. (2003) Causal inference using influence diagrams: the problem of partial compliance (with discussion). In *Highly Structured Stochastic Systems* (eds A. Frigessi and S. Richardson). Oxford: Oxford University Press. To be published.
- Dawid, A. P., Didelez, V. and Murphy, S. (2001) On the conditions underlying the estimability of casual effects from observational data. To be published.
- Deming, W. E. (1982) *Out of the Crisis*. Cambridge: Massachusetts Institute of Technology Press.
- Dishion, T. J., McCord, J. and Poulin, F. (1999) When interventions harm: peer groups and problem behavior. *Am. Psychol.*, **54**, 755–764.
- Eales, J. D. and Jennison, C. (1992) An improved method for deriving optimal one-sided group sequential tests. *Biometrika*, **79**, 13–24.
- Holford, N. H. (1999) Target concentration intervention: beyond Y2K. *Br. J. Clin. Pharmacol.*, **48**, 9–13.
- Kreuter, M. W., Strecher, V. J. and Glassman, B. (1999) One size does not fit all: the case for tailoring print materials. *Ann. Behav. Med.*, **21**, 276–283.
- Lai, T. L. (1973) Optimal stopping and sequential tests which minimize the maximum sample size. *Ann. Statist.*, **1**, 659–673.
- Lauritzen, S. L. and Nilsson, D. (2001) Representing and solving decision problems with limited information. *Managmt Sci.*, **47**, 1235–1251.
- Leach, L. P. (2001) Putting quality in project risk management: part 1, understanding variation. (Available from http://www.advanced-projects.com/CCPM/Papers/Variation_Part1.PDF.)
- Legedza, A. T. R. and Ibrahim, J. G. (2001) Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. *Statist. Med.*, **20**, 867–882.
- O’Quigley, J., Pepe, M. and Fisher, L. (1990) Continual reassessment method—a practical design for phase-I clinical-trials in cancer. *Biometrics*, **46**, 33–48.
- Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Pearl, J. and Robins, J. (1995) Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proc. 11th Conf. Uncertainty in Artificial Intelligence* (eds P. Besnard and S. Hanks), pp. 444–453. San Francisco: Morgan Kaufmann.
- Pollard, D. (2001) *User’s Guide to Measure Theoretic Probability*. Cambridge: Cambridge University Press.
- Poulin, F., Dishion, T. J. and Burraston, B. (2001) 3 year iatrogenic effects associated with aggregating high-risk adolescents in cognitive-behavioral preventive interventions. *Appl. Devlpmnt Sci.*, **5**, 212–224.
- Prochaska, J. O., Velicer, W. F., Fava, J. L., Rossi, J. S. and Tsoh, J. Y. (2001) Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. *Addict. Behav.*, **26**, 583–602.
- Project Match Research Group (1997) Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *J. Stud. Alc.*, **58**, 7–29.
- Project Match Research Group (1999) Comments on project MATCH: matching alcohol treatments to client heterogeneity. *Addiction*, **91**, 31–34.
- Racine, A. and Dubois, J. P. (1989) Predicting the range of carbamazepine concentrations in patients with epilepsy. *Statist. Med.*, **8**, 1327–1338.
- Robins, J. M. (1994) Correcting for non-compliance in randomized trials using structural nested mean models. *Communs Statist.*, **23**, 2379–2412.
- Robins, J. M. (1997) Causal inference from complex longitudinal data. *Lect. Notes Statist.*, **120**, 69–117.
- Robins, J. M. (1998) Correction for non-compliance in equivalence trials. *Statist. Med.*, **17**, 269–302.
- Robins, J. M. (1999) Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: the Environment and Clinical Trials* (eds M. E. Halloran and D. Berry), pp. 95–134. New York: Springer.
- Robins, J. M. (2000) Robust estimation in sequentially ignorable missing data and causal inference models. *Proc. Bayesian. Statist. Sect. Am. Statist. Ass.*, 6–10.
- Robins, J. M. (2003) Estimation of optimal treatment strategies. In *Proc. 2nd Seattle Symp. Biostatistics*. To be published.
- Robins, J. M., Greenland, S. and Hu, F.-C. (1999) Rejoinder to comments on ‘Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome’. *J. Am. Statist. Ass.*, **94**, 708–712.

- Rousseau, A., Marquet, P., Debord, J., Sabot, C. and Lachatre, G. (2000) Adaptive control methods for the dose individualisation of anticancer agents. *Clin. Pharmkin.*, **38**, 315–353.
- Sanathanan, L. P. and Peck, C. C. (1991) The randomized concentration-controlled trial—an evaluation of its sample-size efficiency. *Contr. Clin. Trials*, **12**, 780–794.
- Senn, S. J. (1996) Some statistical issues in project prioritization in the pharmaceutical industry. *Statist. Med.*, **15**, 2689–2702.
- Senn, S. J. (2001) Individual therapy: new dawn or false dawn. *Drug Inform. J.*, **35**, 1479–1494.
- Sobell, M. B. and Sobell, L. C. (1999) Stepped care for alcohol problems: an efficient method for planning and delivering clinical services. In *Changing Addictive Behavior: bridging Clinical and Public Health Strategies* (eds J. A. Tucker, D. M. Donovan and G. A. Marlatt), pp. 331–343. New York: Guilford.
- Thomson, A. H. and Brodie, M. J. (1992) Pharmacokinetic optimization of anticonvulsant therapy. *Clin. Pharmkin.*, **23**, 216–230.