International Congress Series 1220 (2001) 67–78

PK/PD approach to dose selection

Lewis B. Sheiner

University of California San Francisco, Box 0626, UCSF, San Francisco, CA, 94143-0626, USA

Abstract

The utility of a given outcome is the subjective value of that outcome. Decisions differentially affect the probabilities (not the utilities) of possible outcomes. The expected utility of a decision is the average utility of all possible outcomes consequent upon that decision, each one weighted by its probability under the decision. The optimal decision maximizes expected utility. Whilst utilities are subjective, the probabilities of outcomes are objective. Thus, the goal of empirical investigation of dose response should be to provide the particular probability distributions of outcomes as a function of dosage regimen that are required to compute optimal (dosing) decisions. Given the presence of an indication for treatment, and the decision to treat with a given drug, the probability distributions in question are functions not only of dosage (amount and timing), but also of clinical circumstances, i.e., factors that affect individual pharmacokinetics (PK) and/or pharmacodynamics (PD). Let the set of these distributions be called, collectively, the response surface for the drug. Given a mapping from responses to utilities, the response surface is clearly a sufficient condition for optimal dosage decisions. Just as clearly, however, regulatory authorities cannot require a purely empirical estimate of it, as this would entail studying all practically realizable dosage regimens in all possible clinical circumstances, a manifestly impractical task. Perhaps because of the success of empirical hypothesis testing as a means of establishing drug efficacy, that same paradigm has been applied to regulatory requirements for dosing (labeling), with unfortunate results: given that the desired goal is impractical, ad hoc and incomplete strategies have evolved as substitutes. A scientific model-based view, which represents the response surface as parsimonious parametric functions of key PK/PD features, and estimates the surface by pooling data across many different studies despite their different designs resolves the conflict between the empirical demand for data and the decision-theoretic demand for a complete and continuous response surface.

Keywords: Dosage; Response surface; Utilities; Modeling; PK/PD

E-mail address: lewis@c255.ucsf.edu (L.B. Sheiner).

0531-5131/01/$ – see front matter
PII: S0531-5131(01)00287-4

1. Introduction

Dose selection is logically viewed as a decision problem: given a set of circumstances (indication, drug, clinical status of patient) what dosage will yield greatest net benefit? Decision theory (see, e.g., Ref. [1]) provides a formal system for choosing among decision options. Whilst it is not my intent to suggest that physicians or regulatory agencies should advocate formal computation of doses, it is nonetheless instructive to consider the prescriptions of formal decision theory to gain insight into the key information required for good decisions, no matter how they be made.

2. Decision theory

Decision theory starts with the notion of utility. Utilities are subjective scores or values assigned to future states of nature (hereafter, "outcomes") so to produce an interval-valued preference ordering on those outcomes. Thus, the utility of future good health is greater than that of future ill-health. Utilities are more than ordinal: a quantitative difference in utility can in principle be mapped, given a utility on money, to a monetary award to be paid to an individual who suffers a shift from good to poor health by someone held responsible for that shift (this kind of computation is undertaken daily by our tort system). While empirical information may be useful in setting (personal) utilities (e.g., the incremental yearly health care costs in dollars associated with ill-health, in the example above), utilities are essentially subjective (for example, the utility of money: some people care more for it than others, and would sacrifice more time, health, or other goods for it). The utilities of certain future health states, namely those affected by treatment with a given drug, are then a key first element in making dosage decisions regarding that drug. As utilities are subjective, however, they cannot be a matter for detailed regulatory concern, and therefore will not be discussed further. Rather, we must examine the other elements called for by decision theory to find those involving empirical information, as it is regulations and standards for such information that are a legitimate public concern.

Given utilities on all outcomes relevant to the choice of dosage (which will become clear in a moment), decision theory dictates that the best decision is the one that maximizes expected utility (EU). The EU of a particular choice is the average utility of all possible outcomes consequent upon that decision, each one weighted by its probability under the decision.

Formally, for decisions D_i, $i=1, \ldots$, outcomes $Y_j, j=1, \ldots$, and utility function $U(\cdot)$, the optimal decision, D_{opt} is given by

$$D_{\text{opt}} = \arg\max_i(\text{EU}(D_i)),$$

Where

$$\text{EU}(D_i) = \sum_j Pr(Y_j \mid D_i)U(Y_j).$$

The term $Pr(Y_j|D_i)$ above is the probability of outcome Y_j given decision D_i. Different choices (decisions) can differentially affect EU only via differences in these probabilities:

any j such that $Pr(Y_j|D_i)$ is independent of D_i indexes an outcome unaffected by the decision, and hence one that is irrelevant to it.

The definitions above make clear, at least in theory, what are the legitimate obligations of drug developers and regulators with respect to dosage selection: the developers must enumerate the relevant Y_j (i.e., identify outcomes—efficacies and toxicities—causally connected to drug use), and estimate $Pr(Y_j|D_i)$ for dosage decisions D_i (combinations of amount, route, timing—that is the entire prescribed time-course of drug administration), whilst regulators must assure that all relevant outcomes have indeed been enumerated, and that the estimates of $Pr(Y_j|D_i)$ have sufficient precision such that D_{opt} is essentially invariant under all plausible remaining values of $Pr(Y_j|D_i)$.

3. A simple example

To make matters concrete, imagine the simplest possible case that retains realistic features. Let a complete treatment regimen consist of a single administration of a drug (e.g., a vaccine), and let the dosage choices be only the magnitude of the single dose. Further, let there be only two outcomes, binary efficacy (yes/no) and binary toxicity (yes/no). Let the (dis-)utility of toxicity equal that of efficacy, but with reverse sign (this is not as arbitrary as one might think: for example, for the vaccine example, efficacy might be preventing a potentially lethal infection, whilst toxicity might be causing a potentially lethal adverse reaction). In this simple case, the EU of a given dosage is simply the probability of efficacy at that dose less the probability of toxicity at that dose.

In this simple case, one might imagine that a straight-forward approach to dose-finding could be used: perform a parallel dose clinical trial with all reasonable doses and estimate the two probabilities in question for each dose from the results. The simple example just given allows the following two observations on dosage selection.

3.1. The goal of "dose selection" is not suggested doses

The trial information relevant to dosing is the estimates of the key probabilities $Pr(Y_j|D_i)$, not suggested doses per se. Conceivably, two individuals could value the two outcomes unequally. For them, the optimal doses might differ. Both, however, could compute their personal optimal dose given their personal subjective utilities and the common objective outcome probabilities given the doses. Further, the probability of being exposed to the disease against which the vaccine is a prophylaxis might differ from region to region or, being exposed, the probability of becoming infected might depend on personal habits (e.g., smoking). The EU will be less at any given dose in circumstances with low risk than in one with high risk since $Pr(\text{efficacy}|\text{dose}) = Pr(\text{efficacy}|\text{dose, exposure, infection, death}) \cdot Pr(\text{death}|\text{infection}) \cdot Pr(\text{infection}|\text{exposure}) \cdot Pr(\text{exposure})$. The optimal dose for one region or person is not necessarily optimal for another due to these objective reasons. As one cannot anticipate all utility functions, nor all risk circumstances, an important first point emerges: developers and regulators should not attempt to choose doses, they should attempt instead to obtain and provide the objective information (probabilities and outcomes) that will allow individuals who may differ with respect to

utilities and risks to make optimal individual dosage decisions based on those different utilities/risks.

3.2. The "curse of dimensionality" prevents a strictly empirical approach

The parallel dose trial cannot in fact uncover all the objective information (probabilities) required for optimal dosage decisions. Such a purely empirical approach is doomed to failure due to the "curse of dimensionality". Just as disease risk may vary from region to region, or with variation in personal habits, as above, so too will the probability of efficacy and toxicity at a given dose vary with individual pharmacokinetics (PK), which determines drug exposure given dose, and pharmacodynamics (PD), which determines responsiveness given exposure. Thus, even for the simple example above, the apparently simple univariate probability functions, $Pr(Y_j|D_i)$ are really multivariate functions $Pr(Y_j|\text{PK, PD, } D_i)$, denoted "response surfaces" hereafter. The "curse of dimensionality" as it applies to dose selection is this: no practical (set of) parallel dose trial(s) can possibly enroll enough individuals in each cell of a $d \times q \times r$ table of d dose magnitudes by q PK "classes" by r PD "classes" (even granting the radical simplification that both of the latter "variables" are categorical and univariate, rather than continuous and multivariate, as they in fact are) to estimate outcome probabilities as precisely as will be required for optimal dosing decisions. I believe that in the face of this apparent impasse, developers and regulators have retreated to ad hoc and incomplete empirical approaches rather than recognize that it is unsophisticated empiricism that is the problem, not the magnitude of the task. In the remainder of this paper I shall attempt to show how a scientific model-based approach to data gathering and analysis can provide a more complete solution.

4. Learning (estimation) vs. confirming (testing)

Several years ago, I observed [2] that drug development could be seen as two cycles of a general paradigm for the advancement of empirical sciences, first outlined by GEP Box [3], and which may be called the "learn/confirm" paradigm. In this view, one first undertakes empirical exploration with the goal of greater understanding ("learning"), and then one exposes the conclusions drawn from this learning phase, to rigorous empirical test ("confirming"). Both phases involve interrogating reality, that is, doing experiments that involve fixing conditions and observing outcomes, but the designs, analyses, and goals of these experiments differ markedly. The goal of learning is a predictive model of the outcomes of interest as a function of both controllable inputs (decisions, or actions) and uncontrollable inputs (covariates). This description clearly fits the goal of dose selection as presented above. The goal of testing is to (fail to) falsify a specific hypothesis, thereby establishing increased credibility for it.

Learning experiments select units for study that vary with respect to covariate values, administer varying dosages, comedication, etc., and carefully and serially observe the outcomes of interest. In the analysis phase, a model for the data incorporating prior knowledge in the form of specific assumptions is proposed and its free variables (parameters) are fit to the data. The number of parameters in the model is usually kept

small (a rule of thumb is that there be no more than one parameter per 10 experimental units studied) so that reasonably precise estimates of them can be obtained. This parsimony with respect to parameters does not reflect a similar parsimony with respect to model domain/range: the model typically relates a considerable domain of inputs to a likewise considerable range of outputs (both domain and range often include the entire real space of dimension equal to the product of the dimensions of the input variables and those of the output variables, respectively). A simple example makes the point: a linear relationship between x (input) and y (outcome) delivers a prediction of any y for any x using just two parameters (slope and intercept). The choice of a linear relationship is dictated by (and must be justified with reference to) prior knowledge of the subject-matter science. The analysis mode is likelihood or Bayesian (see, e.g., Ref. [4]), that is it conditions on the data and estimates a complete probability model for it. Conditioning on the data usually implies conditioning on the actual design (i.e., using the "as treated" principle), rather than on the intended design (i.e. the "intention to treat" principle). If the two designs differ considerably, then an additional model for the actual design given the intended design (and other covariates) may be required (for more discussion of this, see Refs. [5–8]).

Confirming experiments select homogeneous units for study and administer as few different actions as is compatible with the test to be undertaken. Only a few outcomes are recorded, and these are observed, usually, only at the beginning and end of the trial. In the analysis phase, credibility is assigned to a null model that asserts that outcomes are independent of treatment assignment and that is as free of assumptions on the data (i.e., on the real world) as possible (in the simplest of circumstances, it can be entirely free of such assumptions). The analysis mode is frequentist, that is it conditions on a (null) model for the data, and computes the probability of the data under that model. Inference depends primarily (exclusively if possible) on the probability model on the data induced by the hypothesized model and the design (usually the randomized assignment), and is typically according to the "intention to treat" principle. The strict empiricism of the confirming approach, eschewing to the extent possible all assumptions about how the data are generated by real-world "causes" (i.e. exactly what the learning phase seeks to elaborate), is justified by the testing goal: doing so avoids the inferential circularity that would result if one assumed the very things one sought to test.

Thus, learning builds on past experience to extend tentative knowledge beyond the current level, whilst testing provides an objective validation check on those new learnings. Learning is grand and encompassing, and draws powerful and sweeping conclusions. The price is that these conclusions and predictions must remain tentative and uncertain, depending as they do on an extensive network of assumptions. In contrast, confirming is certain, but of limited scope, usually providing just one bit of information per confirming trial—"hypothesis falsified" or "hypothesis not falsified".

5. Dose selection through learning

Applying the above background to the dosage selection problem, a solution appears to be in sight: if one can tolerate the uncertainty inevitably associated with learning-based

models (i.e. scientific models of PK/PD), then the curse of dimensionality is lifted. Parsimonious but powerfully predictive models of the response surface may be available from experiments of practical size by using study design and analysis methodology suitable for learning, rather than confirming.

Before discussing further how learning designs and analyses can lift the curse, it will be useful to consider whether we can accept learning's uncertainty "price" in so important a matter as dosage decisions. The arguments for this proposition are as follows. First, the performance of the current dosage selection system is so poor, and pre-approval dosage selection so uncertain, that the net uncertainty associated with a principled learning approach is almost certain to be less, not more, than that tolerated at present. Based on an extensive survey by the Georgetown Center for Drug Development Science of 354 evaluable drug labels out of 499 drugs approved in the US during 1980–1999, it appears that formal labeling changes for dosage (80% were dosage reductions from initially approved doses) occurred in about 25% of new chemical entities within the first decade after they were approved (J. Cross, H. Lee, C. Peck, personal communication, 2000). One may take this figure to be a downwardly biased estimate of the actual probability that incorrect dosage information is provided by the pre-approval drug development process, as the FDA requires changes in labeling only if the current dose appears to be unsafe. Such action will not be triggered by doses that are discovered, post-approval, to be above those required for full efficacy, but that are not associated with clear evidence of excessive toxicity. Excessive doses are an almost inevitable result of the current drug development process that seeks in early phases of drug development to study the largest safe doses so as to elicit as strong an efficacy signal as possible (for testing), and then does little in later development phases to search for smaller yet satisfactory doses.

Nonetheless, if most approved drugs are safe and effective, which we have little reason to doubt, then no great harm is being done by today's very uncertain (and upwardly biased) dosage selection process, and perhaps then there is no reason to strive to make that process less uncertain. There are several counter-arguments to this proposition. First, excessive doses must increase the risk of toxicity, however slightly, and the lack of "bodies in the streets" does not mean that more subtle harm is not being done. One hopes, by making dosage selection better, if still imperfect, to avoid some few but very costly errors (seriously dangerous initial dosage for some unusual individuals), and to make a modest improvement on the average, with attendant modest savings in time, money, and discomfort on an individual level, but perhaps appreciable savings across a population. Second, since most drugs are priced by the milligram, excess dosage at the time of approval inevitably leads to serious revenue loss when the correct and smaller effective dose is ultimately discovered post-approval. While it is not immediately apparent that this constitutes a problem for any but the shareholders of pharmaceutical firms, there are potential public costs. To the extent that missing the right dose at the time of approval increases downside profit risk, pharmaceutical manufacturers must set their prices or market targets (or both) higher to compensate. The first adds public expense, and the second deprives the public of effective remedies for indications of low prevalence.

A second reason to accept some (but not excessive) uncertainty in the dose selection process is this: at the level of treatment of individuals, considerable uncertainty in initial

dosage is inevitable as inexplicable inter-individual kinetic variability alone is often on the order of magnitude of 50% or more [9]. Thus, dose titration is a sensible and routine part of medical practice, limiting any harm that error in initial dosage may cause. This observation supplies the same support as above for leaving the current imperfect system in place, but the same rebuttal also applies.

6. How modeling helps

Thus far, I have attempted to establish an epistemological framework for thinking about the empirical information requirements for rational (i.e. justifiable within decision-theory) dosage selection, and have asserted that these requirements might be met using learning studies/analyses, as opposed to the current manifestly inadequate approach of using confirming studies/analyses. In this last section, I try to justify that assertion.

As I have already mentioned, first and foremost, models provide a parsimonious representation of the input–output relationship (system) for which predictions are desired, and, if they are securely based in valid scientific knowledge, some credibility for interpolation between extant data and extrapolation beyond it. They do considerably more than this, however: most notably, they markedly enhance investigative efficiency by (i) increasing the amount of information recoverable from any given set of data, and (ii) allowing principled merging of data from many sources and designs.

Modeling allows more information to be extracted from a fixed quantity of data by turning noise into signal. Information is synonymous with variation: the total amount of variation in a set of drug-response data is fixed. This is an upper bound on the "information" the data contain. Not all variation is "information" however. Total variation can be regarded as the sum of two parts. The first is the part that correlates (to use the word non-technically) with variation in input, especially controllable input such as dosage. This is called "signal". The second part is the part that's left over, the part that cannot be linked to input variation. This is called "noise". "Information", roughly speaking, is proportional to the ratio of signal to noise. Indeed, it is the goal of modeling and estimation to find the functional forms and constants (parameters) that capture the signal as fully as the noise permits. As total data variation is fixed, models can only increase information by transferring noise to signal, and this is precisely what they do. If, for example, all variation in output ascribable to variation in dose input pattern is ignored (as it is in an ITT analysis), then all such variation is noise as far as that analysis is concerned. If actual dose input is recorded, and output is linked to this, then, to the extent that output varies in response to actual doses ingested, noise is moved to signal, and one learns about dose–response. This observation, by the way, explains why confirming designs generally avoid the wide variation in subjects and inputs that would also make them suitable for learning: since the analysis will not allow assumptions on, for example, the relationship between age or dosage and response, all covariate and un-randomized input variation create noise, not signal, thereby reducing study power at any given size.

Modeling allows data of different precision, design and quantity to be merged. The key idea here is that of "exchangeability". Data are exchangeable if the joint probability

model for all of them is unchanged under a permutation of data indices; that is, exchanging the position of one datum for that of another. It is usually appropriate to model exchangeable data as independent and identically distributed (iid). These two properties imply that each datum's contribution to the goodness of fit criterion (the data likelihood, for example) conforms to the same model form indexed by the same parameters. For iid data, the only thing that differs from the contribution of one datum to the next is the value of the observables (i.e. the inputs and the outputs). Explicit reference to covariates is crucial to establishing exchangeability. If the time of observation of an output is not an explicit part of the model, then two observed outputs are not exchangeable, i.e. do not necessarily arise from the same model. When the model correctly incorporates the effect of time on output, however, then conditional on time the two observed outputs are exchangeable (arise from the same model); only the variable, time, differs in the likelihood contribution, not the value of any parameter. Thus, does each iid datum contribute an independent additive increment of information to the growing knowledge of a single flexible model. To achieve exchangeability, models must deal not only with variation in design (inputs), but with other more statistical features such as random inter-individual variability not apparently related to input values (handled by so-called hierarchical models [4]), and so-called heteroscedasticity, that is differences in precision of observations (handled by explicit modeling of residual variability). For a fuller discussion of these and other statistical issues raised here, see Sheiner and Steimer [10]; an instructive example of the use of modeling to design dosage can be found in Sheiner et al. [11]. It nicely illustrates the points presented in this section.

7. Conclusion

A principled intellectual framework for drug development defines the information required to assure the public that a given drug will, under specified conditions of use, likely lead to net positive utility. Certain parts of this information structure (the totality of which may be denoted the drug label) require greater certainty than others. Great certainty is neither possible nor necessary for that part of the label dealing with individual-specific dosage. It is not possible because the number of variables interacting with dosage to yield response is so great that the "curse of dimensionality" prevents strictly empirical estimation of the (probability) relationships among them, the so-called response surface. It is not necessary because most drugs have a fairly wide therapeutic range, and because individual-specific feedback dosage adjustment can limit any harm done by inappropriate initial dose selection. This freedom allows one to advocate taking a scientific model-based ("learning") view of dose selection, a view that gains its power in generality and efficiency by sacrificing unequivocal certainty. Such a view implies, in turn, that dose selection studies should be inclusive with respect to both patients and dosages, and be analyzed under a full set of scientifically valid assumptions, according to a likelihood or Bayesian inferential mode. If this new paradigm for dosage selection is adopted, we may confidently expect ever less expensive drug development programs to yield ever safer and more efficacious regimens.

Acknowledgements

The work on which this manuscript is based was supported primarily by NIH (USDHHS) grants GM26676 and GM57183.

Appendix A. Discussion 6

A. Breckenridge: A topic that causes huge frustration in drug regulation is the problem of dosing in children. One often thinks there's a kind of conspiracy to complicate this problem. What is the contribution of modelling to rationalise this?

L. Sheiner: Essentially what modelling does is decrease the demands on the amount and structure of the data. I suppose at some point we might, completely theoretically, be able to apply directly what we learn in the adult to the child. While this goal appears attainable with respect to pharmacokinetics, as it is simply a scaling problem, the pharmacodynamic side may be a bit more of a problem. For now I would say that what we might consider as a strategy for paediatric extension of drugs studied only in adults is to use the drug in children as best as physicians think they can to start with; monitor that use, and use this observational data to improve understanding. Call this phase IV, if you will. Now observational data has serious problems; for example, one tends to observe most intensively those people who are the most problematic, and less the people who are not; this leads to biased "designs", making causal inference very difficult. But by adding a few extra observations without invading children too much, these inferential problems can be mitigated. Nonetheless, one still can't take the (ideal) same number of observations in everyone; one has to go along with the treatments that seem to be beneficial, and so on. All those things limit the control one has over the experimental design, and this can only be dealt with through adding assumptions about how these differences affect the data. Models provide the implementing methodology for these assumptions; but only science can supply the assumptions themselves. Modelling, then has a very natural application to the problem of extending pharmaceuticals to the paediatric population, but to realize that application it would require a change in our drug development system; an acceptance that we won't necessarily know the dose very well in children at the time a drug is approved; that we'll learn that later, through structured observational studies. Even this will require a system for data gathering, and a willingness to accept basing our conclusions on scientific assumptions as well as empirical data.

N. Holford: Using the ketorolac example, you didn't really bring in the issue of variability too much. The response surface appeared to be essentially without variability in it. How does variability come into the response surface?

L. Sheiner: The response surfaces you saw in those four pictures were probabilities, so they are inherently variable. The probability that pain relief will be greater than a given level may be 80%, but any individual will or won't have such relief with that probability. If the outcome were not a probability of an event, but the value of a variable such as a blood pressure, there would indeed be an "error" bar around it, to express variability.

N. Holford: It was that kind of variability I was asking you to comment on; clearly, that was a probabilistic model, and very unusual in that regard; but if it had been blood

pressure instead of pain then you could've had a response surface on blood pressure and then perhaps you could translate that into a further probability type of surface when we're talking about the probability of a particular blood pressure response.

L. Sheiner: Right, it would be a distribution and it would be very hard to draw the picture. That three-dimensional picture I showed you would have to have uncertainty zones around the surface to represent individual variation.

P. Joubert: One of the problems I see, also with all the experts I have in my group, is a major focus on the concentration effect relationship for the favourable dynamic effect with perhaps less interest in the unfavourable one. Here I think, probabilistic approaches have been very valuable. I am talking about very simple logistic regression on a simple binary response for an adverse event. There are a couple of examples where prospectively we defined very clearly the unwanted adverse event that we could not tolerate on the market. Then by defining a minimal probability that would be acceptable, and relating that to the concentration effect relationship for efficacy, it helped us to make decisions. So I am hooked on probabilistic approaches, particularly for safety issues.

L. Sheiner: In PK/PD modelling we started with continuous variables because they appear to have such high-information content. At first blush a continuous number that is measured reasonably well, say with two significant digits, has about eight bits of information, whereas a binary response is just one bit of information. That means that you need eight times as many patients with binary data vs. continuous data to achieve the same precision of estimation of a parameter, clearly favouring the use of continuous data. The problem with this view, first of all, is that many clinical responses come as time to event data, or binary data, or categorical data (as did the pain relief data I discussed). We've just got to accept that that's the way these data arise and make the best use of them. This is an instance of my first law of modelling: make the model fit the data, not the data fit the model. Secondly, perhaps we are being fooled about continuous variables anyway, except perhaps when they are carefully measured pharmacological responses or biomarkers, as Meindert presented. I say this because the reality of biological variability may mean that even a two-digit precision number may reproducibly indicate only five or six distinct categories—very high, medium-high, high, etc. As we in PK/PD try to deal with non-continuous data, we are aided by the fact that there is considerable experience with such data in other fields, and hence well-developed statistical procedures for doing so. The trick is to find the link between the pharmacology and the outcome; for example, in an analysis of time-to-event data that Eugene Cox, Stuart Beal and I did, and that will be coming out in JPP shortly, the link is something called the hazard function, the instantaneous probability of the event occurring. We thought that it was the natural place to think of the drug acting. I think the modelling technology is there, the statistics are there, and what we need to focus on is the science: writing the right kind of models relating exposure to the right kinds of intermediate variables (here so-called "indirect" PD models will prove useful) that then generate the observable responses. A remaining problem, however, is the low information content of binary or time-to-event responses. This means that one can't build very complex models, and balancing complexity against mechanistic realism may prove to be an interesting challenge.

M. Reidenberg: Back to your original question on modelling, I think we'd covered the issues of genetics well. The other thing to look at is the environment, and the examples I

can give are digitalis toxicity, where we know that you've got to look at potassium levels and calcium levels to relate concentration to effects; for benzodiazepines, we know alcoholics are very resistant to the effect. Another factor is age and the whole question of homeostasis or as George Shriner used to say, homeostenosis, as we get older and older. An elderly adult with a particular condition, a tumour or anything else, can't react to an adverse event or a stress the way a young healthy person can. It will be somewhat difficult to know how to model that, but doable. And I think, as Lew and I were talking, the big thing that the modelling requires is that you think through explicitly what you intuitively feel, and then try to weigh it. I think that as we look at these various variables, it's possible to articulate them, identify them and even to a certain extent weight them in a clinically meaningful way.

G. Levy: With many classes of drugs, but I think particularly about anti-hypertensives, it is customary to start with a low dose because of concern for postural hypotension, and so on. But that's also true for a number of other drugs. Is it more rational to start with a modal dose, in other words a dose that is more likely to be appropriate for a particular patient rather than start everyone with the lowest dose and work your way up?

L. Sheiner: This is a clear example of where decision theory might help. At least it would give you a clue as to how to approach that question. The best example that I know of in this regard is phenytoin. Because of its non-linear metabolism, there will be some people who start at a reasonable dose, but who will develop very high levels, and perhaps become toxic. So it makes sense to start everyone at a dose lower than the one that would give the typical individual a level in the middle of the therapeutic range, which would be the optimal starting dose if the drug level distribution were symmetric (and the "cost" of under and over-dosing were equal). Indeed, practice has evolved so that we do start phenytoin out at about the dose level that will produce a modal concentration at the low end of the therapeutic range, not the middle. In contrast, for a drug that's almost non-toxic, like penicillin, we routinely give a hundred times more than we need, and that also makes sense in terms of decision theory, here because the cost of under-dosage (inefficacy) so far exceeds the cost of over-dosage (only monetary). In this regard, we do not, generally, think about the actual costs: the cost of under-dosage for a chronic condition, for example, is not inefficacy, but rather simply a delay in efficacy, as we will titrate the dose upwards if we fail to see an adequate effect. Perhaps taking a somewhat more formal decision theoretic approach to dose ranging might help by alerting us to the relevant considerations: the probabilities and costs of various outcomes as a function of dose.

M. Reidenberg: I was going to say basically the same thing that Lew said: that with your example of hypertension, there's no rush, and the side effects are a lot worse than a little delay. With an infection, for example, with gram-negative sepsis, we'll start people with an average dose of an aminoglycoside, accepting that we're going to get some nephro-toxicity but that the antibiotic will enable them to survive the infection.

References

[1] D.B. Petitti, Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine, 2nd Edition, Oxford University Press, New York, 2000.

[2] L.B. Sheiner, Learning vs. confirming in clinical drug development., Clin. Pharmacol. Ther. 61 (1997) 275–291.
[3] G.E.P. Box, Use and abuse of regression, Technometrics 8 (1966) 625–628.
[4] M. Davidian, D.M. Giltinan, Nonlinear models for repeated measurement data, Chapman & Hall, New York, 1995.
[5] J.D. Angrist, G.W. Imbens, D.B. Rubin, Identification of causal effects using instrumental variables, J. Am. Stat. Assoc. 91 (1996) 444–472.
[6] D.B. Rubin, Comment: Neyman (1923) and causal inference in experiments and observational studies, Stat. Sci. 5 (1991) 472–480.
[7] D.B. Rubin, Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism, Biometrics 47 (1991) 1213–1234.
[8] L.B. Sheiner, D.B. Rubin, Intention to treat analysis and the goals of clinical trials, Clin. Pharmacol. Ther. 57 (1995) 6–15.
[9] L.B. Sheiner, T.M. Ludden, Population pharmacokinetics/dynamics, Annu. Rev. Pharmacol. Toxicol. 32 (1992) 185–209.
[10] L.B. Sheiner, J.-L. Steimer, Pharmacokinetic/pharmacodynamic modelling in drug development, Annu. Rev. Pharmacol. Toxicol. 40 (2000) 67–96.
[11] L.B. Sheiner, S.L. Beal, A. Dunne, Analysis of non-randomly censored ordered categorical longitudinal data from analgesic trials (with Discussion), J. Am. Stat. Assoc. 92 (1997) 1235–1255.