Decision-making in refractory epilepsy: an analysis based on computer simulation

CHRISTOPHER D WARD

From the University Department of Clinical Neurology, Radcliffe Infirmary, Oxford, UK

SUMMARY A program developed to simulate management of patients with recurrent seizures was tested in 24 subjects, including 13 experienced neurologists. Identifiable errors included excessive reliance on plasma anticonvulsant levels, misinterpretation of random effects, inefficiency in selection of follow-up intervals, and undue reference to arbitrary data, such as the initial dose, in the selection of “optimum” drug dosage.

One of the most burdensome problems of outpatient neurology is the management of refractory epilepsy. Many of the variables impeding drug response such as spontaneous changes in the severity of the underlying disorder, and poor patient compliance are outside therapeutic control. The irrationality of the patient is often proved, and still more frequently suspected, but that of the doctor has rarely been impugned. Experimental manipulations of variables such as patient compliance are rarely possible in clinical practice but are readily amenable to computer simulation. The performance of clinicians faced with an inherently solvable problem has an obvious bearing on their approach to the complexities of real life.

Methods

Between 15 and 24 subjects (termed “doctors”) took part in the various studies. They included 13 experienced neurologists (seven consultants or American board-certified neurologists from three centres, and six career-grade registrars and senior registrars, comprising Group I); and 11 individuals with less experience of epilepsy management (seven qualified doctors with largely general medical experience, and four medical students, comprising Group II). The doctors were asked to “manage” four fictitious patients. Each patient was taking either phenytoin or carbamazepine. No relevant clinical details, such as seizure type, were supplied. Doctors were instructed to alter the dose of a given anticonvulsant until its benefit was deemed maximal, and thus to establish an “optimal” dose for each patient. It was suggested that the purpose was to gain adequate seizure control or, if that was not possible, to establish the therapeutic efficacy of the specific drug as a preliminary to some future decision (for example to substitute or add another drug). They were told that the optimal dose might be nil or more in different patients. They were asked to minimise visits to the clinic, blood tests, toxicity, and seizure frequency. They were warned that non-compliance was possible. At each clinic “appointment”, a seizure diary supplied the dates of all seizures, or of the first 10 if the frequency was high, and clinical toxicity, if any, was automatically reported. Before a plasma level could be obtained, doctors were asked to judge the patient’s recent compliance, and to assess non-randomness (clustering) of seizures. This arrangement corresponded with the tendency of clinicians to form judgements during the outpatient interview, prior to blood tests. The plasma level was then available, if requested. The “therapeutic” range was quoted in parentheses beside each plasma level result (40–80 μmol/l for phenytoin; 25–50 μmol/l for carbamazepine.) Finally, treatment changes were made and a follow-up interval was selected.

The program was written in BASIC for the portable Epson HX-20 computer with 16K memory expansion. Program design is shown in the fig. Characteristics of patients are shown in table 1. For each patient, there was a functional relationship, based on published data, between dose and plasma level. Plasma level was subject to random fluctuation (standard deviation 5 μmol/l). Seizure frequency was governed by a Poisson process. Although small deviations from the Poisson distribution were required in order to permit the effect of treatment during a previous follow-up interval to be reflected in the period immediately following a treatment change, there was no marked non-random clustering (for example in relation to menstruation). The mean frequency of the Poisson process was related to plasma level by an arbitrary patient-specific function. For Patient 1 the relationship was parabolic, so that high doses induced frequent seizures. For Patients 2a and 2b, the relationship between plasma level and seizure frequency was linear. They differed only in the dose of anticonvulsant which they were receiving at the start of the management period. In Patient 3 there was no relationship at all between anticonvulsant dose...
and seizure frequency. Compliance was complete in all patients.

The method used to give criteria for inappropriately short, or inappropriately long follow-up intervals utilised the data available to the doctor at the point when the interval was to be selected, and assumed a Poisson distribution. One estimate of expected seizure frequency was obtained by dividing the total number of seizures which had occurred to date by the total length of follow-up. Another estimate was obtained by assuming that the mean frequency would remain the same as that observed during the most recent follow-up period. Using both estimates, the probability of no seizures, or of more than three seizures occurring, was then calculated for the follow-up interval selected by the subject. A “short” follow-up interval was defined as one within which there was a 90% or greater chance of no seizure occurring. A “long” interval was defined as one within which there was a 50% or greater chance of four or more seizures occurring.

Results

Results for Groups I and II are described together. Those for Group I are summarised in Table 2.

1 Judgement of randomness Although no non-compliance or clustering occurred, these were suspected on 15–24% of clinic visits, as a result of misinterpretation of randomness.

2 Use of blood tests and induction of clinical toxicity Total remission was unobtainable in any patient, and therefore the choice of optimal dose was presumably influenced by evidence of toxicity. Heavy reliance was apparently placed on plasma level estimations, which were requested at 65–95% of appointments, although clinical toxicity was only obtainable above the stated upper limit of the “therapeutic” range. By contrast, less use was made of clin-
ical toxicity as a guide to the maximum tolerated dose. In one patient (with a relatively high seizure frequency) most doctors (including 10/12 in Group I) selected the highest dose below that which had induced clinical toxicity. In another patient, despite the persistence of two seizures annually when treatment was optimal, only three of nine Group I doctors induced clinical toxicity.

3 Selection of follow-up interval By the conservative criteria defined above (see Methods), unduly short intervals were selected at 15–38% of appointments in patients with seizure frequencies of two per year or less, but in only 4–6% of appointments for Patient 3 with six seizures per year despite treatment. Unduly long follow-up intervals were uncommon (0–6%) when seizure frequency was low (Patients 1 and 2), but common (14–45%) with a higher seizure frequency Patient 3).

4 Selection of optimal dose Doctors tested the therapeutic effect of the drug by making successive changes in dose, usually increments. In Patients 1 and 3, the optimum dose could be established objectively. For Patient 1, the minimum seizure frequency without clinical toxicity occurred at a phenytoin dose of 425 mg. The median dose selected by doctors was 400 mg (mean 390 mg, SD 39.9). For Patient 3, with no relationship between dose and seizure frequency, the optimum phenytoin dose was 0 mg. The median dose selected by the doctors was 325 mg (mean 282 mg, SD 106). Only three doctors (including two of 10 from Group I, and one medical student in Group II), selected 0 mg. The maximum tolerated dose of carbamazepine for Patients 2(a) and 2(b) was 2000 mg, and since they had a linear response to increasing doses, with low seizure frequencies above 1000 mg, “optimal” dose was to some extent a matter of subjective judgement. For Patient 2(a), with an initial dose of 200 mg, the median optimal dose was 800 mg (mean 821, SD 422). For Patient 2(b), with identical characteristics except that the initial dose was 1200 mg, the median optimal dose selected was 1400 mg (mean 1478, SD 216) (p < 0.05, median rank test).

5 Comparison of Groups I and II Ignorance of pharmacokinetics, with inappropriate dosage increments, was confined to Group II, but all other errors were frequently represented in both Groups, with little tendency for the more experienced Group I doctors to be exempt. In both Groups, misjudgements of compliance and clustering were equally frequent. Blood tests were requested at most appointments in both Groups, although less frequently in Group I (65–75%) than in Group II (84–95%). Frequency of episodes of clinical toxicity was almost identical in the two groups. Mean follow-up interval lengths did not differ significantly (t test). The frequency of inappropriately long or short intervals was higher in Group I (40% of appointments) than in Group II (17%). There were no significant differences between the median optimum doses selected by the two Groups.

Discussion

The study shows that clinicians, including those with neurological experience, are liable to certain specific errors when making quantitative judgements. These include, first, errors of interpretation of data. One such error is over-interpretation of random patterns, as when a pure Poisson process is perceived as forming a meaningful pattern (non-compliance; clustering). Other errors stemming from misreading of random data include failure to recognise trends such as dose-response relationships, and failure to establish a rational relationship between mean frequency and follow-up interval. A second important class of error is reference to external assumptions rather than to internal evidence. One such assumption was that the “therapeutic” range quoted with each plasma level result would predict desired and adverse drug effects. In Patient 1, the frequency of requests for plasma measurements presumably reflected the assumption that phenytoin toxicity would occur above the quoted upper level of 80 μmol/l. Such reluctance to subject a patient to reversible toxicity might seem justifiable, but the result of this policy was the selection of a suboptimal dose, since Patient 1, like many real patients, could tolerate 90 μmol/l without ill effect. Reliance on plasma level estimations cannot be defended in Patient 3, for whom the concept of a “therapeutic” range had no practical meaning. An empirical approach was undoubtedly preferable as a means of establishing that the drug was of no benefit.

“Anchoring” of quantitative judgements to irrelevant data has been recognised as a characteristic error in other contexts. In Patient 3, the initial dose of 200 mg supplied an anchorage which deterred clinicians from recognising that the true optimal dose was as low as 0 mg. Similarly, the initial dose of carbamazepine acted as an irrelevant reference point or anchorage for the selection of optimal doses in Patients 2(a) and 2(b). If doubts about the toxicity of high dose carbamazepine were operative in Patient 2(a), they should have led to a reduction in dose in Patient 2(b). If there was no perception of anticonvulsant activity in those patients, then there was no justification for increasing the dose and establishing an optimal dose as high as 1400 mg in Patient 2(b). In summary, most errors stemmed from insufficient reliance on empirical data.

The lack of important differences between experienced and inexperienced subjects is not surprising if it is accepted that fallacies in human judgement stem
largely from inherent human limitations,\textsuperscript{2} to which even trained neurologists are subject. It is surprising that the Group I doctors made heavy use of plasma level estimations, rather than using empirical criteria such as reversible clinical toxicity. There was no evidence that plasma level measurements contributed to the efficiency of management. One theoretical option, to establish a dose-response relationship with two or more closely-spaced plasma samples, as a preliminary to a period of follow-up at a (hopefully) therapeutic level, was only utilised by one doctor, in Group II.

Before conclusions can be drawn from this study, one important question is: did the simulation facilitate, or obstruct, rational decision-making? The model was certainly an oversimplification, but this resulted in conditions largely conducive to more efficient management: seizure occurrence conformed to a Poisson distribution; all data were presented in a standard format with obligatory reporting of clinical toxicity; all seizures were reported; and compliance was full. Although the relationships between dose and plasma level were plausible, those between plasma level and seizure frequency were undoubtedly idealised, except in Patient 3. The excessive ease with which plasma levels could be measured resulted in redundant information, but was not otherwise misleading. None of these factors would have reduced the possibility of efficient management. Most doctors regarded the program as a reasonable approximation to clinic conditions.

A more radical departure from reality was the total absence of clinical details. The advantage of this design was that one important aspect of epilepsy management could be studied in isolation: the empirical judgement of quantitative relationships. In general, the value of clinical clues in the prediction of response to treatment is variable, and often it is negligible. All the patients in this study were fully compliant and showed no non-random clustering of seizures. Most patients claim to comply fully with treatment, and clinical information would have been unlikely to reduce the frequency of false suspicions of non-compliance. Similarly, any seizure type may, in principle, occur at random intervals with no tendency to meaningful clustering: no information was likely to reduce false suspicion of non-randomness. The converse of these statements would not have been true: had any of the patients been non-compliant, or had non-random (for example menstrual) seizures, then clinical information might have been predictive. Clinical history may sometimes explain the failure of seizures to respond to drugs, but it does not necessarily do so, and the absence of information about Patient 3 did not rule out the possibility of non-response. The results of the study have especial relevance to those patients in whom response to anticonvulsants is relatively unpredictable, even when clinical evidence is available. An empirical approach is required for these patients in particular, and is a desirable basis for the management of all epileptics.

Finally, one might ask: was the doctors’ response realistic? Some doctors claimed that their selection of follow-up intervals and doses was less scrupulous than in clinical practice; they pointed out that the computer might encourage a game-playing approach which would have been out of place in an epilepsy clinic. This could have led to unrealistically long follow-up intervals and to large swings in prescribed doses. The data show, however, that the major error was unduly short follow-up intervals. Large (even frivolous) dose changes would have established the lack of effect of phenytoin in Patient 3, whereas the observed selection of optimal doses, as well as the rarity of toxic episodes, clearly reflected the habitual conservatism of the doctors. Some suspected that their professional skill was under scrutiny and admitted that the computer made them anxious to perform well: undoubtedly they would have had similar feelings in an epilepsy clinic.

Doubts about the direct application of these results to the specific context of epilepsy do not diminish the importance of general inferences which can be drawn from them. Many clinical situations call for judgements of randomness, identification of trends, and arbitrary selection of sampling intervals; such decisions are required, for example, in the use of antihypertensive agents, anticoagulants, migraine prophylactics, and any treatments designed to reduce the frequency of relapsing disorders. The errors demonstrated by this study suggest that better use could be made of available data, and that awareness of the shortcomings of intuitive judgements might lead to a reduction of the burden both to patients and to outpatient and laboratory staff.

The cooperation of neurologists at the Radcliffe Infirmary, the National Hospitals for Nervous Diseases, and elsewhere, is gratefully acknowledged.

References
