SLEEP STUDIES IN CLINICAL PHARMACOLOGY

IAN OSWALD

University Department of Psychiatry, Royal Edinburgh Hospital,
Morningside Park, Edinburgh EH10 5HF, Scotland

Introduction

Prescriptions for drugs that alter sleep are, with those for antibiotics, the commonest prescriptions of clinical practice. Some drugs that alter sleep are intended so to do and others, prescribed in the hope of altering mood, will generally alter sleep too, while yet others, prescribed for somatic ailments, may have unsuspected effects on the brain observable during sleep. I shall below take as a primary example the study in man of a drug expected to have hypnotic actions, following studies in animals and preliminary human testing.

There are methods for subjective evaluation and others for objective evaluation. It is the former that are of first interest for clinical practice and, although imprecise where any one individual is concerned, these methods are relatively cheap. In the case of hypnotic drugs, more than any other class of compound, one must take an interest not only in the effects of their administration, but in the effects of their withdrawal, and if one is to conduct relevant studies then one will be influenced by the fact that in clinical practice these drugs will be taken not only as occasional single doses, but often too for consecutive nights during many weeks or even years. During these weeks of intake there will be time for cumulative actions and time for the development of tolerance, with then the certainty of withdrawal effects after the drug ceases.

Subjective evaluation of sleep

A worth-while study must embody the principles of the double-blind trial. The pills containing the active drug or drugs should not only look like the blanks, they should also taste like them. They should preferably be capsules that are properly sealed, made available in individual blister packs for each individual subject, and to cover a period of days. A pharmaceutical company that gives an impression of professionalism in these matters gives confidence to the investigator that they can also be relied upon to have put active and inert pills in those parts of the design that the written schedule says they should have appeared in.

The subjective judgements can be very simple, such as saying that night 2 was or was not better than

night 1 and this is perhaps all that can be asked of ill patients in hospital. Most people who take sleeping pills are, however, not in hospital but living active lives, and greater sophistication can be sought by the use of ratings, for example, into five categories, from sleeping 'excellently' to 'very badly', having 'no dreams at all' to dreaming 'very much and vividly', and so on (e.g. Spiegel, 1973). In this case there should be subsequent use of non-parametric statistics for these rankings by each individual.

Visual analogue scales

Visual analogue scales provide another valuable method. In these the patient or volunteer is presented with a line of 100 mm length across a page, one end bearing a statement such as 'Best sleep ever' and the other end a statement such as 'Worst night ever'. The instructions request him to make a mark in the middle of the scale for what he considers to have been for him a normal night and to make his mark some way towards one or other end of the line if he feels the night was a particularly good one or a particularly poor one. He thus creates his own scale and range of values. Like all methods that rely on introspection and the use of abstract concepts for displaying introspection, the method prospers when there is an average or better level of intelligence and education. It is wise to use any potential subjects in a small pilot study before a main study, as there are occasional individuals, albeit intelligent and educated, who will jump from one extreme of the scale to the other in a seemingly insightful manner. Most people, however, find the method simple and acceptable and will fill in these ratings day after day for many months.

In the use of visual analogue scales for research it is helpful to have a top copy of flimsy but opaque paper and a lower card copy that will be marked simultaneously, bearing a faint scale of numbered mm, so that the point of the mark along the scale can easily be read off and then written into a box on the lower copy, for eventual computer handling. If a reasonably large number of subjects is used then the raw data in mm are suitable for later calculations. If only small numbers of persons have been used, then it means that those who make only small excursions from the

mid-point contribute less to the eventual scores than those who are more ready to make flamboyant excursions. One approach to overcoming this problem is to determine each individual's standard deviation during one sequence of nights, such as placebo nights, and then to express every single night of the study as a percentage of that standard deviation, and to use these percentages for the eventual analysis. Figure 1 shows an example of a study in which 12 people made ratings each morning of the quality of sleep they had experienced the previous night and did so for six weeks, during which they received placebos each night for the first two weeks, lormetazepam 1 mg nightly for the next three weeks, and placebo again for each night of the final week. Statistical analysis, using Friedman's analysis of variance and the Wilcoxon test showed a significant change in subjective evaluation in the direction of better sleep during the drug period, compared with the baseline, and a significant impairment or rebound of sleep quality, to a degree worse than that of baseline during the withdrawal period (Oswald, Adam, Borrow & Idzikowski, 1979).

As a further illustration of the use of such methods, my colleagues and I recently extended them to one hundred middle-aged volunteers, selected because of their subjectively poor sleep, who each took a capsule at bedtime each night for 32 weeks. In the first 4 weeks the capsules were blanks for all subjects, as they were in the final 4 weeks. In the other 24 weeks, 50 of the subjects received lormetazepam 2 mg, 25 received nitrazepam 5 mg, and 25 received blanks throughout. Subjective sleep quality, assessed by a visual analogue scale, was improved by both drugs throughout the whole of the 24 weeks of administration, with a rebound on withdrawal that, on graphical display, appeared to have a duration of about 3 weeks before it ended.

In order to assess the time taken to fall asleep, the scale shown in Figure 2 was used, being a modified log scale. In the case of lormetazepam there was once again a significant improvement throughout the period of drug intake, and both lormetazepam and nitrazepam were associated with a significant lengthening of sleep latency after withdrawal, again persisting for an

Figure 1 Display of data originally derived from visual analogue scale self-ratings of the quality of sleep by 12 subjects, mean age 53 years. Data are expressed as percentages of the baseline standard deviation (see text). Lormetazepam 1 mg nightly significantly improved the subjective quality of sleep but there was a significant impairment compared with baseline following withdrawal of the drug.

Figure 2 A scale that has proved of practical value for the self-rating of how long it takes to fall asleep.

apparent period of 3 weeks. In handling data such as these, graphical display, by 5-point moving means to smooth the curves was helpful in conveying the outcome. In order to obtain a more precise and economical method of handling the data for statistical comparison it had been decided in advance to compare the first 4 days of the fourth baseline (placebo) week with the first 4 days during the fifth week (lormetazepam, nitrazepam, or continued placebo), the first 4 days of the 28th week (lormetazepam, nitrazepam or continued placebo) and the first 4 days of the 29th week, this last being the first week of withdrawal. The mean of each 4-night period for each individual was found and his or her mean value for the baseline period was subtracted from each of the later mean values. The resultant data, being then normally distributed, were used for analysis of variance and Scheffé tests, to examine the null hypothesis that there would be no difference between behaviour in the case of those receiving placebo throughout and those receiving active drug. Visual analogue scales, used as here, can be confidently recommended as a method, there being many reports in the literature of their successful use (e.g. Zealley & Aitken, 1969; Oswald, Lewis, Dunleavy, Brezinová & Briggs, 1971; Bond & Lader, 1974; Hindmarch, Parrott & Arenillas, 1977).

Other subjective judgements

There are of course, other subjective features of sleep that can be sought. In the above-mentioned study of 100 subjects, Yes/No answers were given to whether they felt they had experienced vivid dreams during each night. Again, at the end of the study, the interviewing physician, who had remained blind to the nature of the pill taken during the weeks 5–28, was asked to make a subjective judgement of whether the subject's sleep would benefit by continuing or not continuing on the same medication. In this particular study these subjective judgements by the observer discriminated significantly in favour of lormetazepam. Discriminations such as these could only be hoped for when a large number of subjects had been used. In the same study the subjects completed a 25-item check list each evening concerning the simple alternative of the absence or the presence of such possible side-effects as lethargy, dizziness, irritability, dry mouth, diarrhoea and so on, all side-effects that might be associated with hypnotics, and there was no increase in the rate of any of these subjective experiences in association with either of the drugs. This conclusion that there was no increase was possible, however, only because of the large number of subjects employed and the fact that a placebo was included in the design for continuous intake throughout the whole period.

Daytime skills and feelings

Apart from seeking subjective estimates of unwanted side-effects of hypnotic drugs, it is important to undertake day-time objective testing of as comprehensive as possible a range of performance skills (Hindmarch, 1980). Many of the published studies in this field are unsatisfactory. Too many have reported on studies involving only about six subjects, often much younger in age than those who usually take hypnotic drugs, often after only a single-dose instead of sustained intake over a week or more, and often with perhaps a dozen or more comparisons being the subject of statistical examination, with one or two of these described as having reached the $P < 0.05$ criterion, or all results being reported as negative. In this field, as indeed in all research, it is important first to demonstrate that the instruments of measurement are sensitive. Unless it is shown that tests of the kind used, on the sort of population used, are sensitive to some standard drug, then negative findings carry no weight and sporadic comparisons reaching the $P < 0.05$ criterion are difficult to interpret as other than attributable to chance.

Hypnotic drugs make people sleepy, and the Wilkinson vigilance test, originally developed for people who were sleepy owing to brevity of sleep, is one that requires the subject to sustain alertness for at least 30 min and one that I regard as important in the evaluation of hypnotic drugs and their effect on day-time performance. Tasks of this duration have rarely been used in drug studies, though my colleagues and I found the vigilance test to be a task sensitive in discriminating between the intake of flurazepam 30 mg nightly during 3 weeks, compared with placebo or lormetazepam in doses of 2.5 mg or 1 mg nightly. Visual analogue scale ratings each morning of how vigilant or alert the same subjects felt themselves showed that they recognized that during their period of flurazepam intake they were aware of impairment for the first 10 days, but they thought they had returned to normal thereafter. The objective measures, however, showed a greater degree of impairment in the third week on intake of flurazepam than in the first week, confirming the value of concurrent subjective and objective evaluation and that subjective evaluation alone is not to be relied upon (Oswald *et al.*, 1979).

Hypnotic drugs are also anti-anxiety drugs, and drugs that are sold for the treatment of anxiety are drugs that can also be used as hypnotics. It is therefore of value to combine any study that involves the subjective evaluation of hypnotic efficacy with subjective ratings of anxiety, that can be made each evening using visual analogue scales. By this means drugs such as diazepam have been shown to reduce anxiety in anxious patients (Tansella, Tansella &

Lader, 1979). The withdrawal of a modern benzodiazepine, fosazepam, after 3 weeks of nightly intake has been shown to cause significant increase of daytime anxiety above withdrawal levels as a rebound sustained during a three-week period (Allen & Oswald, 1975).

Electrophysiological measures of sleep

The electroencephalogram or EEG is currently the best available indicator of the presence of sleep. As sleep succeeds wakefulness there are progressive changes in the appearance of the EEG and these appearances are divided into stages 1 (drowsiness), 2, 3 and 4 on the basis of their appearance (Rechtschaffen & Kales, 1968). This staging has a validity beyond the EEG technique. Sleep is a state of unresponsiveness and there are diminishing degrees of responsiveness associated with the EEG defined stages, for example, motor responses to auditory signals during sleep (Williams, Morlock & Morlock, 1966), amounts of spontaneous scratching by itchy patients (Savin, Paterson, Adam & Oswald, 1979), whole-body oxygen consumption (Brebba & Altschuler, 1968) and a variety of other measures, all fall with, and consistently follow, the sequential EEG stages 1, 2, 3 and 4. They rise again, to about the same level as stage 2, during paradoxical sleep, also known as rapid eye movement or REM sleep. The latter is a different state of the organism, characterised by rapid, jerky eye movements, by profound loss of skeletal muscle tone, high cerebral blood flow, penile erections and irregularities of respiration and heart rate. Recordings of eye-movements and muscle tone customarily accompany the EEG recording throughout the night, while measurements of respiration, or spontaneous electrical potentials in the skin (Haider & Oswald, 1971), or other measures may be made according to special interest.

Methods

The use of the EEG is a method suited to the detailed study of small numbers of subjects. It is precise but expensive. The techniques make heavy demands on the energies of the investigators and the co-operation of the subjects. Subjects may have to forgo home comforts, renounce alcohol and come repeatedly at night for months to an expensive, hotel-like laboratory in the study of the effect of a drug on sleep compared with that of a placebo. The patients or volunteers would be expected to attend the laboratory at perhaps 21.30 h, to have the necessary electrodes fixed to the head and to go to bed with lights out for some fixed period of time, perhaps from 22.30 h till 07.00 h. A fixed duration available for sleep is important in order to keep down the variance in a measure like total

sleep duration, one of the important measures to be made. It seems to be the practice in some laboratories to give subjects only 7 h or less in bed, but it will be found that if a longer period is given, more sleep is taken, and it is desirable not to curtail the available duration compared with what would be taken at home merely for the convenience of technicians who may be eager to get away to their families in the mornings. It may be remarked that chronic all-night sleep studies need a high degree of dedication on the part of the research workers and at Edinburgh we rely upon graduate research workers with a direct personal involvement in the successful completion of the project and its eventual publication. Where sleep studies are such that individual subjects attend during many months, it is the special relationship between subject and research workers that sustains the subject's confidence and her willingness to continue, and it is this that makes drop-outs rare.

The eventual all-night recordings, still most reliably made on paper, are again still generally scored by visual inspection. Those unfamiliar with the field often assume that this takes an undue amount of time, but the time involved is small compared with the investment of time in marshalling subjects, being there in the evening and at night with them, and generally shepherding them through a study. It may be that in the coming years the recording of the data on tape and their visual display on a screen may come to have the proven reliability that is necessary in this kind of research, but that time has not yet come. Likewise the data can be analysed by computers more expensive than those housed within the human cranium, but the computer cannot be expected to recognize unusual features that may appear as a result of the action of a drug, such as when the human eye enabled the observation to be made that fenfluramine caused nocturnal bruxism (Lewis, Oswald & Dunleavy, 1971). Monoamine-oxidase inhibitors (Akindele, Evans & Oswald, 1970) and clomipramine (Dunleavy, Brezinová, Oswald, Maclean & Tinker, 1972) are examples of drugs that cause anomalous pictures during sleep, combining some of the appearances of REM sleep with a sustained high level of muscle tone, a combination that would have defeated a naive computer. Nevertheless, there are useful functions for computer analysis that cannot be matched by the unaided human. Thus, Johnson, Hanson & Bickford (1976) used an automatic technique for the measurement of total sleep spindle activity (waves between 12.25 and 15.5 Hz) of amplitude at least 5 μ V, and they used an automatic detector for counting K-complexes. They were then able to quantify the large increase in spindle activity and the diminution in the rate of appearance of K-complexes caused by flurazepam. Subsequently Johnson, Seales, Naitoh, Church & Sinclair (1979) used off-line computer analysis to

determine the amount and distribution of delta waves (slow waves of between 0.5 and 2 Hz) with a peak-to-trough amplitude of at least 10 μ V. They were thereby able to confirm that the slow wave components of sleep are considerably reduced by the intake of flurazepam and not merely redistributed through the night as had been suggested by others. It will be realized that these examples are refinements of sleep scoring only of interest to the research worker, for the clinician will want to know the answers to the simple questions of how a drug affects the duration and brokenness of sleep. There are advanced methods being developed for automatic scoring that cover nearly all the requirements, such as that of Gaillard and Aubert (1975) and there are other computer-assisted techniques primarily for the automatic analysis of brain electrical rhythms after drug intake, (e.g. Saletu, Allen & Itil, 1974) that do not bear a close relation to the more conventional scoring of sleep states. The more a technique analyses data in a form difficult for clinicians to understand, the more it is simply a research tool, and even such standard measurements as the percentage or duration of REM sleep in the night, or the periodicity of the ultradian rhythm (the NREM-REM cycle) during the night, and the effects of drugs on these measures, cannot at the present time be interpreted in terms of immediate clinical relevance. It can be supposed only that sleep having natural characteristics is better than sleep that is in some ways distorted. The clinician would probably recognize the value of the EEG in indicating that a drug he may prescribe for other than its actions on the CNS may yet have an action on the brain, shown, by, for example, reduction of REM sleep (e.g. debrisoquine, Dunleavy, Maclean & Oswald, 1971).

Subjects

The subjects of greatest relevance will be patients, but many of these will already have been receiving CNS-acting drugs and such drugs or their withdrawal will be bound to complicate any possible effects being looked for from a new drug and, because clinical progress cannot be predicted with confidence, it is not usually possible to predict that most patients will be able to change from one drug regime to another or to placebo at fixed times in the future.

If one is testing a sleeping pill then certainly patients who complain of insomnia are ideal, but it has to be remembered that most such complaining patients sleep much longer than they think and take much less time to fall asleep than they believe (Carskadon, Dement, Mitler, Guilleminault, Zarcone & Spiegel, 1978). There is always uncertainty about how reliably such patients will give up any sleeping pills they may have been having in recent times. In the study by Frankel, Coursey, Buchbinder & Snyder (1976) a group of

patients with 'primary insomnia' was compared with a group of controls and the authors simply reported that all the insomniacs had 'agreed to discontinue' hypnotic drugs during the 4 weeks preceding the start of the study. Experienced clinicians would expect a high proportion of patients of this kind to be unreliable and likely to go on taking their drugs secretly. They will even do so when they come to the laboratory—a fact that can be betrayed by the increased fast activity at about 20 Hz in the drowsy and REM sleep EEG records, as well as some enhancement of the spindle activity. These disadvantages and uncertainties must be balanced by the thought that volunteers who say they are poor sleepers but do not use drugs will be a less-relevant population. The greatest advantage of volunteers is the flexibility, to change them to placebo or another drug precisely in accord with a planned design.

Owing to the disruption of their lives, volunteers will expect payment, breakfasts when they have attended the laboratory, taxis to and from their homes and conditions of comfort equivalent to those at home. Young men of student age generally sleep so well that it is not easy to disturb their sleep and their sleep is so prolonged and continuous there is little scope for determining whether an hypnotic drug might improve it. The older brain seems more sensitive to drugs, for example nitrazepam (Castleden, George, Mercer & Hallett, 1977) or flurazepam (Greenblatt, Allen & Shader, 1977). Middle-aged and older volunteers have the advantage of greater reliability and regularity in their habits of drinking, times of going to bed, amounts of exercise, and their relative immunity from unpredictable absences on account of transient romantic diversions. At Edinburgh we have found volunteers aged 40–70 years to be much more satisfactory than younger adults. We ask them to sign forms of consent after written and oral explanation of the study to them. We utilize their help with the agreement of their own family doctors and our hospital ethics committee.

Study design and analysis

As Kay, Blackburn, Buckingham & Karacan (1976) in their review of the human pharmacology of sleep emphasized, a great many of the published reports of the action of drugs on sleep have weaknesses in their design, often because of too few subjects, absence of cross-over procedures, absence of dose-response data or invalid statistics. It should not be assumed that this is simply because those who have undertaken the work have lacked competence or awareness of the limitations of their design, but rather because of the great labour and expense of the research.

An initial study will be exploratory and the design simple. Regard will be had to what little may be

known about the likely time for absorption of the drug, the likelihood that it may be rapidly eliminated or cumulative in its effects. There may be little data to go on in the case of, for example, newer benzodiazepines that are active in very small dosage of the primary compound, and with no available information about the metabolites owing to the difficulty of measurement. A compromise is necessary between the conflicting desires for simplicity, for dose-response data, and for the desirability of some degree of chronic intake. The traditional approach would involve the use of a Latin square with multiple doses of the drug and placebo as one of the conditions, on single nights of each condition. Such designs were popular in the days before there was recognition of the rebound effects that follow the withdrawal of an hypnotic drug, with the consequence that so-called placebo nights were often really withdrawal nights. There is no perfect answer, but if one wanted to do some preliminary testing in man with a new hypnotic then one might decide on a sample of ten individuals and ask them to attend the sleep laboratory on three nights each on two successive weeks. On the first of each trio of nights they would always get placebo and the night would be treated as an adaptation night, though with all laboratory procedures. On both of the next two nights each subject would receive either active drug or placebo, and in the subsequent week would receive the other treatment, the order being balanced among the ten subjects, perhaps with subjects attending the laboratory in pairs, each member of the pair having the opposing conditions. The all-night recordings would eventually be coded and scored 'blind'. The code would be broken and the data treated by taking the mean of each condition, and a *t*-test then used for statistical comparison. Fewer than ten people would compromise the weight that could be attached to any statistical evaluation. The data could be further analysed by comparing the first and second nights of recording in relation to one another for the drug condition, or again the differences between the drug and placebo condition on the first night could be compared with the differences between the drug and placebo condition on the second night. Such a design would be reasonably likely to give an indication of whether the dosage that had been chosen for this first exploratory trial had some measurable effects and whether they looked about the same on the two nights.

A more ambitious exploratory study would utilize two or more doses, each dose again to be taken on two successive nights and if, for example, three different dose-levels were chosen, again for comparison with placebo, then one would use a 4×4 Latin square, with attendance at the laboratory on three consecutive nights during each of 4 weeks by each volunteer and using sixteen volunteers. The general

rule is that if one asks a single, simple question one may get an answer. If one seeks to answer several questions from the one experiment, one may end up with equivocal answers to all the questions and the likelihood of getting definitive answers will depend on the use of an adequate number of subjects. The simple design just mentioned ignores the potentially confusing effect of a drug that may persist for several days, or persisting withdrawal effects.

It will be noted that one night was allowed in the above design for purposes of adaptation to the laboratory. The first night in the laboratory is accompanied by more broken sleep, less REM sleep and less total sleep. A single night is not strictly sufficient for full adaptation and, of course, after a bad first night the second night would include heavier sleep to make up for the previous night's deficiency. Individuals may be expected to vary among themselves in how rapidly they adapt and anxious patients may be expected to adapt less quickly. The inclusion in the design of at least one adaptation night prior to actual recordings should help to reduce the variance.

Larger-scale designs

A more thorough study will allow more time for adaptation, time for the establishment of what might be considered to be a baseline for an individual, the opportunity for tolerance or cumulative effects to be seen during more or less prolonged intake of the drug, and a period after withdrawal of the active preparation. The withdrawal effects are likely to be detectable for a month whatever the drug, as a consequence of the slow turnover time of brain proteins, but the abruptness and the magnitude of the peak of the rebound will vary according to how slowly the drug is eliminated.

Rebound insomnia after withdrawal of paraldehyde, chloral or barbiturates has long been recognized (Oswald, 1970), and it certainly occurs after benzodiazepine hypnotics such as nitrazepam (Adam, Adamson, Brezinová, Hunter & Oswald, 1976a), though it is hardly a new syndrome as Kales, Scharf & Kales (1978) have lately proposed. The converse of less broken sleep is found after withdrawal of a sleep-disturbing drug (Lewis *et al.*, 1971).

A rapidly-metabolized drug means an almost immediate rebound of high intensity, while slowly-eliminated active compounds, as after phenobarbitone or flurazepam, mean persistence of the effects of drug for days after withdrawal and the slow emergence of a less extreme withdrawal effect, reaching a peak as late as 2 weeks after withdrawal. If only the first few withdrawal nights are studied, a sharp withdrawal effect may be seen after a short half-life hypnotic, whereas the delayed and more gently appearing rebound after flurazepam may not be dis-

cerned. Rebound effects appear after overdoses of drugs and if one is interested in the phenomenon, patients who have taken overdoses provide illustrations of the varying durations of time to the peak of the rebound (Haider & Oswald, 1970) and as may be seen in Figure 3, which illustrates the REM sleep rebound after an overdose of flurazepam hydrochloride, the peak of the withdrawal effect being seen as late as 16 days after the single dose.

Figure 3 An illustration of the delay to the peak of a rebound where the drug is only slowly eliminated. A woman aged 58 years took an overdose of 1200 mg flurazepam and the peak of the REM sleep rebound is here delayed until 16 days later.

There is no evidence that placebo pills that are believed by the subject to contain an active drug will have any effect on the objective indices of sleep (Adam, Adamson, Brezinová & Oswald, 1976b), though one may suppose that the absence of placebo during a withdrawal period would mean that patients might be afraid of difficulty in getting to sleep and that this itself might keep them from sleeping. Consequently it is customary to use blank or placebo pills in a design that involves adaptation, baseline, active drug and withdrawal. The number of days of drug intake will vary according to the taste and resources of the experimenter. There is no merit in rigidity of design schedules. One may take as an example the following design:

- Week 1 Nightly placebos, 2 adaptation nights
- Week 2 Nightly placebos, 2 recorded baseline nights
- Week 3 Active drug, nights 1 and 3 recorded
- Week 4 Active drug, continued at home
- Week 5 Adaptation on night 16 of active drug. Nights 19 and 21 of continued active drug recorded
- Week 6 Withdrawal. Placebos nights 1, 2, 3, 5 and 7 recorded.

The above may be treated as a basic unit within a larger design so that, for example, in one of our own projects with nine volunteers of mean age 61 years, for each volunteer the drug was at one time flurazepam 30 mg nightly, at another time of year the drug was lormetazepam 1 mg nightly and at another time of year the drug was lormetazepam 2.5 mg. The design was a 3×3 Latin square, repeated three times, so that each volunteer attended the sleep laboratory on a total of 42 nights, each 6 week sequence being separated by a month of non-attendance. Naturally, each volunteer had been free of CNS drugs during the months prior to the study and continued to have no other drugs except those prescribed and no alcohol, except that modest amounts of alcohol were agreed to for the first three weeks of each rest period. It will be seen that each volunteer had to plan his life for 26 weeks and so it is necessary to take account of such times as Christmas, New Year and other holidays in planning.

The withdrawal period

A design such as this is reasonably satisfactory for statistical analysis, but there is no method that is fully informative for the evaluation of the withdrawal period. The course to recovery after withdrawal may involve an initial persistence of the drug and of, for example, continuing increased duration of sleep, then a night about equal to baseline, then several nights of reduced sleep as the full effects of withdrawal appear, and then a gradual return to baseline. If these withdrawal nights are all simply averaged, very little departure from baseline may be found. If nights in only the first week of withdrawal are recorded, then only if the hypnotic is a rapidly eliminated one can early rebound effects be expected to show up on statistical examination, by simply averaging. A graphical display of the data can convey a lot more than statistics, but statistics must be used, so that in the kind of example given above analysis of variance can be used, such as the non-parametric Friedman's analysis of variance and, where appropriate, a Wilcoxon matched pairs signed rank test to examine the difference between, for example, the two baseline nights and the two early drug nights, of any given drug regimen. Likewise there will be comparison of the baseline nights with the two late drug nights, and examination of the difference between the two early drug nights and the two late drug nights, and again the difference between the average of the baseline nights and the average of the withdrawal nights. It is then necessary to examine the data for any differences among the three drugs in their effects upon sleep. One way to do this is to average the two baseline values in each 6-week period of study for each variable and for each subject. For each subject

this average may be deemed to equal 100%, and then the values for each night in that 6 week period can be converted to percentages by comparison with the baseline 100% for that variable and for that subject. The mean percentages for each period of the 6 weeks can then be compared among drugs by ranking them 1, 2 or 3 and the Friedman 2-way ANOVA then used to compare drugs, followed by the Wilcoxon test.

Special features of sleep

In evaluating the individual features of sleep it must be borne in mind that the sleep latency, or time taken to fall asleep, is a measure that has a skewed distribution. Parametric tests, though commonly-used, are invalid, but $\log_e$ transformation will give a normal distribution of the data and allow parametric tests (Johns, 1977). The large element of variability among these latencies means that if only a small sample of subjects is used there is little likelihood of getting a significant positive effect, and even less likelihood if the drug is a slowly absorbed one. Bixler *et al.* (1978) reported on temazepam and found no significant effect on sleep latency. This would not justify any conclusion that the drug is without effect on the time taken to fall asleep, it simply reflects the fact that only six subjects were employed.

The criterion of sleep latency that is usually taken is the duration of time from lights-out to the first stage 2 sleep, namely a sleep spindle at about 14Hz, usually most prominent over the front of the head, and of at least half a second in duration. The latency to the first stage 1 sleep is not satisfactory, since stage 1 is less clearly defined and comes and goes. Sometimes authors have used the latency to the beginning of that period of stage 1 that leads into sustained stage 2 sleep. In a recent study we used all three measures independently and failed to find any evidence to support the idea that 1 g of L-tryptophan would shorten sleep latency (Adam & Oswald, 1979).

The amount of accumulated sleep between first falling asleep and the onset of the first REM period of the night, or REM latency, is likewise not a measure that is normally distributed. It is not only skewed, it has a bimodal distribution, and once again parametric tests are not appropriate.

Distribution in the night

It is desirable to look not only at the total amount of any sleep stage in the night, but also at its distribution. It is common in hypnotic drug studies to find no change in the total amount of REM sleep in the night, but if then one looks carefully it can be seen that REM sleep is suppressed in the early part of the night and enhanced above normal in the later part of the night.

Hour-by-hour cumulative curves offer a useful means of displaying data, as Figure 4 exemplifies. The figure shows the number of minutes of wakefulness and of stage 1 that were present in the course of accumulating a total of 60 min of actual sleep, likewise the number of minutes that intruded in the course of accumulating a total of a further 60 min of sleep for the second hour, and so on.

Figure 4 A method for displaying cumulative data, in this case the cumulative amounts of intervening wakefulness plus drowsiness. Figure by courtesy of Dr Vlasta Březinová, and data from Březinová (1974). Caffeine at bedtime makes sleep more broken. ● caffeine 300 mg; ▲ decaffeinated coffee; ○ no drink.

Other measures

The electrophysiological measures outlined above can be conducted concurrently with studies of body motility through the night, so that an hypnotic drug like flurazepam will be found to reduce motility, and to do so not only by night but by day also (Crowley & Hyding-Macdonald, 1979). A novel approach has recently been adopted by Bonnet & Webb (1979), who deliberately awakened their subject at intervals through the night and measured how long it took to fall asleep again and found that prior flurazepam or pentobarbitone were associated with more rapid return to sleep.

Restorative function and hormones

In the U.S.A. many workers have adopted the term 'sleep efficiency' as an indication of the proportion of time in bed and available for sleep that has actually

been spent asleep, but the term raises the further question of the efficacy or restorative function of sleep and whether drugs affect that restorative function (Adam, 1979). So far it has been hormonal indices that have especially been studied, to give some indication of the probable balance between anabolic and degradative processes in association with sleep and wakefulness. The hormone pattern of sleep is generally anabolic, while the hormone pattern of wakefulness favours catabolism. The catabolic corticosteroids are low during the early night and only rise towards the end of sleep, while growth hormone is high in the early part of the night, in association with EEG slow-wave sleep. In contrast, during wakefulness growth hormone is low, but the catecholamines and corticosteroids are relatively high.

A catheter can be inserted in a forearm vein at bedtime and an extension led out through a hole in the bedroom wall, so enabling blood to be sucked at intervals through the night without seriously disturbing sleep. The system can be kept filled with heparinized saline during the intervals between blood-sucking. It is advisable to sample every twenty minutes or so, spinning down samples of blood and putting the plasma as quickly as possible into a deep freeze. It is taxing work and it requires at least the nearby presence of medical personnel. Spasm in the vein is an occasional problem and sometimes the nights have to be repeated. Consequently, this kind of research is expensive and has not often been undertaken using substantial numbers of subjects. However, it has been shown that a modern minor tranquil-

izer, benzocetamine, will bring about a significant fall in the concentration of corticosteroids in the plasma during sleep, with a rebound above normal following withdrawal (Ogunremi, Adamson, Brezinová, Hunter, Maclean, Oswald & Percy-Robb, 1973). Although the benzodiazepines in chronic use greatly reduce the proportion of slow-wave sleep, the secretion of growth hormone does not appear to be affected by nitrazepam (Adam *et al.*, 1976a) or by benzocetamine or barbiturates (Ogunremi *et al.*, 1973).

A less onerous avenue to knowledge of the hormone pattern during sleep and its modification, if any, by drugs, is possible with newer techniques for the measurement of urinary adrenaline and cortisol, but these will give only the catabolic side of the picture and research into the effects of drugs on the anabolic hormones will continue to require all-night blood sampling.

Conclusions

Studies of the effects of drugs on sleep are very much dictated in their design by matters of cost, for they require a lot of organization and dedication over long periods of time, both by subjects and staff. It is thus axiomatic that studies should be well-designed from the start and analysed using appropriate statistics. The exceptionally heavy use of hypnotic drugs throughout the world justifies a full assessment of a new hypnotic drug before it is introduced to general consumption.

References

- ADAM, K. (1979). Do drugs alter the restorative value of sleep? In *Pharmacology of the States of Alertness*, ed. Passouant, P. & Oswald, I., pp. 105–111. Oxford: Pergamon Press.
- ADAM, K., ADAMSON, L., BREZINOVÁ, V., HUNTER, W.M. & OSWALD, I. (1976a). Nitrazepam: lastingly effective but trouble on withdrawal. *Br. med. J.*, **1**, 1558–1560.
- ADAM, K., ADAMSON, L., BREZINOVÁ, V. & OSWALD, I. (1976b). Do placebos alter sleep? *Br. med. J.*, **1**, 195–196.
- ADAM, K. & OSWALD, I. (1979). One gram of L-tryptophan fails to alter the time taken to fall asleep. *Neuropharmac.*, **18**, 1025–1027.
- AKINDELE, M.O., EVANS, J.I. & OSWALD, I. (1970). Monoamine oxidase inhibitors, sleep and mood. *Electroenceph. clin. Neurophysiol.*, **29**, 47–56.
- ALLEN, S. & OSWALD, I. (1976). Anxiety and sleep after fosazepam. *Br. J. clin. Pharmac.*, **3**, 165–168.
- BIXLER, E.O., KALES, A., SOLDATOS, C.R., SCHARF, M.B. & KALES, J.D. (1978). Effectiveness of temazepam with short-, intermediate-, and long-term use: sleep laboratory evaluation. *J. clin. Pharmac.*, **18**, 110–118.
- BOND, A. & LADER, M. (1974). The use of analogue scales in rating subjective feelings. *Br. J. med. Psychol.*, **47**, 211–218.
- BONNET, M.H. & WEBB, W.B. (1979). The return to sleep. *Biol. Psychol.*, **8**, 225–233.
- BREBBIA, D.R. & ALTSHULER, K.Z. (1968). Stage related patterns and nightly trends of energy exchange during sleep. In *Computers and Electronic Devices in Psychiatry*, ed. Kline, N.S. & Laska, E., pp. 319–335. New York: Grune & Stratton.
- BREZINOVÁ, V. (1974). Effect of caffeine on sleep: EEG study in late middle aged people. *Br. J. clin. Pharmac.*, **1**, 203–208.
- CARSKADON, M.A., DEMENT, W.C., MITLER, M.M., GUILLEMINAULT, C., ZARCON, V.P. & SPIEGEL, R. (1976). Self-reports versus sleep laboratory findings in 122 drug-free subjects with complaints of chronic insomnia. *Am. J. Psychiat.*, **133**, 1382–1388.
- CASTLEDEN, C.M., GEORGE, C.F., MARCER, D. & HALLETT, C. (1977). Increased sensitivity to nitrazepam in old age. *Br. med. J.*, **1**, 10–12.
- CROWLEY, T.J. & HYDINGER-MACDONALD, M. (1979). Bedtime flurazepam and the human circadian rhythm of

- spontaneous motility. *Psychopharmacology*, **62**, 157–161.
- DUNLEAVY, D.L.F., BREZINOVÁ, V., OSWALD, I., MACLEAN, A.W. & TINKER, M. (1972). Changes during weeks in effects of tricyclic drugs on the human sleeping brain. *Br. J. Psychiat.*, **120**, 663–672.
- DUNLEAVY, D.L.F., MACLEAN, A.W. & OSWALD, I. (1971). Debrisoquine, guanethidine, propranolol and human sleep. *Psychopharmacologia (Berl.)* **21**, 101–110.
- FRANKEL, B.L., COURSEY, R.D., BUCHBINDER, R. & SNYDER, F. (1976). Recorded and reported sleep in chronic primary insomnia. *Arch. gen. Psychiat.*, **33**, 615–623.
- GAILLARD, J.M. & AUBERT, C. (1975). Specificity of benzodiazepines action on human sleep confirmed. Another contribution of automatic analysis of polygraphic recordings. *Biol. Psychiat.*, **10**, 185–197.
- GREENBLATT, D.J., ALLEN, M.D. & SHADER, R.I. (1977). Toxicity of high-dose flurazepam in the elderly. *Clin. Pharmac. Ther.*, **21**, 355–361.
- HAIDER, I. & OSWALD, I. (1970). Late brain recovery processes after drug overdose. *Br. med. J.*, **2**, 318–322.
- HAIDER, I. & OSWALD, I. (1971). Effects of amylobarbitone and nitrazepam on the electrodermogram and other features of sleep. *Br. J. Psychiat.*, **118**, 519–522.
- HINDMARCH, I. (1980). Psychomotor function and psychoactive drugs. *Br. J. clin. Pharmac.*, **10**, 189–210.
- HINDMARCH, I., PARROTT, A.C. & ARENILLAS, L. (1977). A repeated dose comparison of dichloralphenazone, flunitrazepam and amylobarbitone sodium on some aspects of sleep and early morning behaviour in normal subjects. *Br. J. clin. Pharmac.*, **4**, 229–233.
- JOHNS, M.W. (1977). Validity of subjective reports of sleep latency in normal subjects. *Ergonomics*, **20**, 683–690.
- JOHNSON, L.C., HANSON, K. & BICKFORD, R.G. (1976). Effect of flurazepam on sleep spindles and K-complexes. *Electroenceph. clin. Neurophysiol.*, **40**, 67–77.
- JOHNSON, L.C., SEALES, D.M., NAITOH, P., CHURCH, M.W. & SINCLAIR, M. (1979). The effects of flurazepam hydrochloride on brain electrical activity during sleep. *Electroenceph. clin. Neurophysiol.*, **47**, 309–321.
- KALES, A., SCHARF, M.B. & KALES, J.D. (1978). Rebound insomnia: a new clinical syndrome. *Science*, **201**, 1039–1041.
- KAY, D.C., BLACKBURN, A.B., BUCKINGHAM, J.A. & KARACAN, I. (1976). Human pharmacology of sleep. In *Pharmacology of Sleep*, ed. Williams, R.L. & Karacan, I., pp. 83–210. New York: John Wiley.
- LEWIS, S.A., OSWALD, I. & DUNLEAVY, D.L.F. (1971). Chronic fenfluramine administration: some cerebral effects. *Br. med. J.*, **3**, 67–70.
- OGUNREMI, O.O., ADAMSON, L., BREZINOVÁ, V., HUNTER, W.M., MACLEAN, A.W., OSWALD, I. & PERCY-ROBB, I.W. (1973). Two anti-anxiety drugs: a psychoneuroendocrine study. *Br. med. J.*, **2**, 202–205.
- OSWALD, I. (1970). Dependence upon hypnotic and sedative drugs. *Br. J. hosp. Med.*, **4**, 168–172.
- OSWALD, I., ADAM, K., BORROW, S. & IDZIKOWSKI, C. (1979). The effects of two hypnotics on sleep, subjective feelings and skilled performance. In *Pharmacology of the States of Alertness*, ed. Passouant, P. & Oswald, I., pp. 51–63. Oxford, Pergamon Press.
- OSWALD, I., LEWIS, S.A., DUNLEAVY, D.L.F., BREZINOVÁ, V. & BRIGGS, M. (1971). Drugs of dependence though not of abuse: fenfluramine and imipramine. *Br. med. J.*, **3**, 70–73.
- RECHTSCHAFFEN, A. & KALES, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington, D.C.: U.S. Government Printing Office, NIH Publication, No. 204.
- SALETU, B., ALLEN, M. & ITIL, T.M. (1974). The effect of coca-cola, caffeine, antidepressants, and chlorpromazine on objective and subjective sleep parameters. *Pharmakopsychiat.*, **254**, 307–321.
- SAVIN, J. A., PATERSON, W.D., ADAM, K. & OSWALD, I. (1979). Effects of trimeprazine and trimipramine on nocturnal scratching in patients with atopic eczema. *Arch. Dermatol.*, **115**, 313–315.
- SPIEGEL, R. (1973). Controlled clinical investigations with a phenothiazine derivative used as sleep-inducer in non-psychotic insomniacs. In *Sleep: Physiology, Biochemistry, Psychology, Pharmacology, Clinical Implications*, ed. Koella, W.P. & Levin, P., pp. 91–101. Basel: Karger.
- TANSELLA, C.Z., TANSELLA, M. & LADER, M. (1979). A comparison of the clinical and psychological effects of diazepam and amylobarbitone in anxious patients. *Br. J. clin. Pharmac.*, **7**, 605–611.
- WILLIAMS, H.L., MORLOCK, H.C., & MORLOCK, J.V. (1966). Discriminative responses to auditory signals during sleep. *Psychophysiol.*, **2**, 208–215.
- ZEALLEY, A.K. & AITKEN, R.C.B. (1969). Measurement of mood. *Proc. roy. Soc. Med.*, **62**, 993–996.