J. chron. Dis. 1965, Vol. 18, pp. 77–98. Pergamon Press Ltd. Printed in Great Britain

THE DESIGN AND LOGIC OF A MONITOR OF DRUG USE

D. J. FINNEY

Department of Statistics and Research Group in Biometric Medicine, University of Aberdeen, Scotland

(*Received* 1 *June* 1964)

BY THE *monitoring* of drugs is meant any systematic collection and analysis of information pertaining to adverse effects or other idiosyncratic phenomena associated with the normal use of drugs. It is to be distinguished from the planned clinical trials that should precede the release of a new drug for general use. Such trials are concerned with the efficacy of the drug, relative to existing alternatives (other drugs, placebos, surgical treatment, etc.), and of course with the detection of any high frequency of particular adverse reactions. They involve at most a few thousand subjects (usually far fewer), and therefore any reaction limited to 0.1 per cent of the population or less is unlikely to be identified as associated with the drug. Even this incidence may be sufficiently high to make use of the drug undesirable, at any rate in some sectors of the population, though decision must depend upon the nature of the disease for which the drug is a remedy and upon the availability of alternative drugs.

Monitoring is an important complement of, not an alternative to, formal clinical testing. The information it yields will seldom be so striking as to lead to an immediate change of official policy or of therapeutic practice. The purpose is rather to ensure that observations on a large number of persons who receive a new drug are collated and used effectively; only so can a warning of any untoward consequences be given as early as possible. Research effort must then be directed at the circumstances.

A number of people have recently been giving attention to medical aspects of monitoring, and to the administrative feasibility of alternative plans. This paper is primarily concerned with defining the task, with logical classification of the types of data that might accrue, with appropriate statistical examinations, with the broad outline of computer programming, and with the proper use of monitoring results.

1. DEFINITIONS

The detail of a monitoring program must depend greatly on the drugs and diseases involved, and on the patients recorded; the principles remain constant. Establishment of a standard terminology, independent of the technical nomenclature of particular medical situations, may enable general method to be discussed without the confusion inevitable when each party to discussion refers to a different special problem. The definitions that follow are presented also to meet the needs of statisticians and computer programmers.

A *Patient* is any human being during a period in which he undergoes medical care or management, and, where appropriate, during any subsequent period of related measurement and observation.

A *Patient Set* is any category of patients to which attention is to be restricted (e.g. all adult diabetics, all women treated during the first three months of pregnancy, all patients treated by any physician who is on a certain list).

A *Drug* "is any substance destined for administration to man for use in the diagnosis treatment, investigation or prevention of disease or for the modification of physiological function" [1]. A mixture of two or more chemical compounds in fixed proportions, whatever their separate actions, may be regarded *either* as one drug *or* as several (of which each may be discussed separately), in accordance with the judgment of the investigator as to the more informative approach.

A *Monitor* is any process of systematic collection of information on phenomena associated with the normal use of drugs and of regular analysis and interpretation of this information, undertaken with a view to obtaining evidence about adverse reactions.

A *Drug Set* is any list of drugs that are to be simultaneously monitored. Usually these will be drugs administered for the same or similar purposes, so that comparisons between them are of interest. They may be related chemical compounds; but this is not essential, and often one or two long-established unrelated drugs will be included as standards of reference. If the monitor is to include such diversity of drugs that direct comparisons between all pairs are not of interest, discussion in terms of two or more drug sets (usually non-overlapping, but not necessarily so) will be convenient. For example, one method of monitoring would be to place all newly introduced drugs on a special register for one or two years. In the analysis of records (Section 9), no interest would lie in comparing, say, a tranquillizer with an antibiotic, and the totality of drugs might usefully be divided into several sets.

An *Event* is a particular untoward happening experienced by a patient, undesirable either generally or in the context of his disease. For example, occurrence of a rash, of peripheral neuritis, of aplastic anemia, of post-operative cardiac failure, or even involvement in a road accident could be classed as an event. The term is *not* to be limited either to recognized side-effects of a drug or to incidents that are in some sense unexpected. Every patient in the monitor field (see below) to whom an event occurs will be noted, irrespective of whether that event is thought to be wholly or partly caused by the drug.

An *Event-Type* is the class of events to which the same name is given. Thus aplastic anemia may be termed an event-type, each case of the disease being a single event.

An *Event Set* is a list of event-types regarded as deserving systematic attention. Considerations of practicability limit the size of the set, and are likely to compel concentration on events so serious as to constitute danger to life or major impairment of health. Nevertheless, events outside the set must not be neglected completely, even if

fairly trivial in character; every encouragement should be given to reporting more fully than the minimum required, and the program should be sufficiently flexible to permit modification of the event set if accumulating evidence suggests that other event-types ought to be added. When ascertainment is by patient or by drug (Section 4), *'no event'* (i.e. absence of every other event-type in the set) should be treated as an event-type belonging to the set; this enables rates of incidence for each event-type to be calculated from the number of cases 'exposed to risk.'

A *Record* comprises a complex of information concerning one patient, the drugs he receives, and the events he experiences, with as much ancillary information as may be available (Section 2).

Ascertainment is the method by which a patient is encountered and his record obtained.

A *Monitor Field* is the conjunction of patient set, drug set, and event set over which a monitor is to operate. Information from outside the field may accumulate, and may from time to time be used in a redefinition of the boundaries of the field (in respect of patients, drugs, or event-types), but strict inference is likely to be restricted to the field.

A *Monitoring Center* is a national or international center to which records from one or more monitor fields are sent. It will need medical, pharmacological and statistical advice of a high standard and good computer facilities. Probably one such center will suffice for a country, but provision must be made for feeding national summaries to an international center (Section 11).

A *Reporter* is a physician, hospital, or other individual or institution, responsible for the preparation of records and their regular transmission to a monitoring center.

A *Reporter Set* consists of all reporters who have agreed to co-operate in providing records from a particular monitor field.

2. THE RECORD

Whatever the method of ascertainment (Section 4), the record for one patient can consist of up to seven distinct parts. These are:

R1: Identification.
R2: Attributes of patient.
R3: Past medical history.
R4: Diagnosis leading to recent drug prescription.
R5: Drugs.
R6: Event.
R7: Reporter's comments.

Of these, R6 is the essential, but the quality of the monitoring and the nature of inferences that may be drawn will depend greatly upon the availability of other information with which R6 may be correlated. Explanation of the seven parts follows:

R1 should be a reference number that permits any record to be traced back to the name and address of the patient. This facility will be needed in checking questionable records, in attempts to amplify incomplete records, and occasionally in follow-up studies, but of course it must be treated in strict confidence. One component of the number should indicate geographical area, which may be useful in sorting, and another the date of the record. A reference that expedites checking for duplicate recording of the same case will be very desirable.

R2 will include sex, age, and any other normal attributes that may be reported. However great the effort that is made, the record is sure to be very incomplete under this head, and unfortunately the attributes reported will not be the same for all patients. Possibilities include physical characteristics (height, weight), genetic characteristics (blood groups), parity, social and occupational classes, and so on. For any one monitor field, a list of the attributes considered worth recording will be specified, in order that space may be left for each, even though the entry may often have to be 'not known'.

R3 is analogous to R2, but relates to abnormal history of the subject. From the vast number of possibilities, a list of those most relevant to a particular monitor field will have to be prepared so that space may be provided. Such diverse factors as previous virus diseases, inoculations, known allergies, surgical and accident history need consideration; obviously only a few can be retained, and again many entries will be 'not known.'

R4 will provide a classification appropriate to the drug set, possibly much simplified but certainly compatible with accepted international practice and terminology. Information on any secondary conditions simultaneously diagnosed in the patient should be included.

R5 will have as its main entry the drug that brings the record within the monitor field; information on dose, frequency, and date and duration of dosing should be added. Also under R5 will come information on any other drugs recently administered to the same patient. A patient may be recorded separately for two or more drug sets.

R6 will usually contain a single event experienced by the patient, or a statement of 'no event.' Of course, a patient may be reported as having more than one event, and R6 must provide for such reports. The date of the event, or of its first recognition, will be needed.

R7 is intended to permit and encourage the reporter to remark on any features of the case that seem to him interesting, but to distinguish his opinions and impressions—valuable though these may be—from the factual information under R1 to R6. Indication of whether he or any other medical adviser considers the event likely to have been drug induced may be valuable, and should be included here.

Any monitor field will require that a standard form of individual record be devised. Space must be allotted for all items of R1 to R6 judged appropriate to the field, even though information on some will often be lacking. All this information must be coded numerically, and a major preparatory task will be to prepare suitable codes in terms of which the record can be put on one or more punched cards. Almost every item will need to allow 'not known' as an alternative to specific information. Several different systems for encoding chemical compounds are at present under development; an ideal system for R5 will be one conducive to easy grouping of related drugs that are believed

likely to be similar in action. The coding of R6 will also need careful planning. The nature of R7 is such as to make it unsuitable for coding in any detail; probably a code indication of whether *anything* is recorded there will be a useful reminder that reference to the original record can be worth while. In some circumstances, different parts of a record may be initiated by different people. Provided that all carry the same identification in R1, the parts can be linked at the monitoring center so as to produce a complete record for storage.

Under ideal conditions, indefinitely long time-intervals between administration of a drug and an event in the same patient would be permissible. In practice, long intervals will increase risks of confusion on account of the number of drugs received by a patient, and of course will decrease the hopes of complete and reliable recording. Where teratogenicity is to be considered, observations must cover an interval of about a year, or even longer if the event-type is not observable very soon after birth; moreover, an event observed for a fetus or for an infant must be regarded as an extension of the experience of the mother and associated with *her* R1–R7. The desirability of good long-term follow-up, however, must not be allowed to delay introduction of monitors that may be adequate for shorter intervals. To run a monitor that might detect drug carcinogenesis after several years will require very careful organization. Important though this quality of monitoring may be, the difficulty of securing it must not be an argument against early operation of a monitor that could detect drug induced hepatitis or cerebral hemorrhage. Monitoring should first be tried under conditions that give it the best chance of success.

The coding and punching of cards may be done either at the monitoring center or before records reach this stage; at the center, however, all records must be transferred to magnetic tape or to some other storage medium convenient for rapid access and high speed analysis. The whole pattern of coding and storage must be so flexible that not only can new records be added at frequent intervals but also new items can be incorporated. Entries in R7, interim analyses of the monitor field, and information from other research may occasionally suggest improvements in reporting; change must not be so frequent as to permit discontinuities to reduce the value of the monitor, but rigidity of provision for record storage ought not to be an important consideration.

3. THE EVENT SET

One criticism raised against monitoring proposals is that the decision whether or not to regard certain occurrences as events is difficult and different reporters will not agree in their practices. This may be true, but it is equally true that for many event-types the decision is neither difficult nor subjective. The aim of monitoring is to study 'serious' adverse reactions to drugs: although opinions will differ about the lower limits of the class to be regarded as serious, the existence of a problem of definition is no excuse for refusal to establish a practice of reporting for unambiguous cases. No one will question the propriety of regarding cataract, hepatitis, phocomelia, or leukemia as serious events, and early experience of large-scale monitoring should be gained on events such as these. It cannot be too strongly emphasized that a reporter is not required to judge whether an event was drug-induced, though he may usefully express an opinion in R7. *He must be prepared to report all cases within the monitor field,* irrespective of this judgment, but his difficulties of classification of the less serious event-types should not be allowed to impair his recognition of the more serious.

Probably each reporter will need to be given a check-list of the event set, perhaps supplemented by a list of event-types not within the set that may later become of interest. This has been done for the drug monitor introduced by the American Medical Association. A check-list will reduce the risk of an event being overlooked, indicate broadly the gravity of events for which search is being made, and aid provision of adequate space for records. There are two dangers. Some reporters who know that one or two event-types are expected to occur will 'find' too many cases; however, if the event set is kept fairly diverse, the bias from 'finding what one is looking for' should almost disappear. Secondly, if too much emphasis is placed on a listed event set, reporters may consciously or unconsciously fail to record events outside the set. Monitoring must be designed to operate effectively when the nature of any adverse reactions to a drug is wholly unknown, and narrow limitation on the event-types to be reported would defeat its objects. On this count, also, it appears that the event set should normally be very broad, including event-types that seem *a priori* unlikely, and that reporters should be urged to record even more widely than this, regarding the check-list as merely a guide. The alternative of providing no check-list would reduce some risks of bias, but would almost certainly lead to under-reporting.

If the net is spread widely, many events recorded will have no dependence on a drug, being either symptoms of the disease or rare but spontaneous reactions of the individual to his environment. Others may be consequences of overdose, perhaps in patients who have low safety thresholds, rather than idiosyncratic reactions, and as such may be well-known. If a particular event-type may occur for more than one reason, some confusion and following of false trails is inevitable. Of course, if an event-type could occur only as an overdose phenomenon, it would be excluded from the event set, but the certainty that justifies this is likely to be rare.

4. ASCERTAINMENT

Three methods of ascertainment of records must be distinguished. They can lead to records of the same form, but the problems of organization for securing them may be entirely different. They will be identified as P(atient), D(rug), and E(vent):

P: Within a population (e.g. a geographic region, or the cases entering a group of hospitals), all members of the patient set are noted and recorded. For these patients, information pertaining to the drug set and the event set is recorded, with emphasis on completeness of recording for major items.

D: Records of prescriptions, (e.g. for the whole population of a region, for the dispensaries of selected hospitals, or from selected general practitioners) are either sampled or fully scanned for instances of drugs within the drug set. The patients receiving these prescriptions are located, and recorded if they are within the patient set; information pertaining to the event set is added to the records.

E: Within a population, all instances of events (other than 'no event,' of course) within the event set are noted. The patients experiencing these events are located and recorded as before. This ascertainment will scarcely be satisfactory unless the population is very readily observable (e.g. the patients within selected hospitals) or the event-types are very clearly recognizable.

The development of computers with large memories and quick access to stored information is making possible systems of medical record linkage, whereby complete

medical records for a segment of a population can be maintained centrally. Growing interest in these will evidently provide opportunities for incorporating good information on drugs, and so operating monitors with P ascertainment. Within a hospital, D ascertainment should not be difficult. KOLLER [2] has made an interesting proposal for its wider operation. He proposes that, by agreement, manufacturers should identify every kth packet of each drug in a drug set (where k might be 50 or 3—or even 1, so that every packet is identified). Any doctor with a patient to whom one of these packets is dispensed would be expected to make special observations on the course of that patient's illness, and to send a record to the monitoring center. The practicability of this method may depend upon national practice in dispensing. The British National Health Service system of handling drug prescriptions could possibly help in producing similar information. Examples of E ascertainment are the well-known American Medical Association's Registry of Blood Dyscrasias and, apparently in a less formal manner, the short-lived Mediphone project [3].

The drug set is likely to be restricted to a few members, perhaps at most those introduced within a specified recent period with a few long-established one as standards of reference. If the event-types are fairly rare, D ascertainment will give far fewer records than P, and E again fewer than D. Probably D is the most reasonable for wide adoption, since it could be operated without great labor, at least in hospitals. Of the vast amount of recording needed for P, much will be seen to be irrelevant as soon as attention is focused on a particular drug set, but this ascertainment needs discussion as the most comprehensive approach. As will be seen from Section 9, E has grave disadvantages, but for certain patient sets or event sets it may be all that is practicable at a particular time. Whatever the method of ascertainment, practical considerations may dictate adoption of a time-limit on the interval between drug and event, at least for the main analyses.

5. CLASS OF RECORDS

The information content of a record will almost inevitably vary substantially from one patient to another, because of the difficulty of obtaining reports on all characters judged interesting. The nature of the statistical analysis that can be undertaken, however, will be largely governed by the class into which the records fall in respect of content. Three main classes may be distinguished:

III: Only R1 and R6 are reported, except perhaps for sex and age in R2. This situation is particularly likely with ascertainment E, but could occur with P; with ascertainment D, presumably at least the name of the drug would also be recorded. For example, mortality or morbidity records may be scanned for instances of an event-type; hospitals and physicians might be asked to report all cases encountered, and there might be special reasons for reliable information on antecedent circumstances being unobtainable. Evidently pure III reporting cannot give direct information on drugs.

II: R5, and probably R4 are also reported. This may occur under any method of ascertainment, but is perhaps particularly likely with D: instruction may be issued that all users of certain drugs are to be recorded, whereupon it will be natural also to record diagnostic details, and that any subsequent untoward events are to be reported.

I: Ideally, of course, information under R2 and R3 will also be recorded. Although with P or D ascertainment some such information ought usually to be available, there will be situations in which practically no general information on the patient is available except for age and sex.

A monitor may have to be begun with records of class III, and perhaps with E ascertainment, but one would hope that this will evolve into II and I without too long a delay. With this evolution may come a change to D ascertainment, which for most drug sets will have many advantages; for some patient sets, further evolution to P may be possible.

6. SCALE OF OPERATIONS

The arguments for conducting a planned clinical trial, comparing a new drug with a placebo or a standard drug before deciding whether or not to adopt the new new drug for general use, are now widely understood and will not be discussed here. Although such a trial is primarily concerned with the efficacy of a drug, obviously any indications of adverse reactions will be recorded and, if sufficiently serious, will condemn the drug or at best suggest that its use be restricted. Why then is there need for monitoring of a drug in general use?

In part this is because a clinical trial is conducted under rigid conditions, with fixed dosage and strictly defined categories of patient. When a drug is released for general use, inevitably conditions are somewhat relaxed: variations in dose will occur; the drug will be given to patients who have recently had other drugs; the classes of patients and of diagnostic circumstances for which the drug is used may be widened, in ways that seem in no way to conflict with what the clinical trial justified but that have unexpected important consequences. The ages of patients, their occupations, their diets, and the severities of their diseases are all likely to be more varied than they were in the clinical trial. The recently discovered association between cheese eating and adverse reactions to tranylcypromine is a striking instance. Only a good monitor can provide any systematic check on the possibility that changed conditions of drug use are affecting its efficacy or its toxicity.

A second major consideration is the number of cases required for detecting adverse reactions. A controlled clinical trial might have 200 subjects on each drug; it then has about an even chance of demonstrating clearly that those receiving one drug manifest an event-type more frequently than those on another if the true incidences of these events are 5 per cent and 11 per cent, and a poorer chance if the rates are closer together. Even with 1000 subjects on each drug, rates as different as 1.5 per cent and 3.0 per cent (or as 0.1 per cent and 0.8 per cent) have only an even chance of detection. Far greater numbers are needed if increases in frequencies of events lower than these but still large enough to be of great importance to the community are to be detected. For example, with 15,000 subjects on each of two drugs there will be an even chance of detecting a difference between rates of incidence of 0.1 per cent and 0.2 per cent, or of 0·02 per cent and 0·08 per cent. These figures assume the most favorable circumstances: since records on this scale cannot be as well-kept or as balanced as in a small planned trial, the discriminating power in practice will be appreciably lower, but the examples give some idea of what may be achieved in a good monitoring program.

The rate of incidence of adverse reactions that can be tolerated in a drug evidently depends upon the disease against which it is to be used, upon the prognosis for in-

dividual cases with this drug or with alternatives, and upon whether or not a reaction is reversible. A rate thought reasonable for a drug that halts the progress of leukemia is likely to be quite improper for an anesthetic. Hence attempts to set a standard for the minimal relative frequency of an event-type, or an event set, that is to be judged important are useless unless made with reference to a particular drug and a particular disease. Nevertheless, both clear objective thinking and reliable records are needed as correctives to the very different opinions that can be expressed on one set of circumstances. An extreme instance is provided by *triparanol*, which was introduced as an inhibitor of cholesterol biosynthesis, widely used in the U.S.A., and withdrawn from use in April 1962 because of suspected harmful side-effects. In evidence submitted to a U.S. Senate Sub-Committee [4], the Deputy Commissioner of the Food and Drug Administration implied that 4 cases of cataracts in 300,000 patients who had received the drug "raised substantial doubt as to the safety of the drug"; this was in association with results from tests on animals, and did not claim to refer to complete ascertainment of all cataracts in these patients. Yet, one month later, in discussions concerned with the question of removing the drug from the market, the manufacturers were content to claim that "the side-effects of all types reported to us to date total substantially less than 1 per cent of the patients treated" [5]. Even when allowance is made for the inclusion of many different event-types, the contrast between rates of about 0.00001 and perhaps 0.005 is striking!

7. QUALITY OF RECORDING

As stated in Section 2, records will need to be stored on magnetic tape or some similar rapid-access filing system. From there, they will be taken for analytical scrutiny and tabulation along lines described in Section 9. *The quality of results coming out of a computer cannot be higher than the quality of the records put in.*

To maintain high standards of accuracy and completeness must therefore be regarded as the first essential. To begin with modest requirements in respect of class (Section 5) is better than to fail to fulfil something more ambitious. To have trustworthy information recorded and submitted to fairly primitive statistical procedures is better than to accept slipshod records uncritically for analysis by sophisticated techniques, yet the aim must be to have both records and statistical analysis of high quality. Limited but conscientiously compiled records will always be preferable to an indiscriminate mixture of sound information and 'garbage.'

Any attempt to begin monitoring with a great number of reporters, indiscriminately chosen and of widely ranging abilities and enthusiasms for the task, is doomed to failure. Even for those who are keenly interested, maintenance of good recording practice will require constant care and attention to detail; any who are pressed to help against their inclination may begin satisfactorily, but the reports they send will probably becomes less and less reliable. Many physicians agree that the standard of patients' records is deplorably low, even in good hospitals. Documentation and ready availability of a patient's medical history are becoming increasingly important to the quality of medical care. Surely standardization of record systems, which is not inconsistent with allowing any clinician space for personal comments and opinions, and inculcation in medical students of a belief in personal responsibility for recording, are overdue reforms.

For a particular monitor field, the wisest immediate course seems to be in the first instance to recruit reporters who volunteer to help on account of genuine interest in the project. One would then hope, by demonstration of the benefits of monitoring and by education in the requirements for recording, to increase steadily the number of participants. Although ultimate responsibility for recording must rest on the clinician, almost certainly the clerical staffs of hospitals need reinforcing with persons competent in the day-to-day handling of records, for whom the chasing of missing items has a high priority. This is part of a larger problem of the management of medical records.

This is not the place for discussing the psychology of good relations, great though its importance is. Evidently all concerned with reporting should be encouraged to think about the purpose of the project, to suggest improvements from time to time, and to comment on any trends or differences suggested by their own scrutiny of records. The value of this last is necessarily limited by the small number of cases seen by one physician or even in one hospital; the compensating benefit from personal knowledge of patients must not be ignored, as a source of impressions that cannot be put into a formal record. Possibly volunteer regional groups of physicians may be encouraged to discuss their own records and to assist in feed-back of ideas and tentative suggestions, but local machinery must not impede the rapid flow of records to the monitoring center. Whether a small payment should be made for each satisfactory record needs consideration, though probably any payment should be only an approximate refund of clerical and postal expenses.

8. VALIDATION OF RECORDS

Before a record is transferred from its written or punched card form to its permanent store, it must pass through a validation procedure. In large part, this should consist of a computer scrutiny directed at discovering errors of recording. For example, the computer can readily signal that a record purports to refer to a patient over the age of 100, a pregnant male, a child under 10 in employment, or other indication that a mistake probably or certainly has occurred. Certain events may be referred back for reconsideration because they are unlikely in conjunction with specific items of R1 to R4 or because of inherent improbablity.

Of course, improbability is not in itself a cause for rejection of a record, especially since monitoring aims at detecting apparently improbable associations. It may be good reason for inspection of the whole record by someone with suitable medical experience, and possibly for asking the reporter to check the record. One danger inherent in any monitoring is that keen reporters will too often 'find' what they are looking for. Neither human nor computer scrutiny will detect all errors, but combination of the two may be made a good safeguard. By steady improvement, the computer validation program should be made to assume maximum responsibility for detecting the inconsistencies and mistakes most likely to be encountered, leaving to the human agency further cross-checks and general search for unpredictable errors.

Either at the beginning or at the end of validation, each record should be checked for duplicate appearance in the monitor.

9. STATISTICAL ANALYSIS

As soon as a monitor is ready for receiving records, there is a clear case for using a high-speed computer to attend to the storage, to bring stored records up to date if

additional information accrues, and above all to prepare analytical tabulations. These tabulations must themselves be stored and brought up to date as records are added to the store. Moreover, the computer can maintain a continuous scrutiny of the tabuations, and signal any major inequalities as soon as the evidence accumulates; without delay, medical expertise and additional research effort can be directed at the important points.

The appropriate statistical analysis depends upon ascertainment and class of records. Various possibilities are now considered:

E, III

Rates of incidence of event-types, relative to population numbers or to numbers in the patient set, cannot be estimated. All that can be done is to study the total number of cases of a particular event-type or event set, searching for any time-trends that may be informative. Maintenance of the same reporter set (e.g. a fixed panel of hospitals or of physicians) over a considerable time should give some assurance that totals of events in successive periods are comparable; there can be no certainty of this, but at best it will give some foundation for monitoring. Over a period in which the number of available reporters is increasing, simultaneous use of more than one reporter set will enable continuity to be imparted to the analysis.

Possibly some form of 'quality control' chart, based on the absolute frequency of an event-type in successive months, will be as useful as anything. No great refinement seems called for: merely a plotting of number against time would ensure that any major change was noticed. The scrutiny might be improved by passing the records through a computer program that produces estimates of linear trends and their standard errors, such as the trend since an arbitrary zero of time and the trend over the most recent 8 months. The intent is *not* to obtain exact tests of statistical significance; it is to extract sensitive indicators of the kind of disturbance that will occur if a new drug that is being introduced is increasing the frequency of a rare event.

If strong evidence of a change in event frequency is found, of course this does not immediately incriminate any drug. However, consideration of any recent changes in therapeutic practice may suggest an interesting association, and special pharmacological investigation can then be directed at it.

If geographic region (R1) and sex and age (R2) are recorded, the records may be classified to correspond to these, in the hope that analysis may disclose a trend not apparent for the whole body of records or may throw light on important features of a general trend. The practice of certain drug firms in limiting pilot marketing operations to particular areas may help the evidence.

Some who have written on monitoring have been disposed to assume that no good can be achieved unless percentage rates of incidence can be estimated. The simplest statistical procedure of all is to compare totals in two successive periods of equal length. If, under identical conditions for reporting, n cases in the earlier period are followed by $(n+d)$ in the later, and if a significantly positive d is regarded as *prima facie* evidence that a closer study is desirable, it is easy to see what minimal value of d

serves as a warning. Some typical figures are:

n	Minimal significant d at probability: 0·05	0·01
0	5	7
1	6	9
2	7	10
5	8	12
10	10	15
20	13	19
50	19	28
100	26	37
500	55	78

Examination of trends over several periods should increase the power of the test. In practice, however, formal tests of statistical significance are likely to be less important than other consequences of the discipline of recording and reporting events. The very fact that well-documented instances of events of a particular set are reported to one center should result in concomitant circumstances being viewed more critically by those who make the reports, and should ensure that a skilled medical scrutineer at the center becomes suspicious much earlier than would anyone under the present casual and unplanned pattern of letters to medical journals.

The method is far from ideal, and is unlikely to detect anything other than gross effects, but under the conditions of E, III no better formal analysis seems possible. Despite the deficiencies, it could give rapid warning of a new teratogen. Indeed, had it been in use in West Germany in 1958, suspicions of thalidomide would almost certainly have been aroused much earlier. TAUSSIG [6] stated that phocomelia was an exceedingly rare malformation before 1959; the incidence appears to have been at most 2 or 3 per year. Yet in 1962 she was able to find records of 17 cases in 1959 at 9 clinics, rising to 124 cases at these clinics in 1960. At no one clinic in 1959 were the cases numerous enough to cause alarm; had they and others been reported to a monitoring center, they could scarcely have escaped comment and further inquiry. In retrospect, it is easy to say that one of the clinics that in 1960 experienced more than 15 cases should have remarked on the phenomenon, but the explanation lies in the absence of any systematic collection and scrutiny of records and the fact that no one was looking for such things. Much is due to LENZ [7, 8] for his discovery in 1961, but a monitor could have signalled a warning 1½–2 years earlier. As KELSEY [9] has pointed out, additional clues were available, such as the manner in which the distribution of phocomelia in Europe followed the pattern of sales of thalidomide, and the incidence of cases among children of employees of companies promoting the drug and of doctors with early access to it. Any complacency induced by a feeling that the lesson has now been learned, and that no such tragedy could occur again, should be exposed to questions about the degree of assurance that no other malformations of infants are in part due to maternal drugs. When the basal incidence of a malformation, or any other event, in the absence of a drug is much higher than that of phocomelia, the situation will present far greater difficulties. Is it not possible that a drug taken during pregnancy may increase the risk of a spastic infant, or that a drug taken before adolescence may induce sterility?

E, II

The essential character of the analysis is determined by the fact of E ascertainment; the impossibility of assessing percentage rates is unaffected by the improved class of records. Consequently meaningful analysis will be possible only if the same reporter set is used over a considerable time, so that trends in the frequency of an event-type are related to incidence rates.

Once again, therefore, the analysis should consist in the production of control charts and perhaps the calculation of indicators of trend. Now, however, an event may usefully be classified according to the antecedent drug, so that charts and calculations may be made separately for each pairing of 'drug, event-type' (using a record more than once if more than one drug preceded the event). A trend that is noticeable in the totality of records may now be seen to appear only in the section of analysis for one particular drug-event pair, and not at all in the remainder. For example, in the simple two-period analysis already described, the records might show:

	n	*d*
All records of event-type	500	70
Drug Q alone	100	60
All other drugs	400	10

This would very strongly suggest an increase in the incidence of the event-type among subjects receiving drug Q; further action would depend upon whether Q was a newly introduced drug, whether the pattern of its use (diagnosis, dose, method of administration, etc.) had recently changed, or whether no changes in the circumstances of use were known to have occurred (this last making explanation of the records very difficult). The general principle is that any measure of the magnitude of an apparent trend will be judged the more significant the lower the level from which it began: an average increase of about 5 per month for 6 months would not amount to great evidence of trend if it began from 500 events per month, but would be highly significant if it began from 15 per month.

E, I

The analysis would be much the same as for E, II. There is the possibility that the analysis in term of trends in drug-event pairs can be subdivided, not only according to items of R4 and R5 but also according to items of R2 and R3. Thus the opportunities for tracking any anomaly to its source are increased. For example, if an event-type reported for newborn infants increased in frequency from 500 to 570, some suspicion would be aroused; if subclassification according to whether or not the birth was in hospital, or according to age of mother, were to show in one class an increase from 100 to 160, the evidence might begin to point strongly at an underlying cause.

D,III

With ascertainment D, information on R5 seems certain to be available, and class III records need not be discussed.

D,II

Completely different forms of analysis are now possible. Within the drug set, incidence rates can be calculated, each being the fraction of all patients receiving a drug for whom a particular event-type (or group of event-types) is subsequently recorded. Moreover, the records can be classified according to different items of R4 and R5, and rates calculated within each class of a one-way, two-way, or multi-way classification.

A general computer program could permit its user to specify any reasonable number of tabulations. For example, doubtless incidence rates for each drug in the set will be wanted, and then rates classified in two-way tables for drug and dose, drug and method of administration, drug and severity of disease or other feature of the diagnosis, and so on. Any items frequently recorded for R2 and R3 will also be used as factors for tabulation. The program should store all of these tables on tape, accept and store new records as they are validated, and bring each table up to date. The program might carry an instruction that each table is to be printed out after every addition of A records (where perhaps A=50). Moreover, the program should permit specification of a critical difference for each table, a magnitude selected by the analyst and easily changed by him; as soon as any difference between two rates in a table exceeds the critical value, that table will be printed out as a warning and a demand for closer scrutiny. Various devices for combining the evidence from different categories of patients may be useful, such as a summation of χ values (appropriately signed square-roots of the familiar χ^2 statistic) or the summarized comparison between totals of observed and of expected frequencies suggested by MANTEL [10]. The primary need is for a simply-appreciated comparison between two drugs that will focus attention on a consistent difference in incidence rates even though the magnitudes of these rates vary with sex, age, dose, and other recorded factors. Significance tests in themselves are not the objective, since these relate to the strength of the evidence rather than to the magnitude of the difference. There must be a recognition that, for the incidence of any specified event-type, large differences between different groups of patients almost always constitute a *prima facie* case for more detailed investigation.

The use of rates removes the requirement that attention be restricted to a fixed reporter set. Frequent extensive changes in that set are obviously undesirable, as they must introduce additional heterogeneities, but a steady expansion in the set should be very satisfactory. The reporters may not be typical of the whole population, but this will affect absolute incidence levels more than comparisons between groups of patients classified in respect of items of R2, R3, R4, R5; a signal from the analysis that a particular point needs investigation will be a rational basis for this action, even from a reporter set that makes no pretence at being representative.

Location of a difference in rates that exceeds the arbitrary critical value is no proof of harmful effects of drugs. As is well known, disproportion in a contingency table can be produced by accidental associations among factors other than those tabulated. For example, an incidence of an event-type apparently higher for drug Q_2 than for drug Q_1 might arise because Q_2 tended to be preferred by physicians for patients treated at home instead of in hospital, and patients at home are exposed to infection and other risks quite different from those in hospital. This bears on the contrast between surveys and experiments as foundations for inference, a subject familiar to statisticians as a source of fallacious reasoning on causation: when categories are

compared in respect of individuals who *happen* to be in them, instead of being randomly assigned to them, an apparent difference between the categories may in reality be due to the circumstances that brought them into the categories, rather than to properties intrinsic in the categories themselves.

Nevertheless, a warning that a difference in rates was now larger than the critical value could be made a signal to the computer to invoke a secondary program. This would cause the records to be read again and many more tables to be produced, all including the suspicious factors. For example, if the warning came in the form of a higher rate of some event-type for females on drug Q_3 than for males on Q_3 or for either sex on any other drug, the secondary program would produce and print 3-way tables (and perhaps even some 4-way tables) of rates classified according to drug, sex, and other available items of R2, R3, R4, R5. This program also would test tables for magnitudes of differences and signal the occurrence of large values. A further program could enable *either* the reference numbers *or* the complete records to be printed out for all patients falling within a specified cell of a multi-way table, in order that individual cases may be scrutinized. Thus the whole analysis would be directed at pinpointing exceptional incidence rates, looking at attendant circumstances, and bringing to light any associations that may be useful leads in the planning of the further investigations that will certainly be necessary.

Only rarely will the analysis just described disclose an incidence of some event-type so striking as to leave no doubt of its being severely inflated by side-effects of a drug; only rarely will the analysis itself justify immediate action to limit the use of a drug. A recent episode in connection with an anesthetic intended for intravenous administration provides an extreme example. A few patients suffered necrosis of the arm, necessitating amputation in order to avoid early death. Quite clearly a very small number of instances of this event-type would arouse suspicion. Even under conditions E,III, a monitor would soon signal a need for investigation. In fact, the event followed the drug so rapidly that D,II was much more likely to describe the conditions. The association with the particular drug therefore appeared immediately; the explanation, a special reaction because difficulties of technique occasionally caused the anesthetic to be injected into the artery instead of the vein, was then perhaps not very difficult to find. Admittedly so striking a side-effect with a clear drug history scarcely needs rigorous monitoring for its detection, but it illustrates the practicability of discovering adverse reactions to a drug without either knowledge of the natural incidence of an event-type or direct comparison with another drug.

Much more commonly, monitoring will do no more than arouse suspicions, and not infrequently further investigations will lead to explanations that wholly exonerate the drug. Though this will be a costly method of protection, it appears to be the price that must be paid for minimizing delay in the detection of adverse reactions. The cost of malformed infants, of drug-induced hepatic damage to patients with relatively minor ailments, and of other drug reactions is also high (Section 12).

The choice of drug set for D,II monitoring may need care. If it is restricted to one family of chemically similar drugs, the monitor could fail to produce any warnings because all the drugs were about equally productive of side-effects, even when the incidence of these events is high. Inclusion of an established standard drug, commonly used for cases of nearly the same kind, will be very desirable.

In contrast with the analysis proposed for E ascertainment, no comment has been made here on trends with time. These are evidently of less interest when rates can be calculated. Contemporary comparisons between drugs and between classes of patients are more valuable because they are not so liable to distortion from seasonal changes or from the effects of changing medical practice. None the less, information on trends in incidence rates, and in the relative frequencies with which alternative drugs are used, may be useful.

The proposed analysis by contingency tables requires that careful thought be given to the form of records in order to expedite the computer processing. Regrettably but inevitably, not every question that is used in constructing a record will be answered for every patient, and the gaps are sure to be irregularly distributed. If every item in the record allows for 'not known' as one of the alternative answers, the program can simply demand full classification in every table that is formed, so enabling all tables to include the same total of subjects. Subsequent scrutiny as a guide to warnings, however, should probably be applied only to those sections of a table for which none of the marginal headings state 'not known.'

D,I

The analysis would not differ essentially from that for D,II. Items now available under R2 and R3 would make possible a wider range of tabulations.

P,III and P,II

If P ascertainment is employed, information on R2 and R5 seems almost certain to be available, and records of classes II and III need scarcely be discussed.

P,I

This is unlikely to be common, because of the expense of P ascertainment except for restricted classes of patients. It may be practicable for a patient-set whose members naturally remain under observation for a period, for example pregnant women attending specified clinics, or for a population in which an intensive medical data linkage project is in operation. The greater potentialities of records from P ascertainment, however, make its use very desirable whenever circumstances permit. The main analysis will be much as for records classed as D,I and D,II. The important gain is that ascertainment through patients will bring in patients who receive no drug or who receive drugs outside the drug set. Inclusion of these in tabulations may provide an important standard of comparison for new drugs within the set, and so aid the detection of situations in which several similar drugs are all liable to inflate the incidence of adverse effects.

No attempt will be made at present to outline further the method of analysis required for P,I. An additional benefit to be kept in mind is that tabulations in respect of items of R2, R3 may help to bring to notice untoward happenings that are not drug-induced. Emphasis on drug hazards must not be allowed to divert attention totally from other sources of preventable misfortunes to hospital patients or to the general population. Moreover, P ascertainment will lead to accumulation of records from which quantitative estimates of the 'natural' incidence and severity of event-types can be formed; these estimates could represent an important advance on existing vague and subjective descriptions of an event-type as 'rare' or 'frequent.'

MELLIN and KATZENSTEIN [11] have recently illustrated the kind of numerical analysis that could be used for drawing attention to changes in incidence rates of an event-type, whether or not related to drug administrations. The setting of probability limits for differences between periods is proposed; as long as rigid dependence upon arbitrary significance levels is avoided, such a procedure may be a useful part of a monitor.

10. AFTER A WARNING

Much has been said about using a computer and statistical processes to draw attention to suspicious circumstances about the action of a drug. What should follow this? The emphasis should be on having the computer give early warning when accumulating evidence suggests that a drug has adverse side-effects and not to wait until the evidence is overwhelming. Consequently, many warnings will later prove groundless. Contrary to what may at first appear, this fact should provide reassurance to drug manufacturers against fears that monitoring will lead to widespread and mistaken condemnation of drugs.

In fact, the monitor is a method of systematizing, expanding, and expediting what happens at present. Instead of relying solely on individual medical practitioners to note anomalous drug reactions in their own experience, and to report these to manufacturers or the medical press in the hope that special investigations will be undertaken, computer scrutiny of a far greater number of records will direct attention to striking disproportionalities in the incidence of side-effects. The computer will analyze the situation further by preparation of additional relevant tabulations, but from that point on, medical expertise will again take over. Only exceedingly rarely would a case against a new drug appear with such dramatic suddenness and strength as to demand immediate actions to prevent its further use: if a drug commonly has serious side-effects, the monitor should give its warning correspondingly early, still before the case is strong. The normal procedure will be to regard the monitor signal as semi-confidential. Its occurrence will be reported to the manufacturer, who will be asked if he has any further evidence. Skilled medical workers in the right field will be consulted. The possibility that the warning originated in some bias of recording will be considered. If medical experts judge that the monitor evidence is sufficient to make desirable fuller chemical, pharmacological, or clinical investigation, they will so advise. Evidence from toxicity and clinical trials will have been collected at the time of first introduction of the drug, and may have been submitted as part of a licensing application; reference to this may be helpful. At this point, the logical position is no different from that obtaining when one or two letters in the weekly medical press have reported an unusual conjunction of drug and event-type, and so have stimulated a special inquiry to elucidate the matter.

The gain is that monitoring can be more comprehensive and surer, and therefore ought to allow this point to be reached with less delay. Once attention is drawn to a particular question, determination of whether or not a drug is dangerous in certain circumstances may be relatively easy; in this era of rapid development of new drugs, the real problem lies in directing attention to the right questions. The number of possible complexities of adverse drug reactions far exceeds the availability of men and resources for intensive study of each. A well programmed computer is admirably

suited to the task of screening a great number of records in respect of wide range of possibilities, and so of conserving medical skill and resources for medical tasks.

Under present conditions, early suspicions about adverse reactions to a drug are usually reported in the medical press, and rapidly brought to wider notice in daily newspapers. This involves adverse publicity for a manufacturer when evidence is still very slight. The growing use of drugs and the increased awareness of risks of adverse reactions are bound to lead to more numerous suspicions of particular events. In some instances, immediate publication may be desirable, but routine channelling of reports to a monitor would ensure that manufacturers seldom suffer from publicity given to suspicions that later prove false.

Large numbers of chemical compounds are screened by simple tests in order to reject those with no apparent therapeutic efficacy, as a prelude to more intensive examination of those that pass the screen; that there will at this stage be many false positives does not detract from the merits of the screening in concentrating true positives. In the same way, after drugs have been released for general use, monitoring is a screening process applied to drug-event pairs. That it gives many false warnings does not detract from its merits, provided that it also warns when a true situation of adverse reaction is encountered, and provided that responsible statistical and medical investigators advise on the next step appropriate to each warning given.

11. COMPATIBILITY OF RECORDS AND METHODS

Monitoring plans are likely first to develop at the national level, but this ought soon to be followed by an international program. If it be accepted that the primary object of monitoring is to direct attention to drug situations that merit further research with minimal delay, the desirability of producing tabulations based on records from several countries is obvious. In respect of a particular drug, five countries might have evidence that in no one instance suggested need for deeper inquiry; in combination, the evidence might strongly indicate an association between an event-type and the drug, even though possibly a rather more extreme probability level might be demanded as a first signal from an international monitor.

International combination of information will be practicable only if there is reasonable compatibility of national practices. Ideally one would like the record to have the same design for every country working in the same monitor field, so as to simplify processing at an international monitor center, though doubtless the programs could take account of small differences. Even if record design were identical, however, this would do no more than simplify the mechanics of data-processing and statistical analysis: it would not guarantee the validity of any conclusions. More important is agreement in classification of diagnosis, in terminology and standards for events, in measurement and description of drug doses, and so on. Establishment of an international center must not be delayed on account of a perfectionist attitude to these points, but efforts to secure reasonable compromises should be well worth while. Some suggest that no thought should be given to international monitoring until several countries are operating national centers; this policy is almost sure to lead to poor compatibility under every head. Informal consultation between countries at the stage of developing national monitoring seems the only hope of avoiding difficulties later. If several countries well-advanced towards systematic monitoring could form a nucleus of agreement on major aspects of compatability, not only would *they* gain

but the advantages to others of aligning themselves with this group would soon be apparent.

The emphasis here on an international center is not intended to discourage national scrutiny and analysis of records, from which many useful comparisons may emerge. Whether national centers should transmit all records or only a wide range of summary tabulations to an international center may be debated, but there must be no unnecessary delay in this transmission. Nor is there any intention of assuming uncritically that associations of event-types with drugs are the same in different countries. Classifications based on items of R1 should enable national differences to be examined, and indeed international collation of records is the only satisfactory way of studying points of interest and importance in this connection.

The general question of international action to provide safeguards against dangers in the medical use of drugs has been discussed elsewhere [12, 13]. The international portion of a monitoring organization should be a major component of a plan such as was there described. The World Health Organization should assume responsibility for this plan. Indeed, W.H.O. should act immediately to call together an expert committee on the principles and practice of monitoring. The recent proposal to create a World Medical Research Center, in which the handling of medical information is to be one of the main activities, seems very relevant, but the urgent need for monitoring can scarcely wait on this still uncertain project. If W.H.O. cannot take the lead, creation of alternative machinery for coordination of monitoring will be essential.

12. COSTS OF MONITORING

An ambitious and comprehensive system of drug monitoring such as has been described in this paper will be expensive. It will require experienced medical staff both for the initial scrutiny and validation of records and for the study and interpretation of any warnings emerging from a monitor; doubtless these will need reinforcement by advisory committees of experts in particular fields of medicine. It will require a statistical group competent to handle the problems of analyzing heterogeneous records with many gaps and irregularities, to take precautions against biases, and to prepare the many necessary summaries. It will require computer facilities and ancillary data-processing equipment, and of course staff skilled in the writing, day-to-day use, and steady improvement of a complex series of computer programs.

Even on a national scale, a monitoring enterprise will be costly, both in initial capital expenses for equipment and in operation. For the international organization that is urgently needed, the cost will be correspondingly greater. To estimate costs with any accuracy is difficult because so much depends upon the rate of flow of records to be processed. Evidently provision must be made for rapid growth. Any reasonable estimate of costs, however, will surely be a small figure relative to the potential savings. Another look at the thalidomide story indicates how much could be saved by monitoring of a single drug, admittedly in an extreme instance. A conservative estimate of the number of cases of thalidomide-induced phocomelia in Western Europe is 7000, and a monitor could have reduced this total to not more than 2000. Without any allowance for the immeasurable emotional strains, medical and social care of these children must average at least $10,000 per case (MINTZ [14], suggested a lifetime cost per case averaging $300,000, but this figure is perhaps extravagant even under North American conditions). A very good international monitoring organiz-

ation might run for quite a long time on $5,000,000, one-tenth of the very conservative estimate of thalidomide costs.

13. AN EXAMPLE

The need for monitors as safeguards against harmful effects of new drugs has been emphasized. Their importance must not blind anyone to their limitations and the difficulties inherent in using them.

A recent paper by WALFORD [15], on evidence for teratogenic effects of antibiotics, illustrates the point. In the present terminology, the data were of type P,II, though the information available (or at any rate used) in the records seems to have been slight. The investigation was retrospective rather than current monitoring, but the principle is unaltered. From several medical practices, records of pregnant women were extracted and classified in respect of infections (R4) and drugs received (R5) during the first 12 weeks of pregnancy. WALFORD developed the main argument in terms of the total of abortions, stillbirths, malformations, and neonatal deaths (R6). The relative frequency of this total 'pregnancy wastage' was higher for mothers who had received antibiotics than for those who had had febrile illnesses without antibiotic therapy or who had had neither illnesses nor drugs of any kind during the 12 weeks. The difference was statistically significant, but the author was properly cautious in drawing conclusions.

Many explanations other than the simple one of blaming antibiotics must be kept in mind. Such factors as general health, nutrition, and social class might be related both to the incidence of infections needing antibiotics treatment and to abnormalities; age and parity might for unexpected reasons differ between the antibiotic and no antibiotic groups. Even for this apparently straightforward example, the number of items from R2, R3, R4, R5 that could affect the interpretation and for which cross-tabulations ought to be attempted is great; moreover, the records are almost certainly incomplete in respect of many of them. WALFORD in fact made no attempt at further analysis, probably because the number of cases in the antibiotic group (50 out of a total of 1004) would scarcely stand subclassification, but the logical difficulties would remain even though a vastly greater number of records were to be analyzed.

SUMMARY

The thalidomide tragedy and other less dramatic evidence of the dangers that may be encountered in the normal medical use of new drugs have shown the need for establishing systematic recording of drug administrations and the experience of patients who have received specified drugs. The aim must be to create conditions under which hitherto unsuspected associations between a drug and a reaction in a patient are recognized as early as possible. To this end, emphasis should be placed on the recording of *events*, untoward happenings to patients, rather than on adverse reactions.

Section 1 of this paper contains many definitions concerned with drugs, patients, events, and their recording. These are designed to enable the principles and problems of monitoring the use of drugs to be discussed in objective general terms, without distraction by complications of detail occurring with particular drugs or types of recording. Of course, proper account of this detail must be taken before a monitor can be operated for any one set of circumstances, but there is at present great need for consideration of the common elements and of the advantages and weaknesses of

different types of program. Sections 2–5 extend the definitions by further discussion of events to be recorded, of methods of ascertainment, and of the nature of records.

The success of monitoring will depend upon the magnitude of the enterprise and especially upon the quality of recording (Sections 6, 7). Recording will improve only if those who make records are convinced of the importance of completeness and accuracy, and if they also have adequate clerical aid. A start may well be made with hospitals in which interest in good records and their use already exists, though even these may need additional clerical staff for ensuring a smooth flow of drug–monitor information. Once a good monitor is operating in a particular field, however small, it should become easier to persuade other physicians of the value of giving their cooperation.

The general character of statistical analysis of monitor records is outlined in Section 9 for different types of situation. Even the simplest type of recording, in which events are recorded but very few attendant circumstances, could have drawn attention to the rapid increase of phocomelia in West Germany 18 months before the thalidomide-association was discovered, and once suspicions had been aroused other clues would rapidly have led to thalidomide. A system of control charts could be used to follow trends. If recording is more full, great opportunities arise for continuous systematic tabulations of events, drugs, and characteristics of patients, with a view to early detection of important associations. The expectation is not that scientific proof of causations can be so obtained but that many early warnings of suspicious associations will be given; once such a warning is received, more orthodox processes of research investigation should be able to determine whether an appreciable number of untoward events are drug-induced or whether some bias inseparable from the recording is responsible. At present, the difficulty lies in deciding which of the almost limitless number of possible associations are worth study; a monitor should be invaluable in screening these (Section 10).

The handling of the large number of records likely to be produced, even at the national level, is a task for an electronic computer, equipped with programs that will store the records compactly and will keep up to date all the tabulations and analyses required. If information is to be used to best advantage, it must be collated internationally. All early planning for monitoring must therefore keep in mind the need for compatibility of codes and programs. As soon as possible, the World Health Organization ought to take the lead in promoting and assuming responsibility for an international monitoring organization (Section 11). The costs will be high, but not great, relative to the potential savings.

Any proposals for drug monitoring may easily be criticized because they are not comprehensive enough, because some items in the records will be incomplete or unreliable, because practical difficulties may prevent everything going according to plan, or because the statistical analysis proposed is neither logically perfect nor designed to cover every eventuality. The critic should remember that *at present nothing systematic is being done.* A monitor can achieve some success on clearly-defined issues, where the records are in little doubt, even though there are uncertainties about parts of it. To do nothing now because we have as yet neither the knowledge nor the administrative power to achieve perfection will be unpardonable.

We must make a start *now* with monitoring under relatively easy conditions; we must extend the scale and scope of our monitors as rapidly as increasing experience, reliability of recording, and quality of computer programming permit.

Acknowledgements—A number of friends and colleagues made helpful comments on an earlier draft of this paper. I am grateful to all, but especially to Professor E. A. CHEESEMAN, Dr. J. CROOKS, Dr. R. GOULDING and Dr. I. JUVANCZ for many constructive suggestions. They are in no way responsible for the opinions I express.

REFERENCES

1. MINISTRY OF HEALTH: *Safety of Drugs: An Invitation to all Medical Practioners*. H.M. Stationery Office, London, 1964.
2. KOLLER, S.: Einführung in die Methoden der ätiologischen Forschung—Statistik und Dokumentation, *Meth. Inform. Med.* **2,** 1, 1963.
3. MANDEL, H. G.: In *Interagency Coordination in Drug Research and Regulation. Part 2: Review of Cooperation on Drug Policies among Food and Drug Administration, National Institutes of Health, Veterans' Administration, and other Agencies*, p. 592. U.S. Government Printing Office, Washington D.C., 1962.
4. HARVEY, J.: In *Interagency Coordination in Drug Research and Regulation. Part 2: Review of Cooperation on Drug Policies among Food and Drug Administration, National Institutes of Health, Veterans' Administration, and other Agencies*, p. 383. U.S. Government Printing Office, Washington, D.C., 1962.
5. BUNDE, C. A.: In *Interagency Coordination in Drug Research and Regulation. Part 3: The Bureau of Medicine in the Food and Drug Administration*, p. 896. U.S. Government Printing Office, Washington D.C., 1963.
6. TAUSSIG, H. B.: A study of the German outbreak of phocomelia, *J. Amer. med. Ass.* **180,** 80, 1962.
7. LENZ, W.: Kindliche Missbildungen nach Medikament-Einnahme während der Gravidität? *Dtsch. med. Wschr.* **86,** 2555, 1961.
8. LENZ, W.: Thalidomide and congenital abnormalities, *Lancet* **1,** 45, 1962.
9. KELSEY. F. O.: In *Interagency Coordination in Drug Research and Regulation. Part 1: Review of Cooperation on Drug Policies among Food and Drug Administration, National Institutes of Health. Veterans' Administration, and other Agencies*, p. 97. U.S. Government Printing Office, Washington D.C., 1963.
10. MANTEL. N.: Chi-square tests with one degree of freedom; extension of the Mantel–Haenszel procedure, *J. Amer. statist. Ass.* **58,** 690, 1963.
11. MELLIN. G. W. and KATZENSTEIN. M.: Increased incidence of malformations—chance or change? *J. Amer. med. Ass.* **187,** 570, 1964.
12. FINNEY. D. J.: An international drug safety program, *J. New Drugs* **3,** 262, 1963.
13. FINNEY. D. J.: An international drug safeguard plan, *J. chron. Dis.* **17,** 565, 1964.
14. MINTZ. M.: Drug tragedy revives hope for cure, *Washington Post*, 19 Aug. 1962.
15. WALFORD. P. A.: Antibiotics and congenital malformations, *Lancet* **2,** 298, 1963.