5 Intervention

A tribal medicine man never teaches an apprentice quite the whole of his knowledge. Allopathic healers, on the other hand, are slowly shedding the cloak of mystery as they go about the business of intervening in the 'natural' course of the ills that befall us. Understandably, there is wistful reluctance on the part of physicians (and their patients) to reduce the therapeutic act to a visible formula. But dispassionate evaluation cannot begin until medical maneuvers are described fully.

SPECIFIED MANEUVERS

In a formal clinical trial, both the standard treatment and the innovation must be specified. The alternatives need to be stated in sufficient detail to permit others to repeat the maneuvers with precision, for it is the everyday application of the results of bedside studies that must be kept in view.

The standard regimen

The 'established' treatment is often difficult to pin down. Most regimens have been introduced informally, and variations on a therapeutic theme are the rule rather than the exception. Nonetheless, a consensus needs to be obtained about the orthodox approach for two reasons. First, if the trial is to be perceived as a proper challenge to accepted treatment, there should be few doubts that this standard was fairly represented. Second, if the sought-for contrast between treatments is to be made as sharp as possible, there should be an effort to reduce the irregularity of results. A compromise should be effected between the need to conduct a trial that is not far removed from everyday practice of medicine and the need to reduce the blurring effects of a loosely defined standard of treatment.

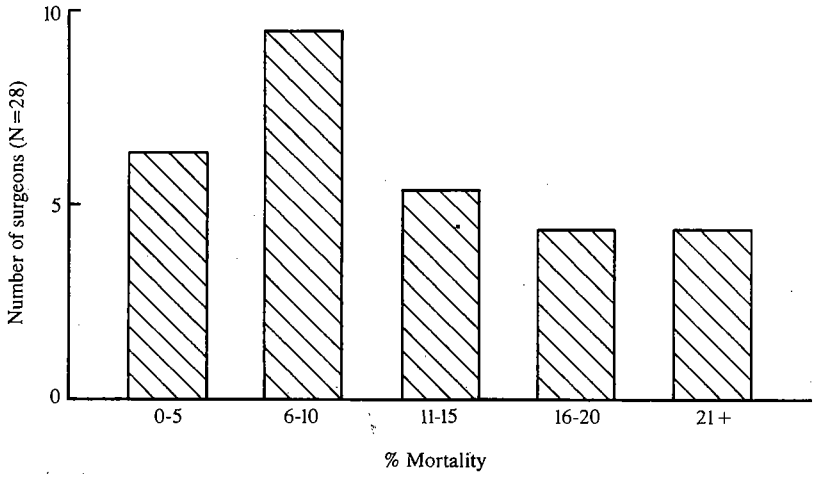
Oxygen monitoring in the RLF trial The 'standard oxygen treatment' of premature infants in the national study of RLF, for example, was defined as supplemental oxygen administered in a concentration over 50 per cent from age 2 days to age 30 days. This stipulation formalized the custom of

routine oxygen administration that had been used widely for years prior to the study in the management of small infants. However, the specified version of the convention called for measurement of the concentration of oxygen in incubators at eight-hour intervals. Since the clinical instruments for measuring this concentration easily and reliably had become available only a short time before the study began, the 'standard practice' as carried out in the study was not the usual practice followed in hospitals throughout the Western world; it was a regulated representation of the accepted approach.

The loss of realism and the gain in precision sharpened the comparison with the new proposal ('curtailed oxygen'—stipulated as use of the supplemental gas only as needed to relieve symptoms of oxygen lack, concentration not to exceed 50 per cent). But artificiality of the 'tight' plan exacted a price: extrapolation of the results of the study had to be made with considerable caution, particularly since the difference in treatment detail between study and everyday practice was not academic. It was later found that prolonged exposure to non-fluctuating concentrations of oxygen was needed to produce abnormal changes in the retinal blood vessels of experimental animals.

Uniform technical performance When interventions involve technical skills, the problem of developing a uniform standard is particularly difficult. The

Wide range of outcomes in cancer surgery



Mortality among 932 patients with cancer of the large bowel in a multicenter trial of surgical excision; the data were collected by L.P. Fielding and other surgeons at St Mary's Hospital Medical School (London). The wide range of post-operative mortality reported by 28 surgeons (each of whom contributed 10 or more patients) suggests that outcomes were to a considerable degree surgeon-dependent.

'same' surgical operation may have widely different outcomes in the hands of different surgeons. Similar problems arise when the non-specific effects of the 'healer' influence outcome (p 66).

It should be recognized, however, that the vagaries here are not unique in any qualitative sense. To a greater or lesser degree, the limitations imposed by irregularities in the therapeutic maneuver are a part of all clinical studies. Although the conflict between the demands of realism and those of order can never be completely resolved, it is important to take the time to develop a plan that will narrow the gap.

Bona fide treatments

Many clinical trials yield 'negative' results; the difference in outcome between groups receiving the standard treatment and others who are treated with the innovation is no greater than would be expected to occur frequently by chance. It is a challenge to planners to increase the amount of information that can be gleaned from a negative result, for there is a strong obligation to patient participants to ensure that a treatment trial is, quite literally, bona fide. The treatments compared should be alternatives that are offered *in good faith* that the results will be relevant to the welfare of those enrolled. Moreover, the predicted outcomes should have substantial implications that bear on current theoretical and applied problems.

Richard Peto of Oxford University has suggested several questions to be posed when planning treatment trials. What value will the results have if both treatment groups fare equally well? Can the design be altered to make a negative result even more valuable? A drug trial, for example, is always a test of the drug in the particular dose and manner given, not a trial of the drug *per se*. Distinctive comparisons, such as determining the largest dose of a new agent that can be given safely, provide more meaningful negative results than less marked contrasts.

Treatment of incipient RLF An instructive 'negative' experience took place in 1950 when the newly available drug ACTH (adrenocorticotropic hormone, which produces heightened activity of the adrenal gland) was used in a desperate attempt to halt the early blood vessel changes of RLF. Following very encouraging results in a pilot experience involving 31 infants, the potent agent was evaluated in a randomized clinical trial.

The formal test results indicated that the new treatment was completely ineffective in preventing blindness, and it appeared quite unlikely that this failure to confirm the preliminary findings could be explained by inadequate dosage of ACTH. All of the treated infants experienced temporary retardation of body growth, and the occurrence of serious infections (a feared complication of prolonged adrenal gland stimulation) was higher than in untreated controls.

This disheartening 'negative' trial revealed that early changes of RLF usually subside spontaneously (a previously unsuspected finding that was supported by observations reported two years later). Additionally, the trial had the immediate positive effect of protecting half of the enrolled infants from the risk of exposure to the potent hormone. And the discouraging experience served to warn others about the dangers of this fruitless approach to the prevention of blindness.

Dosage of treatment agents

The problems of flexible dosage in treatment trials are not easily solved: particularly when the end point for potency is difficult to define, when there are technical difficulties in monitoring on going effects, and when there are multiple effects that must be taken into account. In the national RLF trial many of these difficulties were recognized—in retrospect.

Oxygen dosage in the RLF trial It was decided in the oxygen-and-RLF trial that 50 per cent oxygen concentration in inspired air would serve as the boundary between the two management regimens. This ceiling was chosen for the 'curtailed oxygen treatment' group because it was believed (in the early 1950s) that concentrations in the range of 40–50 per cent were adequate for relief in infants who required treatment with this life-supporting gas. The higher concentrations specified for 'standard' management acknowledged the established practice of keeping very small infants in highly enriched environments on the assumption that this regimen improved respiratory performance, thereby reducing the risk of death and of brain damage in survivors.

Oxygen concentration on either side of the 50 per cent demarcation line was not specified by protocol; in fact, a wide range of concentrations was used. The duration of treatment in the 'curtailed-use' group was on an 'as needed' basis (length of exposure varied from zero to as long as three weeks). When the results of the trial indicated that the risk of RLF was reduced with no apparent increase in mortality among infants who received curtailed amounts of oxygen, many questions surfaced about the relationship between dosage and outcome. For example: Does the risk of RLF increase with rising concentrations of oxygen, with longer durations of exposure, or with some function of concentration and duration? Most important was the question: What is the risk of eye damage when oxygen is administered only during the first two days of life, the period excluded in the trial?

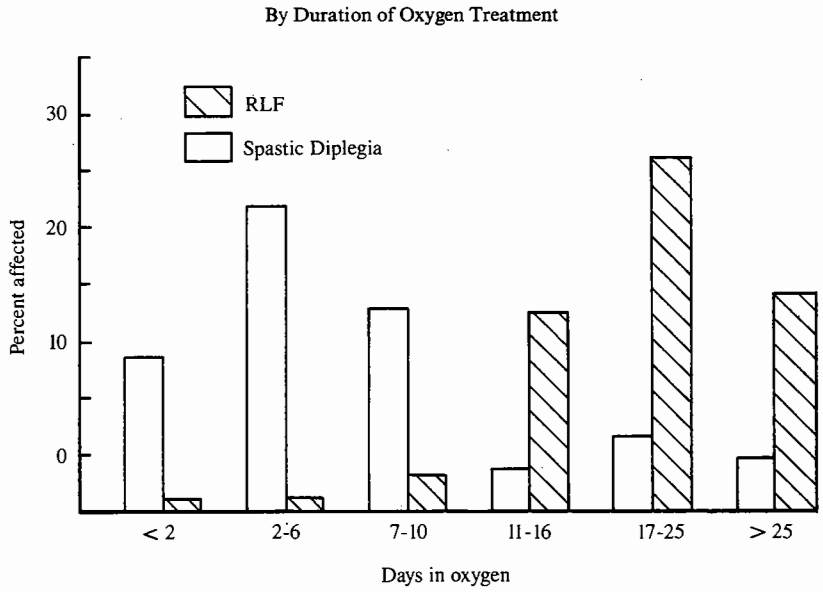
The large body of information collected in the study of more than 700 babies provided an opportunity for a 'data-dredging' search to formulate very specific hypotheses concerning the dosage of oxygen and outcome. And the sifting operation was, indeed, carried out; it was observed, for

instance, that RLF occurred more frequently with increasing duration of exposure, but not with increasing concentration of oxygen. The weakness of such associations, however, was self-evident in the trial design: incremental differences in oxygen exposure (within the two broadly stated treatment prescriptions) were made not by lot but according to individual judgments of treating doctors. The connection between dosage and RLF risk was completely undermined by uncertainty about confounding influences; the sickest babies received oxygen treatment for the longest periods and in the highest concentrations within the wide limits defined in the trial protocol.

Urgently needed tests of the unsolved questions concerning oxygen dosage simply could not be undertaken in babies because of ethical constraints. And studies in other species did not resolve the problems. Although early changes of RLF could be produced, blindness did not occur in experimental animals exposed to oxygen.

In the years following the announcement, in 1954, of the RLF trial results, the questions concerning the exact relationships between oxygen dosage and the well being of premature infants became more complex.

Cerebral palsy and RLF



Among 194 prematurely-born children who received oxygen treatment, cerebral palsy (spastic diplegia) occurred more frequently with short than with long exposures. A reverse trend was noted for RLF.
(From the report of Alison D. McDonald)

Surveys of treatment practices suggested a conflict of outcomes with oxygen curtailment. Short exposures to supplemental oxygen were accompanied by relatively low frequencies of RLF, but spastic diplegia (a form of cerebral palsy) was found to occur more often with curtailed treatments. Additional doubts about the overly simple extrapolations concerning oxygen dosage were triggered by surveys of the trend of mortality among premature infants. The curtailed oxygen practice beginning in the mid-1950s was followed by a halt in the downward trend in mortality rates of babies under 1 day of age. The important question raised by these disturbing observations was: How do the *intertwined* risks of death, brain damage, and RLF blindness vary with increasing exposure to supplemental oxygen?

It was reasonable to postulate, for example, that at the cost of minimal increase in the risk of RLF, there would be a large reduction in the risk of brain damage and of death for infants exposed to doses of supplemental oxygen that lie between the extremes formally tested in the national trial. But ethical barriers, once more, made it impossible to test the thesis rigorously. The issues remain unresolved to the present day.

Limitations of a single trial

The prolonged uncertainties in the use of a familiar substance, oxygen, for an everyday problem such as the management of newborn babies, are not particularly unique. These kinds of complexities are commonplace in the evaluation of interventions in medicine. It is unreasonable to expect to settle all of the important questions concerning treatments with a single well planned trial. Initial decisions about new therapies almost always need to be modified, and it is realistic to plan, from the outset, that a *series* of trials will be required to bring the boundaries of efficacy and the magnitudes of trade-off risks into focus. The plodding pace of clinical trials is so frustrating that there is a strong temptation to wring as much meaning as possible from each experience in the hope of avoiding additional or phased formal tests. Unfortunately, as in the RLF experience, the short cuts are full of unexpected traps.

EVENHANDED TREATMENT

It is completely unrealistic to expect that clinical trials can be carried out by neutral investigators who are not 'betting' on the final results. Personal bias is a given in all studies and it is necessary to devise fairly elaborate schemes to thwart or, at least, diminish its effects.

'Masking'

The infelicitious term 'double blind' refers to the strategy in which neither the physician nor his patient knows which of the alternative treatments

under test has been given. Although this guarantees that the treatments will be administered with an even hand and that observation of results will be unbiased, there are a number of practical difficulties that limit the approach. The disguise may be imperfect or, as in surgical treatment, it may be impossible to achieve. Additionally, it creates workaday problems (especially adjustment of dosage) that are difficult to overcome. And the artificial trial conditions are completely different from those of the real world to which the results are meant to apply.

Other safeguards to ensure objectivity in the application of treatments and in observation of results, such as using third party observers who do not know which treatment the patient received, have advantages over highly contrived measures that keep the physician from knowing which treatment has been prescribed. The term 'single blind' is used when one of the parties is kept from recognizing treatments and the 'masked' class (treating physician, observer, or patient) should be specified.

Experimenters' effects

Although it is not always possible to devise a practical 'masking' plan, the consequences of experimenter bias are not easily dismissed. Two major types of influence need to be considered; both are usually unintentional. The first operates 'in' the observer without modifying the response of patients (the effects, to be discussed in Chapter 6, are felt in the notation of phenomena). The second type of experimenter's effect is interactional. It modifies the actual response of patients to maneuvers and thus can be considered to be an element of these interventions.

The physician component of treatment There is evidence (especially in behavioral research) that the experimenter's orientation leads to interaction with subjects in ways that increase the likelihood that the response will confirm the proposed hypothesis. In medical settings, physicians tend to offer patients cues about what is expected through voice quality and nonverbal body behavior, even though the content of instruction is standardized. The biasing effects of such cues on the responses of patients are not entirely controlled by shutting off communication, as nearly all people have learned ways of acting in the presence of authoritative figures. The self-fulfilling prophecy effects are exaggerated by psychosocial influences: experimenters high in status tend to obtain more conforming responses from their subjects than do lowly researchers.

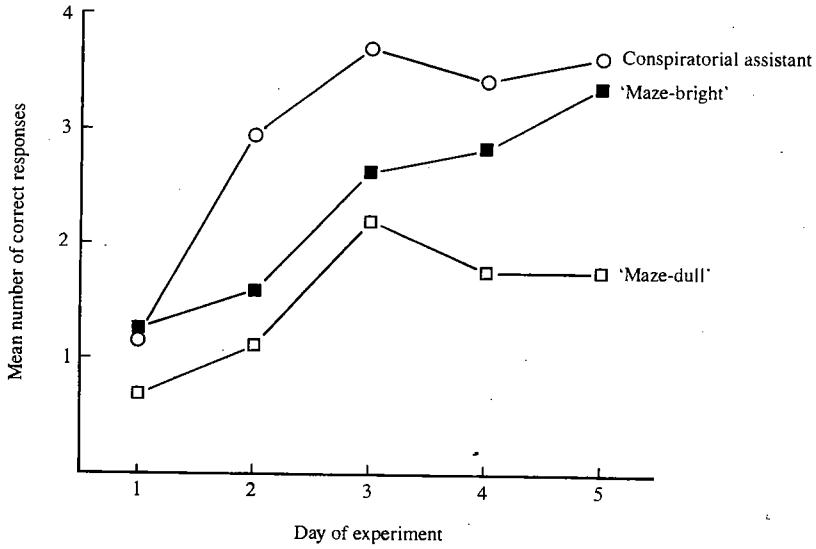
The personality of the experimenter may influence the results of behavioral research: friendly examiners administering standardized tests of intelligence are likely to obtain better intellectual performance than are more aloof examiners or those who are perceived as threatening or strange by examinees. These phenomena have been summarized by the phrase 'the

demand characteristics of the experimental situation' which include all of the cues which govern subjects' perceptions of their roles and the investigator's hypothesis.

When patients trust their physicians they act in ways that are meant to please; compliant patients have a remarkably intuitive ability to sense what is wanted of them and they provide it. In long-term studies, participants often come to identify with researchers and their aims and feel obliged to conform as a self-imposed condition for maintaining the bond. In follow-up studies of children, for example, mothers' caretaking behaviour may be favorably influenced by researchers, and this becomes an unevaluated co-intervention that alters the outcome-of-interest in the child.

The expectancy phenomenon The effects of investigator's expectations have been tested extensively by social psychologists, notably Robert Rosenthal

Effect of experimenter's expectancy on maze learning in rats



In Rosenthal's investigation a group of students were given written instructions concerning an experiment in rats. They were told that continuous in-breeding had led to a strain which was considerably better than 'normal' rats in maze learning ability; these rats were labeled 'maze-bright;' another strain was 'maze-dull,' the students were told. The rats were, in fact, ordinary rats and they were assigned in random order to thirteen groups, five animals in each group. Six students conducted maze-learning 'runs' with 'maze-bright' rats, six used 'maze-dull' animals, and a research assistant who knew the truth conducted 'runs' on the thirteenth group of animals (she was instructed to get as good performance from her animals as possible without violating the formally programmed procedures). Each rat was 'run' for 10 trials per day over a 5 day period. When the animals turned into the darker of two platforms of a T-maze, they were rewarded and the response was scored 'correct.'

of Harvard University. He has found that experimenters may change during the course of their experiments under the influence of responses obtained from the first few enrollees. When the first few subjects react according to expectations, the behaviour of experimenters is modified in ways which influence subsequent subjects to conform to the prediction of the hypothesis.

In one trial, the expectancy phenomenon was evaluated by 'creating' researchers who differed only in the expectations they held about the result of a specific experiment involving laboratory animals. The participants were given rats that were to be taught to run a maze with the aid of visual clues. Some were told that their rats had been specifically bred for 'maze brightness'; other were told that their rats were 'maze dull'; and a research assistant was told the truth: the groups of rats had been formed by random assignment from a common pool. The assistant was urged to obtain a good performance from her animals. At the end of a five-day trial, the results indicated that rats assigned to experimenters expecting brighter behavior showed better learning compared to rats given to investigators expecting dull performance. The conspiratorial assistant obtained the best results of all. The experience suggested that the experimenter who was urged to obtain good performance from animal subjects achieved even better performance than others who were biased to *expect* good performance but were not explicitly instructed to obtain it.

If rats became brighter when expected to by an investigator, it seemed possible that children might become more intelligent when expected to by their teacher. 'The Pygmalion Experiment' was conducted among elementary school children and the results supported the general thesis that one

The Pygmalion experiment

In Rosenthal's exploration of the expectancy phenomenon among human subjects, all the children in an elementary school were administered a nonverbal test of intelligence, which was disguised as a test that would predict intellectual 'blooming'. There were eighteen classrooms in the school, three at each of the six grade levels. Within each grade level, the classrooms had been divided by the school according to ability: above average, average, and below average. Within each of the eighteen classrooms, approximately 20 per cent of the children were chosen at random to form the experimental group. Each teacher was given the names of the children from her class who were in the experimental situation. The teacher was told that the score attained by these children on the 'test for intellectual blooming' indicated that they would show remarkable gains in intellectual competence during the next eight months of school. The only difference between the experimental and control group children, then, was in the mind of the teacher.

At the end of the school year, eight months later, all the children were retested with the same intelligence test. The children from whom the teachers had been led to expect greater intellectual gain showed significantly greater increase in scores than did the children of the control group.

person's expectation can come to serve as a self-fulfilling prophecy for the behavior of another.

The implications for medical studies are fairly straightforward: the 'experimenter variable' (the term 'treatment behavior' has been used to characterize the social acts of those who profess to heal) cannot be ignored when planning bedside experiments.

USE OF PLACEBOS

A difficult question arises when a proposed intervention must be evaluated in a disorder for which there is no generally accepted standard treatment. Should a placebo, either an inert substance or a sham procedure, be prescribed for patients assigned to the untreated control category in a clinical trial? The dilemma is not easily resolved because the arguments are enmeshed quickly in a dense thicket of ethical issues.

Some public attitudes about the question were revealed during the years when attempts were made to reduce the risk of paralysis from poliomyelitis. In the 'double masked' trials of the 1950s to evaluate *prevention* of poliomyelitis, parents readily gave permission for enrollment of their children: 200 745 children received an injection of Salk vaccine and 201 229 received ineffective salt solution. On the other hand, there was considerable resistance to the conduct of a controlled trial challenging Sister Kenny's technique of hot packs which was used widely in this period for the early *treatment* of poliomyelitis.

A number of arguments can be mounted in favor of the use of placebos since inert materials and sham maneuvers may produce bodily effects that are quite striking and, many times, beneficial. There is, in fact, a placebo component in every clinical intervention and it is difficult to distinguish between specific and non-specific effects. Placebo responses may mimic effects produced by potent agents; the force of the phenomenon seems to derive from the ways in which patients interpret the therapeutic behaviour of their doctors. The literal definition of the word placebo is 'I shall please.'

A remarkable example of the complex interactions in medical treatment was cited by Stewart Wolf of the University of Oklahoma. He told of a patient with long-standing and almost continuous asthma who obtained no relief from a series of drugs tried by the treating doctor. When an experimental drug with high promise of effectiveness became available, a supply was obtained by the doctor. The new drug relieved the asthmatic symptoms immediately; when it was stopped the asthma returned. The doctor substituted a placebo without the patient's knowledge; this failed to relieve symptoms. Shifts from new drug to placebo and back again were tried several times with consistent results in favor of the experimental agent. When the pharmaceutical company was approached for an additional supply, the

doctor was amazed to learn that, because of worry about unjustified claims, the company had in this instance provided only a placebo preparation. The experience is a vivid reminder of the blurred distinction between treatment and non-treatment in placebo-control trials when doctors and patients are 'masked.'

Deliberate use of placebos has been a part of the healing art for centuries. Although there is tacit acceptance of the practice, it cannot be denied that patients may feel deceived and denigrated when sham treatments are prescribed solely to please them. ('Dummy,' the British term for a placebo is embarrassingly close in meaning to 'dolt' and 'stupid person'.) But the situation is arguably different in an organized effort to compare placebo and untested treatment. Here the protection of individual patients from unpredicted adverse effects is an uppermost goal and there is, I believe, ample justification for the tactic of placebo control.

The subject is controversial; pros and cons deserve a full debate whenever the use of a placebo is considered for a specific clinical trial. A compromise must be negotiated between the demands of objectivity, the obligation to uphold the dignity of participants, and the need for risk containment.

Multiple interventions

When compound treatments (multiple interventions applied concurrently or serially to the same patient) are tested, it is necessary to make simplifying assumptions to justify the approach. These premises are not badly strained when a single outcome-of-interest is assessed, but they are quickly overwhelmed when more than one result is evaluated.

Cross-over design

When the effects of treatments are temporary (that is, the effects during the first period do not carry over to the period of the next intervention) a cross-over design may be employed in which the patient serves as her/his own control. Comparison of three treatments (A, B, and C) can proceed by randomizing both the order of treatment and the entry of patients into the study:

	Treatment Periods		
	1	2	3
Patient 1	A	B	C
Patient 2	B	C	A
Patient 3	C	A	B

This 3 × 3 set of assignments is repeated so that at the end of the study there are an equal number of patients who have had treatments in each of treatment orders (for instance, A first, A second, A third, B first ...) The effect of order is now controlled and comparison between the three treatments can proceed.

Serial designs

In serial designs, the patient is exposed to different maneuvers according to a predetermined schedule, or successive treatments may be applied on a conditional basis.

Cross-over trials The cross-over design is a scheduled plan that tests for transient effects of alternative treatments at different times in the same patient. The concept is attractive in that it increases the sensitivity of a trial by reducing between-subject variation in treatment comparisons. It must be assumed, however, that the patient returns completely to pre-treatment status with no carry-over effect between treatment. It should also be noted that a trial with an adequate number of patients for a sensitive comparison of treatments may be too small for detecting interaction between treatment and order.

Conditional treatments In one form of serial treatments a second maneuver is applied *on condition* that the initial intervention has produced a specified result. The approach is used in the management of chronic illness and experimental trials that employ this design may mimic reality fairly closely.

For example, infants enrolled in an oxygen management trial who develop early signs of RLF may then receive an additional maneuver purported to decrease the risk of blindness. The interpretation of the results of the compound trial will be highly dependent upon the ability to establish a uniform standard for the classification of 'early signs of RLF.' And care must be taken to ensure that the outcomes are not distorted by the effects of different oxygen treatments administered before the conditional maneuver was employed.

Some biases in executing the experimental maneuver

Bogus control bias:
: When patients who are allocated to the experimental maneuver die or sicken before or during its administration and are omitted or reallocated to the control group, the favorable impression of the experimental maneuver is enhanced.

Co-intervention bias:
: In an experiment, when members of the experimental group receive treatments, attention or care (in addition to the experimental maneuver) which is not provided in equal amount to control, differences in outcome between experimental and control subjects may be spuriously attributed to the experimental maneuver.

Compliance bias:
: In experiments requiring patient adherence to therapy, issues of efficacy become confounded with those of compliance.

Contamination bias:
: In an experiment, when members of the control group inadvertently receive the experimental maneuver, the difference in outcomes between experimental and control patients may be systematically reduced.

Therapeutic personality bias:
: When treatment is not 'blind,' the therapist's convictions about efficacy may systematically influence both outcomes (positive personality) and their measurement (desire for positive results).

Withdrawal bias:
: Patients who are withdrawn from an experiment may differ systematically from those who remain.

(Described by Sackett)

PROCEDURAL BIASES

The details of medical management are specified in the protocol of a clinical trial. When systematic departures and unexpected complications in carrying out the plan occur, they may overshadow other influences on outcome.

Co-intervention

The term 'co-intervention' refers to procedures (such as examinations, diagnostic tests, and auxillary treatments) that are performed on enrolled patients. If these procedures are not performed with equal vigor on the members of each group, confusing distortions may result.

In one survey of the relationship between RLF and the partial pressure of oxygen in the blood of premature infants, the schedule for making these measurements was not specified. When the results were inspected, there appeared to have been a systematic bias toward more frequent measurement of the oxygen status in babies at greater risk of developing RLF. The resulting confusion from this defect in study design could not be unraveled (with any confidence) by means of statistical adjustments.

Contamination

Inadvertent errors in the management of patients cannot be eliminated entirely in the real world, and they occur, inevitably, in the best managed trials. A patient may receive the same or a related treatment as the one used in the comparison group, for example, and a decision must be made about the disposition of this 'contamination' of the orderly trial. There is a strong temptation to shift patients who inadvertently receive the alternative treatment to the 'correct' group on the logical grounds that this event occurred 'by chance'. But a much stronger case can be made for retaining the treated-by-error patients in the allocated group despite the error. In a pragmatic trial, the report of the number of such errors provides a useful bit of information because it provides a basis for estimating how often these may be expected if the regimen is adopted for general use.

Similar problems arise with withdrawals from treatment and with failures to comply with the prescribed course of treatment. All of these irregularities, again, provide important insights about the problems that will be

encountered when the results of the formal trial are translated into everyday actions.

Need for vigilance

Hugo Muench, of the Department of Biostatistics at the Harvard School of Public Health, became a legend in the field of biometry before his death in 1972. He proposed a number of droll 'laws' that provided considerable insight into the problems encountered in conducting clinical studies. A corollary of one of these is appropriate in summing up the intricacies in dealing with the issue of interventions. It reads: 'Anytime that things appear to be going well, you have overlooked something.'

Somewhat the same advice for those planning controlled trials was given by Donald Reid of the London School of Hygiene who recalled the words spoken by the White Queen to Alice in *Through the Looking Glass*: 'Consider what a great girl you are. Consider what a long way you've come today. Consider what o'clock it is. Consider anything ...' These words of caution have additional relevance at the stage in clinical trials when physician observers describe the outcomes of their interventions—the topic of the next two chapters.