PRINCIPLES OF CHEMOTHERAPEUTIC SCREENING

NATHAN MANTEL
BIOMETRY BRANCH, NATIONAL CANCER INSTITUTE

1. Introduction

Progress in the field of cancer chemotherapy has been characterized recently by the development of broad screening programs for the assay of candidate therapeutic agents. A landmark in this field was the Gellhorn and Hirschberg [1] report on "Investigation of Diverse Systems for Cancer Chemotherapy Screening" which was devoted to an examination of a variety of biological systems potentially useful in the selection of antitumor compounds. This report covers the results obtained employing 27 compounds in 74 biological systems and provides much worthwhile information for the formulation of a satisfactory screening program.

Two points from the Gellhorn-Hirschberg report may be worth noting at this time. For one thing, no single tumor system employed appeared capable of selecting all the useful agents tested and it was recommended that a spectrum of tumors could provide a "greatly improved screening system." For another, the results seemed to indicate the general unsuitableness of nontumor systems as screening tools for carcinostatic agents.

While results such as those given in the above mentioned report are essential in the planning of screening programs, equally important is the following of certain guidelines and principles which are appropriate to any screening program. The use of proper guidelines will tend to minimize the cost in time and effort required to attain the objectives of a screening program. In the discussion below a number of these principles will be considered. It should be remarked that, in general, in the use of screening programs the principles which will be discussed below, while not specifically stated, are implicitly followed. The discussion below will also deal with the relationships between screening programs and laboratory development procedures. Implications with respect to a suitable cancer chemotherapy screen will also be covered.

2. Objectives of a screening program

Essentially, the purpose of a screening program is to select for further use or study the more promising materials being processed. A program may be designed

National Institutes of Health, Public Health Service, U.S. Department of Health, Education and Welfare

to select materials worthy of fuller development and investigation in the laboratory; alternatively it may be geared to finding more effective compounds than are presently available or perhaps even to finding especially effective compounds.

A particular screening program may employ stages, the selection becoming more restrictive at each stage. Thus at one stage compounds showing any promise may be selected for further testing while at the next stage some specific high level of effectiveness would have to be demonstrated. The purpose of such a multistage screening program would be typified by the most restrictive stage of testing employed.

Probably the most important question in the design of any screening program is the selection of the stages of screening employed. This is discussed immediately below.

3. Stages of screening

The assay of a series of compounds for therapeutic effectiveness can, in principle, be conducted on a clinical basis. Various factors mitigate against this being done directly. Subjects for testing may be too limited for assaying a large series of materials; clinical testing may be too expensive to be used routinely on all compounds, or the duration of the clinical test may be too long. Of course, there is also the unwillingness to test clinically any drugs for which no basis has yet been established for expecting as good or better results than are yielded by current therapeutic procedures. There is of course also the possibility of undue toxic effects from the new treatments. This can be forestalled by appropriate toxicity studies prior to clinical use. In the opinion of the author such safety testing constitutes an adjunct to a screening program and is not a part of the screen in itself.

When, and ordinarily this is the case, clinical testing is to be preceded by experimental screening, the screening test used should be such that the results it yields are expected to correlate reasonably well with clinical drug effects. The simple elimination by an experimental screen of the vast bulk of compounds showing no promise at all may, in some instances, be sufficient to make it feasible to test the remainder clinically.

The principle illustrated by the use of an experimental screening stage prior to clinical screening is that is can be used to achieve economies in time and effort. But this principle can be applied effectively at the experimental screening level as well as the clinical. Prior to entering materials for testing in some final preclinical screen it may be desirable to test them with some simple procedure which will cull out compounds unlikely to pass the preclinical screen. A procedure requiring half the effort of a preclinical screen which eliminates 90% of the compounds still produces a 40% savings in overall preclinical testing. Such savings can be increased considerably by procedures which combine extreme simplicity with a high rate of rejection of ineffective materials. The use of such

a preliminary procedure is premised on the assumption that compounds capable of passing the final screen will in all probability pass the preliminary screen.

By extension one could in some cases justify employing a series of screening tests making it feasible to test extremely large numbers of compounds, the number of compounds screened at each stage decreasing as the complexity and rigor increase. As the number of compounds tested at each level decreases it becomes possible to employ more sophisticated experimental procedures resembling, perhaps, those which would ordinarily be used in a laboratory research program rather than in a screening program. Such procedures may give insight into the nature of drug action. Primarily, however, they are designed to eliminate compounds not having sufficiently high activity and compounds having nonspecific effects. A characteristic of screening programs employing stages is that they have built-in safe-guards which systematically eliminate compounds of nonspecific activity. Such nonspecific effects may be adequate for passing one particular stage of testing but not the next. The screening program does not differentiate between compounds of insufficient activity and those of nonspecific activity; this is opposed to the goals of a laboratory research program where such differentiation may be of paramount interest.

Some of these ideas are exemplified by one of the more common procedures employed in the preliminary testing of drugs for anticancer activity. The test consists simply of ascertaining whether or not the drug can sufficiently inhibit the early growth of transplanted tumors. Mice are implanted with tumor inoculum, treated daily for say 7 days and the tumors then excised and weighed. It seems reasonable to expect that any material having important anticancer effects should, as a minimum, be able to meet this challenge and inhibit the growth of the tumor to some important degree, say 75%. In some cases this challenge can be met by drugs not having specific anticancer effects, the observed inhibition of tumor growth being a reflection of drug toxicity for the host, host dehydration occasioned by treatment, or perhaps even loss of appetite resulting from treatment. These possibilities can be ironed out later by further investigation for which this procedure constitutes only a prelude.

It should be remarked that limited as the challenge is in the above-described preliminary test procedure it has, in practice, been effective in screening out on the order of 99% of the materials tested; the remaining 1% can then be tested by more exhaustive procedures. The screening out of ineffective materials is so efficient in this case as to give the impression that materials passing the challenge have good antineoplastic properties; all that has been done is to reduce sharply the number of materials to be tested further. The proportion of really effective compounds may have been vastly increased from an extremely low concentration among the materials originally tested but it can still remain rather low among the compounds successful in passing the challenge.

The system of screening in stages is reasonable when at the successive degrees of effectiveness which materials are required to demonstrate the necessary testing procedure is more complex or expensive or for some reason cannot be applied

on a large scale. These conditions may not necessarily always apply. In some instances it may well be that a screening procedure for ascertaining the existence of a high order of activity may be simpler and easier to apply than one designed to detect lower orders of activity. If the intention were to detect compounds with the higher order of activity it would be efficient in such case to employ the procedure appropriate for that order of activity alone. It may be important in designing a screening program to recognize situations such as this where it is desirable to avoid the more preliminary screen. This possibility will be discussed further below and will be illustrated with a specific situation in which the author would suggest immediate use of a greater challenge.

4. Criteria for selection

4.1. *Degree of response.* The use of any particular screening system or of any particular screening stage it employs must be founded on some degree of knowledge of its properties. This knowledge as a minimum must be sufficient to justify the assumption of a reasonable degree of correlation of screening results with the ultimate objectives of the screening program. Further, it is desirable to have some knowledge of the weaknesses of the testing program—what, for instance, are some of the nonspecific procedures to which the test is responsive and what is the extent of such nonspecific effects? This kind of information is of use in prescribing the degree of response required at a stage of testing to warrant further testing at a more advanced stage. Let us consider how such passing levels of response are related to the objectives and structure of the screening program.

4.1.1. A screen may be designed to pass materials showing any potentiality, adequate statistical safeguards being taken to ensure against too many ineffective materials passing or effective materials being overlooked. The justification for passing materials showing some minimal degree of effectiveness is that there is not too high a correlation between test results and ultimate clinical performance. The test presumably can weed out ineffective materials but cannot differentiate too well between degrees of ultimate performance. The ultimate potentiality of any compound which passes must still await further investigation and development.

4.1.2. The screen may be designed to pass only materials showing some specified high degree of effect, this level of effect perhaps being greater than any yet achieved. In this case the tacit assumption is that the correlation with ultimate clinical performance is so high that one cannot reasonably expect a drug to provide satisfactory clinical results if it does not meet the specified level of experimental performance.

4.1.3. The passing level of effect required by a screen may be variable, being shifted gradually upwards as better and better materials are found. When any particular drug is tested it is required to test out at least as good as previously uncovered promising materials. This variable criterion would apply in situations

where the test effect and the clinical effect are anticipated to be reasonably well correlated and, in addition, the available capacity for further testing and development is quite restricted, say to a point where only a few compounds can receive advanced testing. The use of a progressively advancing criterion restricts such advanced testing to the few most promising compounds.

When the critical level of performance is geared to the peak capacity of compounds already tested the possibility of an abrupt shift in the testing program may open. This would apply in the case where the test for a high order of activity is simpler than for a low order. So long as we do not know of any compounds with especially great effects we can be content with selecting compounds with some activity even though of a lesser order. With the discovery of the first compounds which are especially effective we may require comparable activity of subsequent compounds tested. This results in a reduction in screening effort in the case where a simple test exists for the higher order of activity.

4.1.4. In practice, a reasonable compromise exists between the fixed and the variable criteria for selection. Assume that some fixed level of response has been set for determining whether or not a compound should receive secondary stage testing. If in any particular time period the number of compounds passing the primary test exceeds the secondary screening capacity of the program, it is reasonable that the most promising compounds should receive priority in selection for such secondary testing. Compounds tested in previous periods in the primary screen or stage should also be considered in setting priorities for secondary screening. Effectively, the critical point of the test is modified and an equilibrium is achieved between primary screen results and secondary screen testing. If both primary and secondary testing are conducted at constant rates per period of time, the equilibrium will be achieved at a criterion corresponding to some percentage point in the distribution of effectiveness of compounds on the primary screen. If substantially more compounds are being passed at a stage than were initially anticipated and hence more than can be tested in the subsequent stage, it may be advantageous to raise the required criterion of performance. This will permit a reduction in experimental effort in the earlier stage.

In addition to the savings resulting from using multiple stages of screening, additional savings can be obtained by applying similar procedures within a stage. This involves the use of steps or substages at any particular stage of screening. The use of such sequential procedures in screening is discussed by Davies [2] and Finney [3]. Testing in a sequential procedure is carried out in small but essentially similar steps. For many of the materials tested it will become clear quite early whether or not they meet standards—testing is continued only for materials for which this issue is still in doubt. The average amount of testing per material is reduced.

It is the assembly line nature of a screening program which makes such sequential procedures suitable. With a specified capacity for handling test animals, reducing the average testing per material increases the output of tested materials. The time required to get final results for a particular material following initiation

of testing may be increased, but the sequential procedure makes it possible to get around to testing it sooner. Ordinarily, in the research laboratory, it is desirable to obtain results promptly and the possible loss in time from sequential procedures may more than offset savings in experiment size.

The sequential procedure, essentially, sorts out materials into those passing or not passing the test criterion but does not necessarily rank those passing according to their degree of promise. As indicated above, such ranking may be necessary in order to provide a basis for selection of the most promising materials for secondary testing. This can be accomplished by modifying the sequential procedure so that testing through all steps is continued for materials showing promise; an overall score for promising materials is thus obtained on the basis of which they can be ranked. Since relatively few materials will pass the test, the complete regimen of testing will add little to the overall amount of testing. Where too many materials pass it is clear that the passing criterion should be made more stringent.

4.2. *Nature of response.* The progressive selection of materials through a screening program can be in some instances a somewhat simple affair. In the selection of drugs for their therapeutic effects no single parameter attaches to each drug on the basis of which the drugs may be graded and to which all the screening tests in the program will be sensitive. If this were the case a screening program would consist of a series of tests each providing progressive selection for the parameter of interest. In a therapeutic screening program, and also for other special programs, the parameter to which the screen is sensitive may change from one stage of screening to the next, being constant only between the substages of a sequential test. The success of the screening program depends on the correlation of the parameter values for a compound between the successive stages.

This is illustrated by considering a three-stage screening program for obtaining drugs effective against a particular infection. The stages of screening could consist of an in vitro test followed by an in vivo test against an experimental infection and ultimately by testing against clinical infections in man. Similarly, the stages in an anticancer screen could consist of tests of ability to inhibit experimental cancer growth; ability to increase the survival time of animals with experimental tumors; ability to cure animals with experimental tumors; and eventually ability to cure clinical cancers. The parameters measured by the successive tests above are concerned with different though probably correlated abilities.

Another aspect of screening for therapeutic agents is the consideration that must be given to the toxic effects of the drugs tested. Certainly some safety testing will be made prior to clinical use, but to what extent should host toxicity play a role in evaluating drug effectiveness. It may be naive to expect that a screening program will produce effective compounds with little or no toxic effects. Rather the program may produce compounds with important therapeutic effects

but for which some developmental program will still be necessary for minimizing toxic effects without important loss of therapeutic effect.

It does not follow from this that a screening program can ignore the problem of host toxicity entirely. If host toxicity were ignored the problem of finding effective anticancer agents would be reduced to one of obtaining satisfactory detoxifying agents, if possible, for presently known agents. There are any number of agents which will at sufficiently high levels completely inhibit experimental tumor growth. At these levels, unfortunately, they are invariably lethal. An effective therapeutic program can afford to leave for the subsequent developmental program the burden of finding remedial measures for only the less onerous forms of toxicity.

The fact that a therapeutic screening program must take cognizance of drug toxicity introduces certain complications. It is no longer appropriate simply to compare, say, the antineoplastic effect of one milligram of drug *A* with that of one milligram of drug *B*. The comparison must now be made in the light of their toxic effects for the host. To the extent that in any therapeutic program material costs may be considered to be essentially trivial, the comparison of beneficial drug effects should be made only in the light of their toxic effects. This concept was applied by Goldin et al. in two different ways in their investigations into the chemotherapy of experimental mouse leukemia L-1210. In [4]–[7], the experiments performed by Goldin et al. permitted them to make a temporal separation of deaths attributable to drug toxicity and those attributable to progressive tumor growth. This separation in turn permitted evaluating a treatment in terms of its antitumor effect, as measured by the average survival time of mice not succumbing to toxicity, at a specified cost in lethal toxicity. This method, for instance, demonstrated the superior antileukemic specificity of action of amethopterin over aminopterin despite the greater amounts of amethopterin required to achieve a specified antitumor effect [6]. The other procedure employed by Goldin et al. was based on considering overall survival time as an appropriate reflection of both the toxic effects and the antitumor effects of treatment. The measure of performance of an agent on a schedule of treatment was then the maximum average or median survival time it could yield—at doses less than optimal mice died early due to more rapid tumor growth and at doses above the optimal they succumbed early to drug toxicity. Application of this procedure confirmed the superiority of amethopterin over aminopterin [8], [9] and resulted in the demonstration of still higher antineoplastic efficacy for various halogens of amethopterin [10].

It should be noted that neither of the procedures employed by Goldin et al. gives any weight to the factor of material drug requirements in comparing or evaluating alternative therapies. It has been employed by them to evaluate other factors such as altering schedules of treatment [11] or employing drugs in combination [7], [12].

The procedures employed by Goldin et al. for taking into account drug toxicity

illustrate the necessity for obtaining the effects of an agent over a series of dose levels in order to evaluate it properly. Such a titration though not necessarily infeasible could add substantially to the burden of a testing program. A somewhat ingenious procedure for overcoming the necessity for full titration is currently employed by the Cancer Chemotherapy National Service Center (CCNSC) in connection with its primary screening program for anticancer agents [13]. A drug is first tested at what is considered a physically maximum dose. The results obtained are accepted unless excessive toxicity is noted—in such case testing is repeated at a lower dose level, the extent of lowering being related to the degree of excess toxicity observed. If necessary the level of treatment may be reduced still further until finally a result is obtained at a relatively nontoxic level. In effect, the endpoint for a drug is its practical maximum effect, the effect noted at the lower of its maximally tolerated dose or the specified physically maximum dose. The results obtained with such a system depend on the reasonableness of the prescribed physically maximum dose, for which it may be difficult to obtain universal agreement. Nevertheless the procedure does point the way towards means for overcoming the necessity for thorough titration.

5. Parallel and alternative screening procedures

Of the 74 biological systems considered for cancer chemotherapy screening in the Gellhorn and Hirschberg report [1], 15 are based on the inhibition of experimental neoplasms. The results obtained indicated the advisability of using only tumor systems for screening. In addition, the recommendation was made that a spectrum of tumors be employed for screening purposes since not all the useful agents tested were selected by any single tumor system.

It is the paralleling of effort implied by the use of a spectrum of tumors which will be under discussion here. To what extent is it worthwhile to test materials simultaneously in what are essentially several different screening systems, one for each tumor in the spectrum. A related question is the advisability of modifying the method of application of a material in a single tumor system; if a drug is not satisfactory when administered daily, perhaps it will prove so when administered on alternate days. When are such serial modifications of a screen suitable?

Perusal of the results in the Gellhorn and Hirschberg report [1] reveals a fairly high degree of association between the results obtained for a compound with the various tumor systems. If a compound is effective in one system its probability of proving effective in another is increased; conversely, an ineffective compound in one system is more likely to prove ineffective in another. The association of antitumor effects in the various systems shows up despite the fact that they were obtained by a variety of investigators each employing his own techniques and criteria for evaluating effects. With more uniform procedures the degree of association would probably have been greater.

The association between results for the various tumor systems has certain

implications with respect to the conduct of screening programs in parallel. It suggests that it would be wasteful of effort to test a drug with a second tumor system if it has proven negative with the first; a completely untried drug would have better chance of proving effective than one already found to be ineffective with the initial tumor system. Given an unlimited number of compounds to test the most economical way of expending testing effort would require using a single tumor system, presumably the one most fruitful per unit of effort.

The use of a single system of testing however becomes inappropriate when one considers that certain costs are involved also in obtaining a drug for testing. Where a drug has to be newly synthesized the relative costs of testing may well be trivial. In such case, considering overall effort, more thorough testing may be justified; it is not uneconomic not to test a compound relatively thoroughly just because it has proven ineffective on some particular test. An extreme example of this would be the case where one has only a limited number of compounds that can be tested; effectively, the cost of obtaining another compound is infinite. In this case one would be interested in testing each compound in every possible reasonable way, and effectively the screening procedure and the laboratory development procedure become identical.

It is under the conditions indicated above, where materials for testing are hard to come by, that the Gellhorn-Hirschberg recommendation for employing a spectrum of tumors in testing is justified. The fear of overlooking a promising compound through incomplete testing is warranted when the production of the compound alone represents a sizeable investment of effort and it is no doubt with this situation in mind that the recommendation for spectrum testing was made.

The various materials entering a testing program, however, represent a wide range in productive effort. The bulk of materials submitted for testing require no special effort to produce—many of the newer materials submitted may even be obtained fairly routinely. The extent of testing for a material could reasonably be related to its class o f difficulty of production; routine materials could be tested with a single screen, more costly produced materials over some moderate screening spectrum, while the most costly materials would receive thorough testing. The justification for more thorough testing of more expensive materials must also depend, however, on the expectation that they are more likely to be effective in one of the screens—without such expectation one might just as well restrict the testing programs to compounds not presenting any special procurement problems. There is also the logical difficulty of how to handle compounds which, though readily accessible today, represented a production effort some years ago; essentially these are no different from the expensive compounds of today and should be treated in the same way.

There is evidently some need for establishing classifications of materials which will determine how thoroughly they are to be tested. The classifications would ensure adequate amounts of testing for the more interesting materials processed, but would avoid swamping of the program by having the bulk of materials

tested in only one or two screens. If a material proves positive for one of the screens in which it is tested it could be advisable to test it in the balance of the program.

Minor modifications in a screening procedure may be looked at from the same point of view. However, the correlation of drug effects in variations such as changes in schedule or duration of treatment is likely to be so great as to make it unworthwhile to employ more than one scheme in routine testing. Variations in the testing procedure could appropriately be applied as part of the developmental program for drugs which have proven effective. Continued research to make improvements in the routine testing scheme would also be warranted.

Alternative testing procedures could be geared towards other endpoints of effect. The present antineoplastic screening tests are directed towards finding compounds which are therapeutic as indicated by their inhibitory effects on definitely present tumor material. One could also inquire as to their preventive effects; can drug treatment prevent the appearance of spontaneous or induced tumors? If a simple enough assay system exists for evaluating such possibilities, a program emphasizing these aspects could be initiated. In the absence of an assay amenable to widescale application it might still be worthwhile to make the evaluation of such preventive effects part of the developmental program for the antineoplastic compounds found. Though some negative compounds could also be tested, such a program, by and large, would miss compounds for which the preventive effects were not associated with good inhibitory effects.

Other justifications for the use of a variety of screening procedures may be remarked on here. To the extent, for instance, that human cancer may represent a range of disease entities rather than a single entity, it cannot be anticipated that a single experimental tumor will prove selective of effective agents for all forms of the disease. Use of a single screen even where one is dealing with a single disease may be unwise since there is the risk that the particular screen employed may not correlate with the human disease situation. Use of a range of screens provides a hedge against this possibility.

6. An application to cancer chemotheraphy screening

Mention was made above of the procedures employed by Goldin and his associates for taking cognizance of toxicity in evaluating the antileukemic action of drugs. One of the procedures employs the average or median survival time of treated tumor-bearing mice, irrespective of cause of death, as a suitable criterion for taking account of both the antitumor effects and the toxic effects of treatment. The question of level of treatment is disposed of by requiring that the survival time be obtained at or near the most effective dose. Let us consider the suitableness of such an assay in screening.

To the extent that the objective of therapy is to cure, or at least to obtain extensive prolongations in the lives of patients, such a test comes closer to re-

sembling clinical treatment than does a test for tumor inhibition alone. We should expect, accordingly, that the results obtained with such a testing procedure will tend to correlate more highly with clinical results than would a test of tumor inhibitory capacity alone. If the tumor inhibitory capacity of a drug is not reflected in an extension in survival time, it is to all purposes ineffective. From this point of view a test in the nature of that used by Goldin should provide a suitable secondary screening procedure for compounds with demonstrated tumor inhibitory capacity.

Goldin's procedure has been applied successfully employing both early and advanced leukemia. By initiating treatment early with amethopterin, say about a week before the time of death of untreated controls, it was possible to increase the median survival time by two months and to achieve a limited number of cures [8]. Even when treatment was initiated quite late, only one to three days prior to normal time of death, the survival time was extended for two to three weeks [11]. More recently some agents were found capable of producing extremely long survival times and cures even with late initiation of treatment [10].

The demonstration that amethopterin and certain other agents can be employed with some degree of success against advanced experimental leukemia permits us to raise our standards in judging drug effectiveness and we may require comparable performance from other agents prior to their clinical application. Other advantages attach to employing survival time obtained in treating advanced tumors as a secondary screen.

1) Testing would ordinarily be of relatively short duration as compared with treating early tumors—it would be initiated later and discontinued sooner.

2) The advanced tumor is relatively insensitive to nonspecific drug effects. Moderate extensions in survival time accompanied by sharp reductions in tumor size have been achieved by initiating dietary intake reductions early, but not late, in the course of the leukemic process.

3) The testing system would still be appropriate when the required criterion of performance is extensive prolongation in survival time or even cure. In other words, the test can provide a good degree of differentiation between two apparently effective compounds. Two compounds each of which provide apparently 100% inhibition in tumor growth in a primary screen may prove to be distinctly different in effectiveness with a secondary screen of the type suggested. This advantage applies also for treatment with the early tumor.

At this point a question intrudes itself: Can screening with the advanced tumor be used to displace a primary screen in which the degree of inhibition of tumor growth is the endpoint? Continuation of such a primary screening program would have to find its justification in one of the following bases.

1) We are still principally interested in finding compounds with high tumor inhibitory capacity even if it is not reflected in an ability to prolong the survival time of tumor-bearing mice.

2) The test of tumor inhibitory capacity is so inexpensive to apply as com-

pared with the expense of testing capacity for prolonging survival time that it is economical to apply the former test as a primary screen, the survival time test as a secondary screen.

Whether or not the first point above provides a justification for retaining the tumor inhibitory test would depend on the objectives of the screening program. When so justified there can be no disagreement. Let us consider however if the second justification might apply.

The tumor inhibitory test ordinarily requires holding and treating animals on each of seven days, then sacrificing them in order to excise thd weigh their tumors. With the survival time screen, animals are held for a period prior to initiation of treatment, but once treatment is begun, it may be continued for an extended period; all that need be noted is the day of death of animals. Note, however, that the vast bulk of materials tested will probably be completely ineffective in such a screen; if treatment is initiated two days before death would otherwise occur and is maintained on an alternate day schedule, most of the animals tested will require only one or two treatments. It would be only with the rare effective materials that extended treatment would be required. The potential of the survival time screen for economizing on effort is apparent, and from this point of view it may make the tumor inhibitory test unnecessary.

Skipping the tumor inhibitory test may provide some additional advantages by avoiding the losses which would have resulted from the imperfect correlation between tumor inhibitory test results and the survival time test results. The additional cost from testing with both systems because a compound has a nonspecific tumor inhibitory effect is avoided. Also, we are now in a position to uncover compounds the ability of which to increase survival time is not associated with a concomitant ability to inhibit tumor growth, though such compounds may be extremely rare. Finally, we avoid the problem of having some percentage of compounds which are actually effective failing to pass the inhibition test and so not getting to the survival time test.

Use of the survival time test in a screening program to take the place of both a primary and secondary screen may still require a degree of implementing. Some device to achieve the effect of complete titration while avoiding its cost, such as the one employed by CCNSC and described above, may be desirable. And while Dr. Goldin's group has applied the survival time test to leukemia L-1210, a developmental program will be necessary to facilitate its application to other tumor systems. The CCNSC does in fact now employ the survival time test, not as an initial screen on all compounds, but in special studies of selected agents [14].

What has been demonstrated in the foregoing section is the abrupt change which may occur in a screening program when the uncovering of an effective compound leads to a change in the required level of performance. Use of the tumor inhibition test was justified so long as we did not know of the existence of especially effective compounds. Its use could have continued to be justified if its cost were low enough. But in the present instance the cost of testing for

an especially good effect (extension in survival time for mice with advanced leukemia) was even less. The tumor inhibition test may be of no further value to the screening program.

7. Summary

In the foregoing the problem of the efficient design of a screening program was discussed and at points, the design of an anticancer screening program was employed for illustrative purposes.

The chief method for achieving program efficiency described was the employment of a series of testing stages each of successively higher rigor and requiring successively greater testing effort. It was emphasized that this method was not always appropriate—the more rigorous test might be simple to perform and could thus be used as a primary screening test. As more effective materials are uncovered in the course of a screening program the objective of the program may change to a point where it is no longer worthwhile to uncover materials capable of passing only a moderately severe screening test. At this point, provided a simple rigorous test exists, the design of the screening program may require alteration.

Another procedure for introducing design efficiency is through the employment of sequential substages. Each stage of testing is divided into steps and, ordinarily, a compound is tested through the full series of steps only if an earlier decision cannot be made as to whether or not the compound is satisfactory. It was pointed out that for purposes of selection of the most promising compounds for further testing, a sequential test should be such that promising compounds are fully tested. The results of complete testing can then be used to order compounds according to their degree of promise.

The employment of a set of alternative screens, as for instance screens based on a variety of experimental tumors in an anticancer screening program, was discussed in relation to the cost of compound procurement and the degree of correlation of results obtained with the various tests. This brought out the need for classifying drugs according to the extent of testing which they should receive. The developmental program for compounds found positive in a screen might include testing with the alternative screens and with modifications of the test already employed. It may also include testing with procedures directed to come other endpoint of effect but for which routine screening is infeasible.

Screening programs for therapeutic agents must take into account the toxic effects of the agents tested. Some aspects of toxicity may perhaps be disposed of in a developmental program for compounds found positive, but to a large extent the problem of drug toxicity must be faced up to in the course of the screening program itself. This important question underlay much of the discussion; methods for handling it in special instances were given.

Alongside of any good screening program there must be an effective program of research and development. If this point has not been made clearly enough

above it should be re-emphasized here. The research program should lead to understanding more about the test system employed and may point the way to improving or displacing it. The function of the developmental program is to learn the properties of and how more correctly to employ drugs found effective in a screening program.

REFERENCES

[1] A. GELLHORN and E. HIRSCHBERG (eds.), "Investigation of diverse systems for cancer chemotherapy screening," *Cancer Res.*, Supp. No. 3 (1955), 125 pp.

[2] O. L. DAVIES, "The design of screening tests in the pharmaceutical industry," *Bull. Inst. Internat. Statist.*, Vol. 36 (1958), pp. 226–241.

[3] D. J. FINNEY, "Statistical problems of plant selection," *Bull. Inst. Internat. Statist.*, Vol. 36 (1958), pp. 242–268.

[4] A. GOLDIN, N. MANTEL, S. W. GREENHOUSE, J. M. VENDITTI, and S. R. HUMPHREYS, "Estimation of the antileukemic potency of the antimetabolite aminopterin, administered alone and in combination with citrovorum factor or folic acid," *Cancer Res.*, Vol. 13 (1953), pp. 843–850.

[5] ———, "Effect of delayed administration of citrovorum factor on the antileukemic effectiveness of aminopterin in mice," *Cancer Res.*, Vol. 14 (1954), pp. 43–48.

[6] A. GOLDIN, J. M. VENDITTI, S. R. HUMPHREYS, D. DENNIS, N. MANTEL, and S. W. GREENHOUSE, "A quantitative comparison of the antileukemic effectiveness of two folic acid antagonists in mice," *J. Nat. Cancer Inst.*, Vol. 15 (1955), pp. 1657–1664.

[7] ———, "An investigation into the synergistic antileukemic action of amethopterin and 6-mercaptopurine in mice," *J. Nat. Cancer Inst.*, Vol. 16 (1955), pp. 129–138.

[8] A. GOLDIN, J. M. VENDITTI, S. R. HUMPHREYS, D. DENNIS, and N. MANTEL, "Studies on the management of mouse leukemia (L-1210) with antagonists of folic acid," *Cancer Res.*, Vol. 15 (1955), pp. 742–747.

[9] A. GOLDIN, J. M. VENDITTI, S. R. HUMPHREYS, and N. MANTEL, "Quantitative evaluation of chemotherapeutic agents against advanced leukemia in mice," *J. Nat. Cancer Inst.*, Vol. 21 (1958), pp. 495–511.

[10] ———, "Prolongation of the lifespan of mice with advanced leukemia (L-1210) by treatment with halogenated derivatives of amethopterin," *J. Nat. Cancer Inst.*, Vol. 22 (1959), pp. 811–823.

[11] ———, "Modification of treatment schedules in the management of advanced mouse leukemia with amethopterin," *J. Nat. Cancer Inst.*, Vol. 17 (1956), pp. 203–212.

[12] A. GOLDIN and N. MANTEL, "The employment of combinations of drugs in the chemotherapy of neoplasia: a review," *Cancer Res.*, Vol. 17 (1957), pp. 635–654.

[13] Cancer Chemotherapy National Service Center Specifications for Screening Chemical Agents and Natural Products against Animal Tumors, Cancer Chemotherapy Reports, Vol. 1 (1959), pp. 42–62.

[14] A. GOLDIN, J. M. VENDITTI, I. KLINE, and N. MANTEL, "Evaluation of antileukemic agents employing advanced leukemia L-1210, in mice. I," *Cancer Res.*, Vol. 19 (1959), pp. 429–466.