The Evaluation of Prison Treatment and Preventive Detention Programs: Some Problems Faced by the Patuxent Institution

NATHAN T. SIDLEY, M.D.*

I. Basic Logic of Evaluation

Recent years have seen many programs devoted to "treating" imprisoned criminal offenders, programs which have elicited both enthusiastic espousal and vigorous criticism from interested parties. Legal and political controversy have also resulted from these efforts.

In this article an attempt is made to try to help clarify the situation by considering the very fundamental logical ideas involved both in determining what effects of a treatment program might be, and in considering whether or not to implement such a program. Logic alone cannot, of course, make the critical determinations; however, without a common framework of logic, ultimate agreement and reasonable resolution of conflict among disputants are all but impossible, and a rational approach to knowledge and decisions cannot be evolved.

The logic of treatment evaluation will be presented first; then Patuxent Institution, a widely-known institution with a treatment program for offenders, will be examined in the light of those principles. The keys to this presentation are, first, the conceptualizing of the logic of the "Ideal Case," that is, a situation postulated to possess certain properties, and then, second, the comparison of that situation with more practical ones in the real world.

The Logic of Treatment in the Ideal Situation

The basic unit in the overall logic of treatment is the "case." Each case is thought of as having a "diagnosis." To all cases within any given diagnostic class, a specific "treatment" is applied by the individual in a position to make and implement decisions about the case. All cases of a given diagnosis who undergo a specific treatment end up, after the lapse of a period of time, with some specific "outcome." Thus, *in the ideal situation* diagnosis leads to treatment, which in turn leads to outcome. *All members of* any *diagnostic class* are *homogeneous; all separate treatments* (within the specific class of treatments given to each member of the diagnostic class) are *identical* and the *outcome* classes are also homogeneous. Thus, in effect, *diagnosis-treatment-outcome* combined categories are the result when the individuals in each of the component classes are uniform or homogeneous.

Diagnosis, treatment, and outcome can be considered separately and are usually thought of that way. Each represents a separate classification of a pattern either of observations or of operations performed by someone. Such classifications can be made in many ways, as long as they are made consistently. However, for purposes of rational decision-making about treatment (including treatment of offenders in prison), these classifications are valuable only insofar as they lead to homogeneous *total classes*. In

* Director, Court Clinic, Fourth District Court of Eastern Middlesex, Woburn, Mass.

essence a case manager must be able to predict that if he treats the case in front of him in a certain way, a certain outcome will occur. If treating cases of the same diagnosis with the same treatment leads to different outcomes, the treater's rational decision-making is compromised to the degree that those outcomes lack homogeneity. Indeed, if an outcome is unpredictable, there is no rational decision-making at all, but merely a random, a magical, or a stereotyped assignment of treatment to a case at hand.

Medical men have become accustomed to thinking of diagnostic classifications as "diseases," as if there were such a phenomenon as a disease. It is evident on reflection that whatever basis a diagnostic classification in medicine may have, be it similarity of symptoms, findings, pathological anatomy, or other, the classification functions so as to achieve homogeneity of cases in a diagnostic class with respect to outcome under different conditions of treatment. (Thus the "natural history" of a "disease" is the expected course of that disease when there is no effective treatment.)

If, for example, there were a "disease" of a particular pattern of symptoms, etc., such that under any condition of treatment some got well and some didn't, and if it were discovered that there existed some characteristic, the presence of which was associated with getting well and the absence of which was associated with not getting well, two diagnostic classifications would necessarily result, just as different diagnostic classifications would emerge from a single one if a particular agent, such as a virus, were associated with some of the cases within the larger diagnostic group but not with others. The search for homogeneous D-T-O categories occurs in any treatment-oriented art, whether it be medicine, engineering, or brewing beer.

Formation of Categories

It would surely be desirable if somehow homogeneous D-T-O categories presented themselves to mankind. But they do not; such categories must be found. And the only way to find them is by laborious trial and error. That is, every possible combination of diagnostic observations must be tried with every possible treatment modality combination, and each separate diagnostic-treatment combination must be evaluated according to all possible outcome-evaluation combinations.

For example, suppose 5 treatments exist, and suppose individuals presented for treatment can be diagnostically evaluated according to six characteristics, each of which has only two values, present or absent. There are 2^6 or 64 possible diagnostic classes. If each treatment were applied to each diagnostic class, there would be 5 treatments $\times$ 64 diagnostic classes, or 320 diagnostic-treatment classes. Let us assume 2 possible outcomes. There then occurs the *possibility* of 320 diagnostic-treatment classes $\times$ 2 outcomes, or 640 D-T-O classes. (The situation now under discussion is that of *forming* the appropriate classes, when no such homogeneity is available in advance and each diagnostic-treatment combination is likely to lead to several outcome possibilities. Treatment *decision-making* in the ideal situation presupposes that homogeneous D-T-O categories already exist.) Assuming that 6 individuals are necessary to fill out a class for some kind of evaluative assurance, 640 D-T-O categories $\times$ 6 individuals per D-T-O category, or a minimum of 3840 individuals, would be required to establish D-T-O classes in which a decision-maker would still have only minimal assurance because of the small numbers in each class.

Notice that although in the hypothetical example there are 64 possible diagnostic patterns, if any of them yields homogeneous D-T-O categories, further analysis may be indicated regarding the possibility of higher-order conjoint or disjoint categories especially with respect to any given treatment. For example, suppose that eight diagnostic patterns yielded homogeneous D-T-O categories with respect to Treatment #1. That is, if any individual with any of the eight diagnostic patterns were treated with Treatment #1, the outcome would be the same as for other individuals with any of those

patterns treated with Treatment #1. Suppose these were the eight patterns with respect to the six diagnostic parameters:

Pattern Number	*Parameter* A	B	C	D	E	F	*Outcome with Treatment #1*
1	+	+	+	+	+	+	+
2	+	+	+	+	+	−	+
3	+	+	+	+	−	+	+
4	+	+	+	+	−	−	+
5	+	+	+	−	+	+	+
6	+	+	+	−	−	+	+
7	+	+	+	−	+	−	+
8	+	+	+	−	−	−	+

It is apparent from inspection of the Table that the results of Diagnostic Parameters D, E, and F are irrelevant to the D-T-O patterns but that the presence (+) of Parameters A, B, C leads to a homogeneous D-T-O category when Treatment #1 is applied. The A+, B+, C+ Diagnostic category is a higher-order combined category than the first-order categories which take all possible observational parameters into account. That then represents a *conjoint* higher-order diagnostic category.

Similarly there might be a *disjoint* higher-order diagnostic category of such nature, say, that if either A and B are +, or if C and D are +, or if E and F are +, the categories are homogeneous with respect to outcome if a given treatment is applied.

Formation of such higher-order conjoint or disjoint categories may be useful in a search for homogenous categories. It is apparent that they will not be formulated unless they are sought. The process of seeking and testing them appropriately, however, is subject to the same limitations as exist for lower-order, more basic D-T-O categories.

In the medical field, doctors no longer go through such an exhausting exhaustive procedure to formulate D-T-O categories. They take heuristic short-cuts that eliminate many steps. They perform only a small number of the total possible diagnostic evaluation procedures; they subject cases to only a small number of possible treatments, and they usually evaluate according to a small number of criteria.

In such short-cut situations doctors often use a method which is based on the notion of determining whether two phenomena are different from each other. The notion is that of the controlled observation. In brief the logic of the controlled observation is this: If two phenomena are identical and are treated identically with respect to everything except a single parameter, any differences in outcome are results of the application of that parameter of treatment. In practice the difficulty is obtaining identical starting conditions and single treatment parameters, though for some physical and chemical phenomena, one can obtain practical identity.

That difficulty leads to a requirement in behavior studies for statistical controls. A statistical control involves the selection for observation of groups of individuals who are at least as similar as possible if not identical. When a behavioral experiment is conducted, statistical control is provided by randomly dividing subjects into experimental and control groups. It is assumed that the groups are identical at the start even if the individuals aren't. The experiment is conducted as if the individuals are identical. Any differences between the experimental and control groups are regarded as the results either of chance factors or of the application of the experimental procedure to the experimental group. If the differences are not great (i.e. less than a certain level of significance), the experimenter concludes that chance factors accounted for the differences. If the differences are great enough so that it is sufficiently unlikely that chance

factors alone were responsible, the conclusion is that the experimental procedure was the important factor leading to the differences.

Post facto studies observing differences between groups are based upon similar logic. Both groups are considered identical in principle, except with one significant difference which corresponds to the application of an experimental procedure. E.g., members of paired contrasting comparison groups, such as blacks and whites, males and females, smokers vs. non-smokers, etc., are all considered to be equivalent as people, and as people they are divided into two groups according to the comparison characteristic evaluated. Except for that characteristic they are expected to be identical. If secondary observed differences exist between the groups, the conclusion is that the primary observed difference is the important factor related to the secondary difference. E.g., if many of a group of smokers all have lung cancer and few of a group of non-smokers have lung cancer, the conclusion is that the smoking was the important factor related to the cancer.

The real problem in *post facto* studies lies in the assumption that the populations observed are identical in every respect except the primary factor observed. For example, if one were to observe a difference in cancer incidence between a group of smokers and of non-smokers, and if half the smokers also worked in asbestos factories, the conclusion that the primary factor of smoking led to the increased secondary factor of cancer incidence would be highly questionable.

From the standpoint of logic, any lack of identity between two such groups casts into question the conclusion that the primary factor is related to the secondary factor. Of course, it does not prove that the conclusion is not true. (Many propositions which are logically uncertain are, nevertheless, true.) It merely indicates that the evidence does not prove the association.

Whenever statistical control exists, careful examination of the assumption of identity between the groups may indicate lack of identity, and that confounds the conclusions. The greater the seeming relevance of the lack of identity to the outcome observations, the greater the questionability of the conclusions. Although lack of identity can be a problem in such experiments, it can often be considerably controlled in advance of the experiment. In *post facto* studies of different groups in nature, there is no way of controlling lack of identity. If serious confounding problems occur, one must look further to other groups to try to evaluate the potential significance of the confounding lacks of identity.

The preceding discussion has mostly been aimed at the formation of diagnostic categories; the formation of treatment categories represents a similar endeavor. Each potential treatment category represents a combination of treatment operations which can be deployed under the control of the case-manager. Notice, however, that the definition of treatment classifications differs from the definitions of diagnostic and outcome classifications. The latter two represent observations made upon the case, while the treatment classification represents observations upon the case-manager.

A case occurs over a time span. Diagnosis represents the classifications of conditions of the case at the start of case-management, while outcome represents conditions of the case at the end of a period of time of case management. Treatment classifications represent conditions of (or applied to) the case during the period of case-management. It can be seen that conditions imposed upon the case by the manager represent only a portion of all the conditions of the case during the treatment period. A more comprehensive aim of the search for homogeneous categories should be Before-Intermediate-After categories rather than D-T-O categories. Diagnosis-Intermediate-Outcome categories also represent a larger group of categories than D-T-O categories.

D-T-O categories are used in case-management because treatment categories represent the operations which are under the control of the case manager. The others are con-

sidered irrelevant. However, some diagnostic-intermediate-outcome categories may be highly homogeneous (individuals with stomach cancer who leap from 15th story windows form a highly homogeneous category), and if members of those combined categories are not kept separate from others in similar diagnostic categories, case-managers who have to make use of the appropriate categorical information may be misled. It makes a difference whether a patient, within a year after a surgical operation, dies from his cancer, from a post-operative complication, or from an auto accident.

Statistical Categories

Experience does not reveal homogeneous D-T-O cases such as postulated in the ideal situation. Not only is there ambiguity with respect to aspects of diagnostic observations, treatment operations and outcome evaluation, but also the fact alone that multiple influences may affect outcome precludes rigid control and precludes D-T-O homogeneity.

Thus actual working categories are "statistical" D-T-O categories. For example, in dealing with offenders one might be able to say something like "75% of individuals in Diagnostic Class I with no treatment recidivated, while 25% did not recidivate." Or one might be able to say "30% of individuals in Diagnostic Class 2 who received no treatment were arrested for felonies involving physical violence and injury to innocent victims within a period of two years of being discharged from prison and being exposed to the community." Notice that the criteria in the second example are more specific than in the first. Obviously more specific categories require more observational subjects and investigative effort to obtain. The principle that the experience of the past is the only guide to the future leads to the use of figures derived from observation as the basis of an "expectation," or prediction for future observations. The usefulness of offender statistics as a basis for future expectations is likely to be somewhat proportional to the similarity between groups observed in the past and groups to be observed in the future. A high degree of similarity may not always exist.

The above discussion has centered on a strong and simple logical form, the ideal case of D-T-O categories and a weaker practical case of statistical D-T-O categories. A related functional model is used in parole prediction, namely the regression model. In the regression model each individual is evaluated on many characteristics and is given a final composite or resultant score. On the basis of that score, a prediction is made based on correlation of scores with some outcome criterion, e.g., recidivism.

The general approach to formulation of such recidivism prediction scores is this. Just as in the statistical D-T-O effort, a group of prison releases is evaluated on a set of observable characteristics. After a period of time, outcome criterion data, recidivism vs. non-recidivism, are collected. Each characteristic is correlated with the outcome criterion. Those characteristics which have a sufficient correlation are combined, usually in a simple sum (though more complex combination methods can be used), to form an index score. The index score is in turn correlated with the criterion. For example, number of previous offenses, age at first offense, grade attained in school, previous imprisonments, etc. might be characteristics evaluated in each potential releasee. A score of 1 or 0 might be assigned to each item depending on the situation in each case. The sum of the scores of each characteristic evaluated would be the index score. An equation called a "regression equation," relating index score to likelihood of recidivism, can be developed. On the basis of the regression equation a prediction can be formulated for each case.

Furthermore, an estimate can be made as to the likelihood of error for each score. E.g., one might be able to say that a score of 80 or more would be expected to cover 70% of recidivists and only 30% of non-recidivists; a score of 85 or more would cover 65% of recidivists and only 15% of non-recidivists, while a score of 90 or more would cover 60% of recidivists and only 2% of non-recidivists.

Decision-Making

The logic of decision-making requires one other consideration, that of value. For example, for any given diagnostic category, there are many potential treatments, thus several D-T-O categories. Each different treatment is likely to bring about different effects or a different outcome with respect to that diagnostic category. The principle of decision-making is that for any diagnostic category, the decision-maker should select the treatment in the D-T-O category which has the relatively most favorable outcome in relation to the relatively least unfavorable outcome.

The basic problem is comparing outcomes so that one can tell not only 1) how favorable and unfavorable any outcome is, but also 2) the extent to which one is favorable and unfavorable compared to others. Some kind of scales of comparison are required, scales that have strong mathematical properties so that the values can be calculated in a manner analogous to money calculations.

The required means of placing values objectively on different potential outcomes (or on different D-T-O categories) are in general not available. Nevertheless, in practical situations a kind of subjective evaluation and computation does take place and must take place, as, for example, in medicine, whenever a patient undergoes treatment, especially when it is on a completely voluntary basis. The problem, of course, is that there is no uniform comparison scale by means of which one can come up with a "net utility score" for each alternative treatment possibility within a specific diagnostic class. Who can balance a 50% chance of cure with a 20% chance of losing a leg against a 40% chance of cure with a 10% chance of losing an arm? Yet despite considerable imprecision and absence of facts in the medical field, patients decide every day what kinds of treatments to accept for their ailments after considering some kind of informed calculation of the potential risks and benefits. Doctors also decide which treatments to recommend for which ailments, despite incomplete information. The differences between treatments, however, are usually gross ones (there is no real argument between the use of digitalis or appendectomy for cases fulfilling diagnostic criteria for "appendicitis"), and most medical recommendation decisions are easily made.

Because there are generally only two alternatives, release or incarceration, decisions about offenders in relation to potential recidivism might seem even easier to make. Actually they are more difficult. The differences between individuals who will recidivate and those who won't are ambiguous, and no gross reliable criteria exist for differentiation of the two classes. Thus all decision-making in this field must be on a statistical basis.

Furthermore, these decisions are not made by the individuals who suffer personally when the statistical decisions work out badly. They are made by legal authorities (who don't suffer directly), and they are public policy decisions. What should be the basis of public policy decisions regarding offenders or potential offenders?

Let us consider a hypothetical example. Assume that it could be shown that members of diagnostic class Alpha are expected to commit a property felony within two years of discharge from prison, such that on the average, for each 100 men released, $10,000 worth of property will be stolen or otherwise destroyed. All individuals in the class are alike (the definition of a diagnostic class). It is assumed that there is no way to improve monitoring of class members in the community so that the loss can't be reduced. How can one determine whether to parole or not to parole an individual in that diagnostic class (really a group of individuals, for the same considerations hold for all in the group) if he is eligible 2 years before his maximum sentence expires?

One might first consider the financial considerations. The total cost of paroling these offenders is the $10,000 in property loss, the costs in inconvenience to the owners of the property, the costs to police departments in seeking the offenders, the court and other judicial processing costs, and the costs of incarceration for those who are again sent to

prison. There are other potential effects from the ordinary risks that an individual will commit crime or otherwise impose some cost on society.

The cost of retaining those same offenders in prison is the actual cost of incarceration, the costs of maintaining the families of the men who are in prison, the costs in inconvenience to those individuals who would not recidivate and are unnecessarily retained in prison, the potential costs of damage to or by the individuals in question that would occur because they were in a prison, and the potential costs of crimes committed by individuals retained longer, upon their eventual release. Assuming no other considerations, the decision regarding parole is easy in principle to make. All one has to do is add up the costs of paroling and the costs of not paroling and choose the alternative that costs least.

Many of the factors noted above are highly subjective, though. How can one assign a dollar value to the inconvenience suffered by a crime victim, be he a victim of either a property crime or a personal crime? How can one assign a dollar value to being confined unnecessarily in prison for a period of two years?

Another intangible that cannot be assessed accurately but which is a consequence of such legal decisions is the impact upon the public. Above and beyond the immediate emotional reactions people have to various decisions, there are two fundamental considerations. On one hand, people feel they have a right to be protected by their government and legal agencies, and they expect to be so protected. Actions which indicate that the protection function is not being fulfilled are quite costly. On the other hand, however, people wish assurance that they will be protected from arbitrary actions of government itself, that is, that a person will not be deprived of liberty or property unless he is actually guilty of a crime, and that the punishment will fit the crime. There is, of course, no solution to the conflicts between these two fundamental considerations, though eternal vigilance can perhaps help to preserve some kind of equilibrium. In any case the dispositions of such cases are likely to have definite, if not determinable, effects on these fundamental considerations of public opinion.

There is another cost, namely the moral one to the decision maker. It is one thing to observe someone suffering because of the actions of another. It is another thing for a person himself to impose suffering on another. The moral cost to a decision-maker of erroneously incarcerating (or keeping incarcerated) a non-recidivist is a high one that also must be included in the computations.

Thus, when prediction of recidivism is made on the basis of regression scores, the decision-maker must determine where the cut-off point should be. All individuals beyond the cut-off point are retained (thus creating an expectancy of a certain range of "false positives"), and all below the cut-off are released (creating an expectancy of a certain range of "false negatives"). In essence the decision-maker must determine the false positives and false negatives at each possible cut-off score, and appraise the "value" at each of those cut-off scores. He then makes his cut-off decision at the point where the expected value is greatest.

It is apparent that the really critical issues in the determination of where to draw the lines, and thus whom to release and whom to retain, are at this time either essentially unknowable or arbitrary. Making rational decisions in this field is therefore laden with caveats, imponderables, and other obstacles which interfere with making good decisions. In fact one of the major problems in legal decision-making in this field is that there is not a sufficient basis in knowledge to permit even the remote possibility of rational decisions. That non-rational decisions are made anyway is well known. Perhaps, however, ultimately the situation and the results can be improved.

Preventive Detention

The whole problem is also well-exemplified in decision-making related to preventive detention laws. These are designed to protect society from individuals who, if they were

not incarcerated, would commit violent and injurious crimes upon innocent victims. In many states these laws result in individuals being declared to be "dangerous," and they are involuntarily sent to penal or quasi-penal institutions for unpredictable and indefinite periods of time. Yet there are nowhere definite before-after studies which have systematically delineated relevant D-T-O categories with respect to dangerous individuals.

How then are determinations of "dangerousness" made when there is no basis in knowledge for doing so? The process is an inferential one, called the "clinical" method. Indeed, the process ought to be regarded as a mysterious one, for by definition it works in mysterious ways.

The basic process of the clinical method is that "experts" are exposed to many facets of a case. They obtain a "case history" from various sources; they "examine" the individual themselves, and they submit the individual to "tests." (The tests themselves may be simple, or they may be "clinically interpreted" and therefore also mysterious.) The experts then, either singly or following discussion at a "case conference," give their conclusions and opinions based upon their "experience" and their "judgment." The process is a mysterious one because there is no analytic method or algorithm which can process and integrate the information about the case in a specific manner. (If there were an algorithm there would be no need for a clinical method.) The information is processed in the experts' hands in a manner that is incapable of precise description or of analysis. Ultimately the conclusion rests upon the authority of the experts who make it.

Although the fact that experts disagree in their opinions is important, it is by no means the basic problem in clinical evaluation by authority. The elemental problem is that by the clinical method of evaluation, every case is unique. There is, by definition, no way in which the expert can know, in the sense of being able to subject to systematic analysis, the reason for his conclusion. He has no way of knowing, if he looks at the facts of the case on two different occasions, whether extraneous considerations have influenced his conclusions on either occasion. He has no way of knowing whether his basic processing may have changed from one occasion to another. He has, in short, *no systematic way of learning from his mistakes* and therefore *no way of improving his performance.* (Indeed, in many of these cases there is no way for him ever to know if he has made a mistake, for the experts' opinions are all too often the deciding factor in a case. If the psychiatrists say "He's dangerous," the court often incarcerates. There is no way to discover if one has made a false positive determination.) Moreover, there is no way for an expert himself, let alone some outsider not privy to the thinking of any expert, to be able to compare one case with another. Because every case is unique, it must be evaluated by the experts in terms of that specific case.

This is not to say that expert decision-making must be mysterious to outsiders, even if the experts present that view. It is possible that experts may come to conclusions in mysterious ways but that outsiders can correlate their conclusions with simple facts. For example, it is entirely possible that even though a panel of experts may study a case for days and thoroughly evaluate it, an outsider, knowing that the case was one in which a young man slipped into a woman's apartment, raped her and beat her bloody, can reliably predict that the experts will declare him to be dangerous—or indeed that the experts would conclude that any multiple offender who has committed a recent severe assault is dangerous. Note, however, that the mere facts, either that experts form mysterious opinions or that those opinions are often predictable, do not indicate at all whether or not the opinions are correct.

Indeed the final problem with clinical decision-making is that it is often frankly wrong. Many studies exist of the relative efficacy of "clinical versus actuarial" decision-making, well reviewed by Sines[1] and Sawyer,[2] and in almost all situations in which the two decision-making methods have been compared, clinicians have been less effective than statistical combinational methods. Unfortunately neither method has as high an accuracy as is desirable; nevertheless, when the two are compared, the clinical approach

appears to be less effective than the statistical. Sawyer's suggestion that "preferences among methods reflect something other than efficacy of prediction"[3] seems particularly apt.

From the standpoint of the logic and the procedure, preventive detention is the same as parole decision-making. That is, a prediction is made as to whether the individual will, if not detained, commit an undesirable act or not. If it is predicted that he will, he is detained. From the legal standpoint the situation is somewhat different, however. It is ordinary to punish or incarcerate a person after he has been convicted of a crime, and he can be retained in prison until the expiration of his sentence. It is extraordinary to detain (and thereby punish) a person who has not been and may never be convicted of a crime. Although it would be expected, *a priori*, that in such an extraordinary situation unusual attention to safeguards, precision of observations and careful documentation of conclusions within the limits of knowledge available would be adhered to, even in this kind of situation the limits of knowledge and rationality are often transcended in decisions which vitally affect people's lives.

Finally it must be pointed out that the logic and methods for evaluation of treatment for individuals who have been detained preventively are the same as for those who have been convicted of a crime beforehand. The problems are a bit greater if there have not been crimes in those individuals' past histories, but usually there is something like an offense in the past history of an individual before he is considered for preventive detention. Obtaining adequate documentation of offensive behaviors is generally easier, however, if a person has gone through a trial than if he has not. A much larger problem may be faced by the psychiatric examiner when he has to acquire his own information to evaluate an offense than when he has the benefit of a prior trial to provide him with many facts.

We shall look to Patuxent, an institution that is in many respects a model institution, for some indications as to how in practice the evaluation criteria in Preventive Detention programs are fulfilled and how the conclusions follow from the information presented and the logic used.

II. The Patuxent Institution

The Defective Delinquent Statute

The Patuxent Institution was opened in 1955 in response to a Maryland statute of 1954[4] which defined the concept of "Defective Delinquent." The purposes of the law were 1) to protect society from individuals who would commit violent crimes, 2) to provide decent and humane conditions for offenders, and 3) to rehabilitate offenders as well as possible through psychiatric treatment. The main mechanism was committing individuals to the Patuxent Institution for an "Indeterminate Sentence." By means of such a sentence, society could be protected and treatment continued as long as necessary. The range of individuals eligible for indeterminate sentence commitment was limited by the restriction of the procedure to those who had been convicted of a crime and upon whom sentence had been passed. The definition of a Defective Delinquent is "an individual who by the demonstration of persistent aggravated antisocial or criminal behavior evidences a propensity towards criminal behavior and who is found to have either such intellectual deficiency or emotional imbalance or both as to clearly demonstrate an actual danger to society so as to require such confinement and treatment when appropriate, as may make it reasonably safe for society to terminate the confinement and treatment."[5]

Because the definition epitomizes the law and the institution, it must be examined in some detail. The definition is multi-faceted and contains both requirements and dispositions. The first requirement, "persistent aggravated antisocial or criminal behavior," is related to the individual's past history. The word "aggravated" applied to criminal (and antisocial) behavior, for practical purposes, implies offenses in which a victim was or might easily have been killed or significantly injured. The word "persistent"

here implies that an individual must have been convicted of at least two aggravated offenses.

(Here perhaps too much is presupposed. Though it is not defined, the word "aggravated" presumably refers to the anticipation of danger of physical injury or death to someone. An ordinary burglary, for example, would not be regarded as an aggravated crime; however, a robbery with a loaded gun, an arson of an occupied dwelling, a forcible rape, or an injurious assault would probably all qualify as aggravated crimes. Also, presumably, the only way an individual conclusively demonstrates "persistence" is by judicial determinations in the form of convictions.)[6]

The second ostensible requirement is that the individual show an intellectual or emotional deficiency. In effect that is no substantive requirement. It can be argued that anyone who commits at least two aggravated crimes is emotionally deficient, and for practical purposes there is no gainsaying that argument. The second requirement is thus contained in the first and is a pseudo-requirement.

Thus the true second requirement is that the individual *"clearly* demonstrate an *actual danger* to society" (italics added). Presumably, as above, an individual is a danger to society when he will commit some criminal act in which someone might easily be significantly injured or killed. The specification that there be a "clear demonstration of actual" danger implies an unambiguous indication of danger as opposed to a probability or other less definite indication. There are many people who are probably dangerous (i.e., some fraction of a group so considered would commit aggravated crimes); however, it is a strongly doubtful assumption that a group of individuals can be discovered whose numbers would virtually *all* commit aggravated crimes *within a reasonable period of time*. Yet the statute presupposes such a high degree of certainty.

The next part of the definition is interesting in that despite its appearance, it does not specify a requirement. It specifies a disposition. If a person is dangerous, he must be confined and treated. That is the disposition mandated by a finding of danger. Though the disposition is mandated only when appropriate, the grounds for appropriateness are not spelled out in the definition. Confinement has in any case been the disposition of individuals diagnosed as defective delinquent. Interestingly the law does not specify what the institutional disposition should be of those individuals for whom the second part of the statutory disposition, i.e., treatment, is inappropriate (i.e., if they refuse to cooperate). To keep them in a treatment facility seems foolish. Yet for those troublesome cases there seems to be no satisfactory alternative disposition to retention in Patuxent, though Section 16d of Article 31B[7] allows the transfer of inmates to prison.

It is finally implied, but not mandated, that confinement and treatment *may* be terminated when such action is reasonably safe for society. This further implication is also interesting, for it suggests a differentiation between a person who is being considered for incarceration when, at the time of consideration, he is not locked up in the institution, and a person so considered if he is already locked up in the institution. If a person is not already in the institution, the burden is on the *government* to prove *clearly* that he is a danger. If a person is already in the institution, the burden appears to be on *him* to prove that society is *reasonably safe* if he is released. It appears to me that the two different criteria are not strictly comparable. The law seems to demand a weaker criterion to retain a man when he is locked up at the time of determination than when he's not locked up at that time. In effect a person must be more dangerous to be committed for the first time than *to be kept* longer when he's there at the time of possible re-commitment.[8]

Ultimately the diagnosis of "Defective Delinquent" (which in effect is the decision to commit the individual indefinitely to what is basically a preventive detention facility) is made judicially by a judge or by a jury. However, a prior diagnosis, actually a prediction, is made by the staff at the Patuxent Institution, and it is only when the

Institutional Staff concludes that the individual should be adjudicated a defective delinquent that the judge or jury gets the opportunity to make a judicial determination.[9]

The Institution

The Institutional prediction is based on a "clinical" appraisal performed during a period in which the individual evaluated resides at the Institution.[10] The diagnostic study attempts to take into account any "relevant" past background or present factors which might bear on the prediction of aggravated criminal recidivism. As noted above, the statute does not specify what the probability of aggravated recidivism[11] (i.e. "clear demonstration" of "actual danger") should be before the examinee is diagnosed a Defective Delinquent. It is obvious that if the requisite probability is .98, there will be a far smaller number of individuals in the Defective Delinquent Category than if that probability were only .75. Perhaps the legislature was wise in not specifying that, perhaps not; however, without such a specification, there is room for interpretation as to what the probability should be, though the words "clearly demonstrate an actual danger" suggest that the probability be quite high. The virtually inevitable result of the ambiguity, namely overconservatism, has occurred at Patuxent, as will be seen later.

The Institution also provides treatment, which is on a voluntary basis in that an individual does not receive special punishment if he does not participate. For practical purposes, however, to be released from the institution, he must participate. The treatment consists essentially of group psychotherapy, occasionally individual psychotherapy. Additionally, certain administrative approaches provide incentives and punishments.

These are called the "graded tier" system, in which an individual who can spend a certain amount of time without getting into trouble is "promoted" from one degree of restriction to another which is less onerous. There are four such levels. Disciplinary measures include demotion, solitary confinement, increased lock-up time, and other prison discipline methods. In addition, operating as incentives, opportunities for acquisition of educational and occupational skills are provided at the institution, as well as social service aid, outpatient followup, and a halfway house.

Note that treatment is not specified as to dosage but is provided on a continuing basis. There is no "course of treatment" for comparison purposes. Beyond "Defective Delinquent" there is no further publicly stated formal subdivision diagnostically as far as allocation of treatment is concerned; however, some inmates receive only group therapy while others receive both individual and group therapy. There is, in effect, a total treatment program which may vary from one inmate to another.

If a person fails to "progress," he may be retained indefinitely at the institution until he does. As far as can be determined, the institution does not make a diagnosis of "hopeless case"; nor is it clear whether treatment would be discontinued if such a diagnosis were made by the staff, nor whether attempts would be made to transfer him out of the institution.

Evaluation

As was noted above in considering the logic of diagnosis, the clinical method of evaluation is essentially valueless as a method for making comparisons in a systematic or scientific manner. Because it is mysterious and authoritarian by nature, it allows no independent means of verifying a diagnosis. Clinical diagnoses are thus intrinsically unreliable. Furthermore, everyday experience suggests that experts can differ in their diagnostic impressions of a case. Which, if any, of the experts who disagree is the "right" one? In addition, clinicians, unless they have an exceptionally systematic turn of mind, do not tend to follow up on cases in order to ascertain how accurate they have been in their diagnostic predictions and to improve their methods. The writer has never seen in a psychiatric service organization a follow-up conference in this field

analagous to the pathological conferences in general medicine. Such method tends not to be psychiatric style. Furthermore, such a systematic follow-up tends to lean toward objective decision-making and to undermine authoritative or clinical decision-making. In a way it is by nature anticlinical and is likely to be eschewed by clinicians.

The clinical method, however, has been used at Patuxent throughout its 18-year history, and despite the subjectivity of the method in any given case, some observations can be made regarding its overall employment in groups of cases. We will examine three areas of functioning in the Patuxent operation. They are 1) the consistency and stability of inmate population and staff diagnosis throughout the years, 2) the accuracy of diagnosis, and 3) the effectiveness of treatment. Certain inferences can be formed regarding institutional performance despite problems which might exist in evaluating the evaluations of any individual case.

Let us first consider the consistency and stability of the population of individuals evaluated at the institution and of staff approaches to evaluation. From 1955 through mid-1973, 2,098 individuals had been diagnosed by the institution staff.[12] 326 additional persons had been admitted for examination during the time interval, but their cases, for various reasons, had not been diagnosed. The 2,098 individuals were referred by the courts to the institution after conviction and sentencing. There is no necessary expectation of homogeneity in those cases which were referred by various judges over an 18-year span. However, it can be assumed that there was some homogeneity in that judges would tend to refer cases they considered to show a strong likelihood of aggravated recidivism. These are non-psychiatric evaluations, however, by laymen who are unfamiliar with the psychiatric aspects of making recidivism predictions.

Those diagnosed DD by the institution staff are expected to be a highly homogeneous group, of whom almost all would have been expected to commit an aggravated criminal act if given opportunity to do so. Those diagnosed non-DD would be expected to be less homogeneous. However, because of uncertainties in making the determination, one would expect some aggravated recidivism on the parts of those diagnosed non-DD. (Since the legislature provided only two diagnostic possibilities and since the indications for a positive DD diagnosis are strong ones, the doubtful residuals go into the non-DD class. Obviously, the more reluctantly the staff assigns the DD label, the greater will be the expected offending in the non-DD group.)

There are some indications of a lack of homogeneity and of the employment of different diagnostic criteria (as well as different referral criteria by Maryland courts) during different periods in the institution's history.

Perhaps considering a simple dichotomy, namely the early years vs. the late years, is the easiest approach to appreciating this difference. In the first 10 years of Patuxent's existence, 1,101 examinees were diagnosed. Of those, 821 (82 "positive" cases per year from a total of 110 diagnoses per year), or 75%, were regarded by the staff as DD's; and 280, or 25%, were diagnosed non-DD. In the last 8 years 1,007 individuals were diagnosed. Of those, 556 (71 "positive" cases per year from a total of 123 diagnoses per year), or 55%, were diagnosed DD, while 451, or 45%, were diagnosed non-DD. Thus in later years fewer individuals constituting a smaller fraction of those admitted for evaluation, have been diagnosed as defective delinquent than was the case in former years. That fact alone may imply differing criteria for referral to Patuxent, possibly coupled with different diagnostic criteria on the part of the staff. (It is, of course, also possible that the only difference was the staff's changing its diagnostic criteria.)

Interestingly, however, there have been other changes suggesting lack of homogeneity in the evaluation population. In the first 10 years 726 men received court hearings after they had been diagnosed DD by the Patuxent staff. Of these, 595, or 82%, were adjudicated DD by the court or jury while 131, or 18%, were not. In the last 8 years 507 men received hearings and 459, or 91%, were adjudicated DD, while 48, or 9%, were not. Thus when a higher proportion of examinees were diagnosed DD by the staff,

the court concurred with fewer of the diagnoses, while when a lesser proportion of examinees were diagnosed DD by the staff, the courts were more nearly in agreement.

We have not yet established whether the courts, the institutional staff, or both, were changing their criteria. Some data, however, suggest that both the referring courts and the staff have changed criteria, and perhaps the courts and juries that hear the defective delinquent proceedings have too. Thus in the first five years' operation of the institution (1955-59) only 41% of the individuals referred to Patuxent for evaluation had been sent there after a conviction for a crime against the person (murder, robbery, assault, and rape), while from 1970 to 1972 the proportion sent after conviction for such a crime was up to 71%.[13] The average sentences given to the examinees for the immediate pre-referral crimes in the first five years were less than half what they were in the 1970-72 period, i.e., 4½ years as against 10 years. Thus the implication is that the average case sent to Patuxent in the latter period was a more serious case than the average one sent there in the former period. Yet despite the more serious nature of the cases, a smaller proportion was diagnosed as defective delinquent by the Patuxent staff. The most reasonable explanation appears to be that the courts have been sending more serious cases to Patuxent, while the institution's diagnostic staff has also been far more restrictive in recommendations regarding defective delinquency. Since the percentage of agreement with the institutional recommendations by courts and juries has recently been higher than before, it is suggestive that courts and juries making those adjudications are also more restrictive than they were formerly.

The rising crime rates over the years, along with increased numbers of serious crimes, explain why, if judges refer the most serious cases to Patuxent, the severity of crimes and sentences has increased in Patuxent referrals. It does not, however, explain why there have not been many more referrals to the institution. If criteria are the same as before, greater numbers of crimes should mean greater numbers of criminals, both milder and more severe, and therefore more referrals to Patuxent. Yet there weren't. Indeed, it appears that if the institution has ever been filled to its 600-bed capacity, it has not remained that way for long. It almost seems as if an informal quota system exists in which the referrals to the institution skim off enough of the relatively worst offenders in the system to fill the institution to a certain point, and that the quota-filling system is independent of the severity of the criminal acts of the individuals involved.*

One other difference between the early and the late groups is their racial composition. The early group of referrals was 74% white to 26% non-white (mostly black). The latter group of referrals was 44% white to 56% non-white, a dramatic difference.

The above differences are interesting in themselves, perhaps, but in the context of treatment evaluation they pose fundamental problems. If two populations available for comparison are different both before and after the application of treatment measures, there is no way to know whether application of the measures has led to the later difference. Comparisons between early populations at Patuxent and present populations do not necessarily indicate anything other than that the populations were different to

---

* Since the initial writing of the paper I have had the opportunity to examine the "Report of Commission to Study Changes and Basis of Selection for Patuxent Institution, December 9, 1965, to Governor J. Millard Tawes." It indicates that "In 1964 when overcrowding of Patuxent appeared imminent the Commissioner of Correction suspended all referrals from the prison system" (p. 21). The Commission "noticed some reduction in the number of men referred to Patuxent by the courts and supposed that some of the Judges are aware of the limited capacity of Patuxent and are exercising restraint in making referrals" (p. 21). "Under this tight quota the Commissioner naturally tended to give considerable weight, in making the selection, to the protection of the public against dangerous men who might otherwise be released. The result was that in the opinion of the Director of Patuxent the men referred from the prison system tended to be predominantly of the more hardened category and lower than average prospects for successful treatment" (p. 22).

start with. Any set of diagnostic or treatment measures which might have been relevant to the earlier group are not necessarily relevant to the latter group. Expectations as to outcomes in the latter group based on experience with the earlier group may be erroneous.

One final note on change in criteria for referral: it is obviously self-defeating from the standpoint of the purposes of the law. Either the courts and the institution were accurate in fulfilling the law previously, or they are accurate now, or they have not been accurate at either time. In any case, change in criteria neither serves justice nor is it consistent with proper public policy.

The thinking of the institution staff also shows lack of specific consideration of the law. Thus all evaluations of the Institution's performance have been in terms of *criminal recidivism*; these evaluations ignore the statute's aim, the prevention of *aggravated* criminal recidivism, a significantly different criterion. Furthermore, the staff has, *de facto*, but without published consideration of the implications of alternative policies, decided on the meaning of the statutory words, "reasonably safe for society to terminate the confinement." Decisions to send inmates out into the community on parole or a similar program short of total release, but nevertheless free from confinement, seem to have been made at a 35-40% *recidivism* (not aggravated recidivism) level. Dr. Boslow speaks of "calculating success in parole in terms of not committing a crime"[14] and mentions a figure of 35%. Dr. Hodges says "The current finding that 37 per cent of 156 parolees had new offenses suggests the Patuxent staff is maintaining consistent criteria for granting parole."[15] Although the "reasonably safe" language of the statute may mean a 35% recidivism rate, one would feel more comfortable had there been discussion of consideration of alternative policies as well as some rationale for selecting this particular one. In some ways it almost appears as if the staff used implicit criteria for parole, and the figures came out to 35% recidivism level without any prior consideration of cut-off levels. One wonders whether if the staff had tried for a recidivism level of 20%, or of 50%, they would have been able to modify parole criteria and achieve it.

As indicated above, the criteria for placing cases into different diagnostic categories are important insofar as they contribute to homogeneous D-T-O categories. The diagnostic staff at Patuxent faces two separate problems germane to homogeneity of D-T-O categories. The diagnostic problem of the first kind is that of prediction according to the statute. Whatever be the patterns of observation that lead to the conclusion, the diagnosis of defective delinquent must be homogeneous with respect to the outcome, in that persons so diagnosed and not incarcerated or otherwise restrained commit what are essentially crimes of actual or potential personal injury, i.e. crimes of violence, shortly after they are so diagnosed. Furthermore, individuals diagnosed not defective delinquent must be homogeneous with respect to outcome, in that persons so diagnosed and not restrained do not commit crimes of violence.

Both are simple D-T-O categories in which treatment controllable by the evaluator is nil. Thus, these are D-O categories. However, since it is apparent that individuals may commit violent crimes for many different reasons, it is very likely that diagnoses of both "defective delinquent" and "not defective delinquent" are disjunctive diagnostic concepts. That is, patterns a,b,c . . . , or d,e,f . . . , or etc., would indicate defective delinquency, while patterns t,u,v . . . , or w,x,y . . . , or, etc., would indicate non-defective delinquency.

The second diagnostic problem faced by the staff is that of developing D-T-O categories with respect to the treatment possibilities that are or might be offered at the institution. One would not be surprised if there were conjoint or disjoint D-T-O categories. One would hope that there would be different categories for the different kinds of treatment modalities. The notion "defective delinquent" may be quite useful as a predictive concept without being at all useful as a concept upon which to base treatment decisions, or indeed the decision whether or not to treat at all.

We are immediately faced with an insuperable problem. Because the evaluations

are "clinical," there is no way to determine the criteria used by the staff in order to arrive at the prediction that an individual evaluated is likely or unlikely to commit a potentially injurious crime. Thus there is no way to know the reproducibility or, in effect, the reliability of such judgments. Although a clinical diagnosis surely is based upon processing of information presented by the examinee and his records, with a subsequent weighing (i.e., ascribing relative probative importance to different aspects) of the information, no objective means exists to insure that the weighing of the facts is the same by different clinicians or on different occasions.

The non-objectivity of clinical determinations does not necessarily render them invalid, but it does render impossible appropriate comparisions of cases on the basis of anything but the clinical judgments.

Nevertheless, possible criteria do exist to evaluate such clinical predictions. Two approaches are possible. First is to evaluate people completely, then expose them to conditions which check the criteria. In the case of Patuxent the procedure would be to evaluate a number of men thoroughly as to whether or not they were defective delinquents and then to expose both the men who were diagnosed defective delinquent and the men who were diagnosed not defective delinquent to non-constrained conditions of society. The criterion of evaluation would be the homogeneity of each group with respect to committing or not committing potentially injurious crimes. (Note that the criterion is not uncontaminated itself. Not only do some potentially injurious crimes go undetected, but also, ascertaining to what degree the different allegations of the facts of an offense reflect the facts as a group of unbiassed observers might see them, and from that inferring the degree of potential injury, is a highly judgmental job. In any case that process is one potential approach.)

Another potential approach is for someone to try to evaluate the clinicians' evaluations to try to ascertain how in fact the different considerations were weighed by the clinicians and to try on that basis to render predictions more objective. The ultimate criterion for evaluation, however, namely that of exposing individuals with the two diagnoses DD and non-DD to conditions of society and recording the potentially injurious crimes committed by both groups, would still be necessary.

The two major problems in such studies are 1) such an evaluation requires considerable ouput of resources and is beyond the fiscal capability which has been allocated by the Maryland Legislature to the Patuxent Institution; 2) if an individual were evaluated as likely to commit an injurious crime, under the statute he could not be released except by a court order, and judges would surely be reluctant to release into the community a potentially dangerous individual.

Nevertheless there are some pertinent data at Patuxent, though they seem to have been collected more by accident than by design. Many men have been judged by the institution staff as defective delinquents but not so regarded by a judge or jury. Many of those individuals have been released into society, and there has been some opportunity for follow-up evaluation. Therefore some data are available on those diagnosed as defective delinquent, and thus information is available on "false positive" diagnoses. I am not aware of follow-up studies on those diagnosed not defective delinquent, so that unfortunately no information is available on "false negative" errors of diagnosis, a rather important shortcoming.

We turn to the first problem of diagnosis. How accurately does the staff determine whether individuals will commit aggravated crimes if given the opportunity to do so?[16] The nature of the question of accuracy of prediction in the context of the institution raises problems of evaluation. Obviously the only way to tell whether or not a person will commit a crime if given the opportunity is to give him the opportunity. But over 80% of individuals concerning whom that prediction was made were not given the opportunity, because they were retained in custody at the institution. Of course, given reasonable numbers of cases, a sample of one case in five can provide an extremely

precise appraisal of some characteristic of a given population, provided the sample is suitably representative of that population.

Were the one in five who were judged likely to commit aggravated crimes, but who were not retained at Patuxent because courts or juries did not adjudicate them as DD's, a representative sample of the total populations considered by the institution to be DD's? The best method to obtain a representative sample is by random selection. It is obvious that no effort was, or could have been, undertaken to select randomly those men diagnosed DD who would not be adjudicated DD in court hearings. The next best method of obtaining a representative sample is by randomization within a set of strata into which the larger population has been divided. It is obvious that such a procedure could not have been undertaken either. Indeed, no measures could have been taken *a priori* to insure representativeness of such a sample.

The best that can be done *post facto* is to compare the sample with the entire population according to a set of characteristics, and to ascertain the similarity on that basis. Although it is likely that considerable data are available for evaluation on this approach, the only published data[17] indicate that the court-committed DD's are relatively similar in IQ to those staff-recommended to be DD's, but they tend to be a bit different in age (younger) and age at first conviction (younger), both of which factors tend to mitigate in the direction of recidivism, and in number of prior convictions (fewer) and length of sentence for the immediate offense (shorter), both of which factors lean in the other direction *vis-a-vis* recidivism. The writer knows of no way to form a comparative expectation of recidivism for the two groups. On one hand it is possible that the expected recidivism, either general or aggravated, is the same for both groups; on the other hand it is possible that the expectations are different. Indeed the expectations may even be different with respect to different kinds of recidivism (i.e., "plain" vs. "aggravated") in either direction for either group. The unfortunate situation is that any guess is but a speculation. The only way to study such differences is to obtain proper empirical facts, which are now unavailable.

A further complication is that predictions sometimes end up being made for a distant future. A man may well be diagnosed DD, not adjudicated as DD by the court, and then sent back to prison to serve out his sentence. He may be in prison for years before he is exposed to the community and to criminal opportunity. Although long-range divination may be possible, it seems rather unreasonable to ask a diagnostician to predict what another human being will do in a few years, particularly when the experience he will undergo between the time of diagnosis and the time of proving is as unpredictable and varied as the prison experience. It is also desirable that all the predictions be for the same time range. Although it may be possible to make predictions which for some individuals mean only a few months' anticipation and which for others mean years' anticipation, that task too appears to me to impose unreasonable burdens on the diagnostician.

To this point we have concentrated on those persons diagnosed DD by the staff. There remain those diagnosed not-DD by the staff. In contrast to those diagnosed DD, those diagnosed not-DD are much more likely to be exposed to criminal opportunities in the community within a reasonable time after diagnosis. After such men receive negative DD evaluations from the staff, they are returned to prison and handled the same as other prisoners, so that many probably attain reasonably early releases. An observer would, of course, expect some aggravated criminal recidivism on the part of the non-DD group, depending on the cut-off point for the DD-non-DD diagnosis. In any case, information on past-release recidivism and aggravated recidivism might be helpful in assessing diagnostic effectiveness of the staff.

As noted above, the present writer so far has not been able to obtain any follow-up data on these individuals. For whatever reason, this critical information is not available.

No way exists, then, to evaluate the effectiveness of the staff's diagnosis of not defective delinquent.

To return to what the data show with respect to those who were diagnosed DD: One main source of information is Dr. Hodges' study.[18] He compares three classes of those diagnosed DD by the institutional staff and subsequently released to the community for at least 3 years or until re-incarcerated. The three classes are 1) those not adjudicated DD by the courts, called the "Untreated Group"; 2) those adjudicated DD by the courts initially but later released by the courts against the institution's advice, called the "Partially Treated Group"; and 3) those released by the institutional staff, called the "Fully Treated Group." The criterion that is used to evaluate the three groups is "recidivism," i.e. being convicted of a criminal offense of a severity sufficient to warrant reporting to the FBI. (The data on a handful of cases, those charged with an offense without the final disposition's being known, may be eliminated.)

According to Dr. Hodges, 95 of 118, or 81%, of the "Untreated Group" were recidivists, while 111 of 156, or 71%, of the "Partially Treated Group," and 57 of 156, or 37%, of the "Fully Treated Group" also committed new offenses. (There is some ambiguity in the article about the last group. It was mentioned that 31 men "had to return to the institution for further treatment," so that it is possible that the proper figures for the "Fully Treated Group," in order to ensure 3 years' exposure, should be 57 of 125, or 46%. The two figures seen to represent the outside limits.)

Dr. Hodges notes further that some of the recidivistic crimes committed by the various groups were misdemeanors (and thus not aggravated crimes), while some of the felonies were property offenses and some were personal offenses. On the assumption that only personal offenses represent aggravated crimes (indeed, it is likely that even some personal offenses are not aggravated), we recompute Dr. Hodges' figures. Thus 39 of 118, or 33%, of the "Untreated Group," 30 of 156, or 19%, of the "Partially Treated Group," and 15 of 156, or 10% (or at the extreme 15 of 125 or 12%), of the "Fully Treated Group" committed aggravated crimes.

Dr. Boslow presents some figures[19] regarding recidivism concerning the same cases as Dr. Hodges as well as some additional cases. These figures do not distinguish personal offense felonies, but represent only total offenses. He differentiates among those individuals who have been released against the staff's advice by courts, of whom there are two groups, one with and one without conditional release experience, and those who were discharged fully after three years' release to the community on parole status. To summarize his figures, 83 of 186, or 45%, of those without release experience, 39 of 100, or 39%, of those with release experience, and 10 of 135, or 7%, of those fully discharged after three years' parole, committed new offenses. (If the proportion of personal offense felonies to total offenses is the same as in Dr. Hodges' group, about ¼ of those numbers represented aggravated crime recidivists.)

These data indicate that in a *majority* of cases in which the diagnostic staff predicted that an individual would commit an aggravated or violent crime, they were in error. The number of false positive diagnoses of defective delinquent is greater than the number of true positives. Diagnostic accuracy and precision must be severely called into question by such a large number of false positives.

The question must be raised as to whether the present diagnostic process is an asset or a liability. Would as good or better results be obtained by declaring everyone to be admitted to the institution to be a DD, or by using some simple criterion like declaring everyone with 4 previous felonies and a present offense against the person to be DD? Surely one wonders whether the diagnostic process has done a better job of predicting aggravated criminal recidivism than would a random assignment of DD diagnoses. Of course, without data concerning the fate of those declared non-DD, the questions are unanswerable.

In any case, from the standpoint of homogeneity of the diagnostic classification, the

fruits of the diagnostic process at Patuxent are disappointing. A 33% level of homogeneity is in the writer's view an unsatisfactorily low performance. And 33% represents the highest level of aggravated crime commission noted in these groups. In some ways the diagnostic process might be thought to be even worse with respect to men already in the institution, for only 20% of those recommended for retention in the institution have gone on to commit aggravated offenses. Perhaps the discrepancy is that the diagnostic staff has interpreted the law in such a way as to use different standards to evaluate those inside the walls already, from the standards used to determine whether or not a person should enter in the first place. More likely is that the staff is more conservative in judging people already there than in judging those who might enter.

The staff's diagnostic accuracy is perhaps no lower than that of other forms of clinical diagnosis. But the real issue in any diagnosis is not the affixing of a label for statistical purposes. It is the utility of the diagnostic classification for decision-making purposes. And that utility depends on the stakes. In a hypothetical situation in which, of 100 people in a diagnostic class, 40% would die within a week if they were all untreated, while only 10% would die if they were treated, and the treatment had no side effects other than costing fifty dollars, a person in that class would be foolish not to take the treatment. However, if three-fourths of those treated ended up blind, everyone in the class would have reason to consider gambling. When the stakes are high, homogeneity and predictability become much more important.

In a situation in which a likely consequence of receiving a diagnosis of defective delinquent is incarceration of four years or more, 33% does not appear to the writer to be a sufficiently high homogeneity figure to warrant continuing use of the diagnostic process as a basis for making such decisions. Its continued use, however, is not surprising, but rather typical of institutional conservatism. The fact is that mistakes of false negative (i.e., a man declared non-DD commits an aggravated crime) are called to the attention of the diagnostic staff ("Why did they let him out?"), while false positives are undetected because they are retained in the institution. There is virtually irrestistible pressure on the staff to consider doubtful cases to be DD's and to retain them. As a result the false positives multiply in such institutions.

What about the second problem of diagnosis, that is, diagnosis as a basis for assigning treatment? From published records it does not appear that diagnosis is used at the Center as a basis for applying different treatments. There seems to be no indication of any within-institution formal diagnostic classification that leads to different treatment for members of the different diagnostic classes; that is, there are no intra-institutional D-T-O categories. There must be some kind of basis for making individual treatment assignments, since some inmates receive group psychotherapy and some receive individual psychotherapy; however, it appears to be an informal, *ad hoc* assignment system, rather than one based on systematically recorded experience. The same seems to be the case for decisions regarding an individual's changing tiers or going out on work release, etc. Several people are involved in the decision-making process, but the decision rules are not formalized and are perhaps variable from case to case.

Next, the efficacy of the treatment program should be surveyed, for despite its limitations the system may work. Certainly many people believe that it does work. Implicitly the logic of their argument, which uses the figures cited above, is that individuals diagnosed defective delinquents are homogeneous. They are treated in different ways, some receiving treatment in the institution and others receiving no treatment. Differences in recidivism are the criterion and are a function of the treatment. The logic is a standard inductive one.

Two basic problems call in question whether this logic in fact models the situation at Patuxent. First is the question as to whether those individuals diagnosed DD by the institution and not committed by the court are homogeneous with those diagnosed DD by the institutional staff and committed by the court. As noted previously, no clear

evidence on the point is available; there is no way to tell from the present data. Surely, though, if the groups are not homogeneous, the differences shown by Dr. Hodges[20] on Chi-square testing may be merely a reflection of an initial difference between the groups, and not a result of the treatment program.

The second basic problem is that of the many post-diagnostic differences between the groups. Disregarding such issues as dosage of treatment, or group vs. individual psychotherapy, or the limitations upon the benefits of psychotherapy to a person with an IQ of 79 or below (one-fourth of the Patuxent population has an IQ of 79 or below[21]), other differences exist between the groups. Such things as different atmospheres in different prisons, different lengths of time under incarceration, differences between the graded tier system and other administration arrangements, different personnel-to-inmate ratios, differences in availability of social services, differences in parole follow-up, differences in total money spent per inmate, and the differences between indeterminate sentencing and determinate sentencing, all tend to differentiate between the two groups. Which if any of those factors or combinations of factors is responsible for the differences? Without further systematic studies, there is no way to determine that. Speculation alone does not give a sufficient answer.

One other datum presented is worthy of further comment, namely the 7% recidivism figure for those individuals who have been on parole status from the institution uninterruptedly for a period of three years. That is seemingly an impressive figure. However, the real problem, again, is the lack of a control group for comparison. For it is well known that recidivism tends to follow early upon an individual's release from prison custody. There is very little criminal recidivism after three years. Thus any individual discharged from any institution who can last for a period of three years without recidivating has a good chance of continuing indefinitely without committing another crime. It seems very likely that little difference would exist between any group which has survived parole 3 years and the comparable group from Patuxent. Surely data are potentially available on this point.

Finally, a vital question is that of the degree to which operation of the defective delinquency law constitutes crime prevention. Dr. Hodges[22] compares the number of recidivists in the DD diagnosed group with an expected number derived from the experience with the men who were declared DD by the institution staff but not officially adjudicated as such. He found some 263 actual recidivists among a potential 654 recidivists, or a reduction of 391 individuals. He estimates that the group of 391 individuals, plus an estimated 109 or more from 5 more years' operation of the statute, might have committed well over 1500 crimes, which were therefore prevented by the operation of the law. Well might that number have been prevented. Well might they not have. No one can know. Yet it is also obvious that even more crimes could have been prevented if the entire group had been incarcerated permanently! Surely crime reduction is laudable and something to strive for strongly. Yet society must also be aware of the costs of crime prevention. And if that cost of preventing certain crimes is the deprivation of fundamental liberties of some people, the cost is too high. Society may well be better off tolerating many crimes than paying that kind of price to diminish them.

Summary

The Patuxent Institution and the Maryland Defective Delinquency law were conceived in response to several extremely important community needs and correctional requirements: to reduce violent crime by preventively detaining potential offenders and by rehabilitating offenders through treatment, to provide decent and humane conditions for offenders so that their incarceration can be of benefit to them, and to do both in an institution staffed by experts in human mental functioning and correctional rehabilitation, so that a true therapeutic atmosphere might exist therein.

The law and the institution are inspired by two fundamental assumptions: first, that psychiatrists can predict, by means of a "clinical evaluation," which offenders will recidivate and which ones will not; second, that psychiatric therapy has advanced to the point that it has something useful to offer in the treatment of offenders.

The present writer has tried to explicate the logic of evaluation of treatment programs through the basic concept of the diagnosis-treatment-outcome category. Diagnosis is a sub-category of the larger D-T-O category. The evaluation of any treatment program depends upon systematic and controlled experience given appropriate categorization of individuals.

The experience of the Patuxent Institution has not demonstrated its effectiveness. Its fundamental limitation lies in diagnosis: the institutional staff has been known to have made false-positive diagnostic errors to the extent that well over half the individuals incarcerated there may not have been properly retained according to the terms of the statute. No data have been presented publicly which might bear on the question of false negative diagnostic errors.

Certain data indicate that the more participation in the institution treatment program is undergone by groups of inmates, the lower is their criminal recidivism rate. The problem of evaluation of these data is the absence of adequate controls for the differences among the various groups. Thus it is virtually certain that there are differences in recidivism among the different groups whose data are presented. It is not known to what extent those differences are a function of initial differences among the groups or a function of subsequent differences, whether treatment or other.

Although certain types of research have been undertaken at the Patuxent Institution, it is apparent that the entire approach to the diagnosis and treatment issue has been that of a service-oriented organization rather than that of a scientific research organization. Whether through lack of facilities, lack of scientifically trained research personnel, lack of time, or whatever, the result has been that at this time, some 18 years after the inception of the facility, many unanswered questions remain about its true capability and accomplishments *vis-a-vis* its fundamental mission. On one hand there are indications that it is a poor performer with respect to diagnosis, and on the other hand indications that it may be helpful with respect to cutting down recidivism. Yet the problems have not been researched systematically, and easily-obtainable, important data have either not been sought or have not been presented so that outsiders may have access to them. As a result there can only be uncertainty as to the degree to which the organization, and indeed the law behind it, do the job.

Opinions and Recommendations

It is the writer's opinion that in the present state of the art of psychiatry, it is impossible to predict future violent criminal behavior with precision. Inevitably, therefore, in any such endeavor, many mistakes will be made both in releasing individuals who will subsequently commit violent crimes and in retaining individuals who would not commit violent crimes. The present author believes that even with the best of predictive methods, too many mistakes will be made to warrant the use of this kind of preventive detention, on the grounds that the cost to society is greater with such a law than without one. Others may have different opinions about this somewhat arbitrary ascription of values in a rather nebulous and difficult-to-test situation.

There is no real question, however, that the clinical method of evaluation is an inferior method of prediction to the use of objective measures and regression equations with cut-off points. Besides being mysterious, relatively unreliable, and subject to changes in decision criteria from time to time, it has on almost every occasion, when tested against objective methods of prediction, come out second best.[23] The clinical method can and should be dropped as the method of predicting aggravated recidivism. (This does not mean that a clinical examination cannot be a part of an objective

evaluation. It merely means that the ultimate decision procedure should be objective.) In any case, every diagnosis should be put in the form "It is predicted that . . ." with a percentage figure given. Merely affixing a diagnostic label is inadequate.

It is, of course, recommended that the easily available data regarding the possibility of false negative staff diagnoses be researched, as well as the similarities and differences among the groups with different recidivism rates.

Also recommended is further research into the indeterminate sentence and its effects on the inmates of the institution as well as on the public. The assertion that indeterminacy provides more therapeutic motivation to an inmate than does the potential reward of earlier parole from a maximum-length sentence ought to be tested in view of the notorious short-sightedness of many offenders. It surely ought not to be taken for granted as the major important therapeutic motivation.[24]

Surely one would think that if the program at Patuxent can prevent criminal recidivism among the worst criminal offenders in Maryland, it should have some beneficial effect if applied to less serious offenders in other prisons. Perhaps the indeterminacy of sentence is the critical factor, so that a therapeutic program would not work in a maximum sentence prison, yet the possibility that it might should be considered. However, systematic controlled studies of the different diagnostic and treatment factors must be undertaken as a preliminary to any further application of any program to large numbers of people.

Thus in the aggregate, because of the limitations of diagnosis, it is recommended that the indeterminate sentence be eliminated. On the grounds that the observed differences in recidivism between those individuals subjected to the treatment program and those not so subjected may in fact be *because* of the treatment program, the institution should be retained to continue to attempt to rehabilitate difficult offenders. The highest priority, however, should be given to systematically evaluating the effects of the various diagnostic and treatment factors.

References

1. Sines, J. O.: Actuarial vs. clinical prediction in psychopathology. *British Journal of Psychiatry*, 116, 129–144, 1970.
2. Sawyer, J.: Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 66, 178–200, 1966.
3. *Ibid.*, p. 192.
4. Article 31B, Annotated Code of the Public General Laws of Maryland, Volume 3, 1967 Replacement with 1972 Cumulative Supplement. Michie, Charlottesville, Va., 1967, 1972. (Hereafter cited as Md. Code Art. 31B.)
5. *Ibid.*
6. The Maryland Court of Appeals has, by implication, not completely agreed with this viewpoint. "Actual danger to society" has been held not to "limit the type or pattern of offenses to crimes against the person." See *Cowman v. State*, 220 Md. 207, 151 A.2d 903, (1959), *Eaton v. Director of Patuxent Institution*, 240 Md. 35, 212 A.2d 497 (1965). (But see note 13.) Furthermore the statute is civil in nature, thus requiring only a preponderance of evidence for a decision. Thus "persistent aggravated . . . behavior" may be shown by evidence other than convictions.
7. Md. Code Art. 31B.
8. Editorial comment by author: It may be that that differential criteria effect occurs merely because of the existence of institutions, that once a person is committed he falls into a category of persons different from others and should then be subjected to different legal criteria; however, I do not believe that is at all necessary. The criteria for incarcerating a person involuntarily ought to be the same whether or not he is an inmate of an institution at the time the decision is made.
9. Schreiber (Schreiber, A. M.: Indeterminate therapeutic incarceration of dangerous criminals: Perspectives and problems. *Va. Law Rev.*, 56, 602–34, 1970) has pointed out in a paper which makes most of the points made herein, that the lack of precision of the wording of the law, coupled with the felt sense of ignorance on the parts of most judges, attorneys, and jury persons regarding psychiatric phenomena, have in effect made the staff the

arbiters of the law. The staff's impact is heightened by their opportunity to testify to the "ultimate conclusion" as to whether or not the individual under consideration is a defective delinquent. See *Sas v. State of Maryland*, 334 F.2d 506 (1964).

Almost without realizing it, and probably without wanting to, the institutional staff has taken over legislative and judicial functions with respect to the very basic criteria of definition of defective delinquent. Schreiber concludes that this situation is very inappropriate, a point with which I am in strong agreement.

10. The length of diagnostic stay varies, but one person resided there diagnostically for a period of six years. The long stay was a result of his refusal to cooperate with the examining psychiatrists, and the individual was ultimately released after a period of time had elapsed equivalent to the maximum sentence for the crime of which he was convicted. The staff had concluded that they could not affix a diagnosis of defective delinquent to the man without a proper examination, and without his cooperation they could not perform a proper examination. It was recommended that the man be retained indefinitely in the institution, but the U.S. Supreme Court decided that to hold him indefinitely on an order committing him for a diagnostic observation would be denial of due process. See *McNeal v. Director of Patuxent Institution*, 239 Md. 407, 211 A.2d 737 (1965).

The issue of non-cooperation in a diagnostic study is an interesting and fundamental one. To what degree should it legitimately be demanded that an examinee cooperate with examiners in answering their questions in such a legal proceeding? Should a non-cooperating individual be jailed, in contempt of court, until he answers the questions, even if it might mean that he would be jailed for life?

Many factors bear on these questions. First, one must consider how completely an individual could answer questions if he used every resource at his command, and also how an observer could determine (particularly in a non-cooperative individual) what those limits were. Second, one must consider what degree of non-cooperation would be required to regard an individual as sufficiently non-cooperative to be in contempt. Should the degree of non-cooperation for contempt be contingent on the individual's ability to cooperate, or should there be a minimum level? One might say that the contemptuous degree of willful non-cooperation is that degree which prevents the examiner from forming a proper opinion. However, that leads to the most important point, the third: To what degree can the examinee's cooperative behavior in answering questions be a key factor in formulating an accurate prognosis regarding his potentially committing a violent injurious crime? To what degree in fact is a clinical interview necessary at all in arriving at such a prognosis? The Patuxent staff felt that proper prognosis could not be formulated without a cooperative patient. Yet to the writer's knowledge, that is a clinical authoritarian assertion, without systematic empirical precedent. As such its validity is not established. It may or may not be accurate, but there is no hard evidence on the point at this time.

Fourth, there are due process and public policy considerations. In a way, an examination to determine if an individual will commit a violent injurious crime is analogous to determining whether some specific hazard to the public health exists. An interview may be regarded as analogous to a search procedure. It may be argued that compelling a person to answer questions when he may be committed to a prison institution indefinitely is analogous to compelling him to testify against himself in connection with a criminal proceeding. However, it is also argued that these procedures are not criminal proceedings, so that Fifth Amendment protection is not relevant. The individual is not committed to a facility for punishment, but is sent there for society's protection and for his own rehabilitation; thus the Fifth Amendment doesn't apply. A psychiatric interrogation is viewed as analogous to public health authorities' searching a restaurant for a source of Salmonella infection, a procedure which is accepted as part of the police power of any jurisdiction. Though the analogy is imperfect, the most important considerations in its relevance appear to be two. First, there are specific indications for searching, specific, definite, and restricted methods of searching, and specific means of validly reporting results in evaluating a food service facility for contamination. The investigators know exactly what they need and can obtain their information directly with a specific legal mandate to do so. Second, there is a direct and known connection between the results of the public health search and the hazards to the public health. Interrogating a potential recidivist, however, is non-specific and is equivalent to a "fishing expedition," a procedure for which no warrant legally exists. If the examiners were able to restrict their inquiry to the answers to a limited number of specific questions, perhaps this argument wouldn't apply, but such restrictions would hamper the clinical method. Beyond that, whatever the results are of the search, there is no way except inferential ones of determining their reliability. Indeed there is also no way except inferential ones of establishing a connection between the answers and the degree of danger to the public. The art of doing so does not yet exist. Thus, I conclude

that no valid basis exists for compelling a suspected or potential recidivist to cooperate with psychiatrists or to answer their questions in this kind of examination.

11. If experience has indicated that a specific fraction of individuals in some given specific diagnostic group committed an aggravated recidivistic offense, it is "expected" that *in the future* that same fraction, X, of a diagnostic group identical, or at least indistinguishable from the original group, *will commit* an aggravated recidivistic act. That fraction is the "probability of aggravated recidivism" as applied to any individual in the diagnostic class.
12. Maryland Department of Public Safety and Correctional Services; Patuxent Institution: Annual Report 1972. Patuxent Institution, Jessup, Maryland, 1972.
13. The law does not restrict consideration as defective delinquent to those who have committed crimes against the person and in the past persons who have been considered only likely to commit property offenses have been adjudicated defective delinquent. See *Cowman v. State*, 220 Md., 207, 151 A.3d 903 (1959). However, as of 1971 the Federal Court found that the staff "does not recommend classification as defective delinquents any such person." See *Tippett v. State of Maryland*, 436 F.2d 1153 (1971). However, the staff does not seem to require previous conviction of an injurious or dangerous crime as a prerequisite to concluding that the individual will commit such a crime in the future.
14. Boslow, H. M., and Kohlmeyer, W. A.: "The Maryland Defective Delinquency Law: An Eight Year Follow-up." *American Journal of Psychiatry*, 120, 118–124, 1963.
15. Hodges, E. F.: "Crime Prevention by the Indetermine Sentence Law." *American Journal of Psychiatry*, 128, 291–295, 1971.
16. The writer wishes at this point to express an opinion. Despite the fact that diagnosing a person as defective delinquent means predicting that he will commit an aggravated crime, a psychiatrist in a diagnosing position is likely to find the task more palatable when put in the form of deciding whether a person is a DD or not than in the form of making a specific prediction. In psychiatry, of course, we affix labels to cases more as a matter of statistical record-keeping than anything else. And no one really suffers if the diagnosis is "wrong." If someone asks us specific questions like "Will this drug cure my schizophrenia?" we tend, if we are knowledgeable and honest, to hedge on the answer and to say that we can't be completely sure, or that we think it will because it works out that way most of the time, etc. I have never heard of any physician, let alone a psychiatrist, giving a money-back guarantee that he can make a specific prediction about a case. But we don't hedge like that at all if the question is one of affixing a diagnostic label for a hospital discharge summary or an insurance form. I believe that putting the question to the diagnostician, "Is this individual a defective delinquent or not?" results in easier feelings of diagnostic capability than putting it in the completely equivalent form, "Will this individual commit an aggravated crime or not?" I expect that fewer people would be regarded as falling under the purview of the statute if the diagnostic question were put as requiring a prediction rather than as requiring a labelling. Avoiding labelling would also avoid one important aspect of placing the burden of interpretation of the law in the hands of the staff. If they only testify as to predictions and the basis for those predictions, they do not testify as to the ultimate issue and do not thereby make law.
17. Maryland Department of Public Safety and Correctional Services; Maryland's Defective Delinquent Statute: A Progress Report. Patuxent Institution, Jessup, Maryland, 1973. (Hereafter cited as Progress Report, 1973.)
18. Hodges, E. F.: *op .cit.*
19. Progress Report, 1973.
20. Hodges, E. F.: *op. cit.*
21. Progress Report, 1973.
22. Hodges, E. F.: *op. cit.*
23. Sines, J. O.: *op. cit.*, Sawyer, J.: *op. cit.*
24. Hodges, E. F.: *op. cit.*