

63157

r

✓ EVALUATING A DELINQUENCY INTERVENTION PROGRAM: A COMMENT

Andrew C. Gordon
David McDowall
Northwestern University

Michael D. Maltz
Richard McCleary
University of Illinois,
Chicago

In the February 13th issue of Criminal Justice Newsletter, there appeared a three-column summary of "an important evaluation" by the American Institutes for Research (AIR) of the Illinois Unified Delinquency Intervention Services -- "UDIS: Deinstitutionalizing the Chronic Juvenile Offender."

Because there is a discouraged corrections community anxious to believe this evaluation, and because of a national dissatisfaction with deinstitutionalization (with which these findings are apparently consonant), we have been scrutinizing the report and have begun to reanalyze much of it.* Our strong impression is that the major conclusions either are an artifact of the methodology employed or have been exaggerated by that methodology. Moreover, it is misleading to conceive of the AIR report as a comprehensive evaluation of UDIS or as the commentary on deinstitutionalization it has been taken by many to be.

Although our reanalysis is not completed, it is apparent that if the findings and interpretations of the AIR report were to be implemented as policy, some of the best programs for youth might be dismantled and some of the worst affirmed. Since the report and its various summaries have received a wide airing, including a presentation to the U.S. Senate Subcommittee on Juvenile Delinquency, we feel compelled to alert the criminal justice community to several major deficiencies of the report suggested by our pre-

NCJRS

DEC 7 1979

* This reanalysis is being funded by the Illinois Department of Corrections. The first author was also instrumental in developing a tracking system for UDIS.

liminary reanalysis. In our opinion, these shortcomings are of sufficient magnitude to cast a "reasonable doubt" on many of the conclusions of the AIR evaluation report.

AIR evaluates the success of UDIS --- a broad range of alternatives for youths who would otherwise be incarcerated in Department of Corrections (DOC) institutions --- by comparing the records of youths in UDIS with those of youths sent to DOC. The AIR authors conclude that there is little difference between UDIS and DOC in terms of their effectiveness in reducing recidivism. They report a profound decrease in the delinquent activity of youths placed in either UDIS or DOC. In both cases, rapidly accelerating rates of delinquent activity are reported to be drastically reduced (or "suppressed," using the terminology of the report) after youths were "treated" by either of the two programs. The authors further conclude that the differences between UDIS and DOC were slight; that no less severe prior intervention (e.g., station adjustments, probation) had any salutary effect on these serious delinquents; and that the more intense the UDIS placement (e.g., wilderness programs and intensive psychological treatment, compared to foster and group homes), the more successful the impact on recidivism.

The AIR authors have correctly criticized the many studies which call any single post-intervention re-arrest a measure of recidivism equivalent to multiple re-arrests, and substitute a far better measure of program impact: the proportion of offenses reduced in a pre-/post-intervention comparison of delinquent careers. But their methodology has some difficulties which merit serious discussion. Our comments will focus on three major areas: (1) the validity of the major findings, (2) the persuasiveness of the explanations given for those findings, and (3) the relationship between recidivism and the quality of care.

SUPPRESSION OF DELINQUENCY: THE MAJOR FINDING

The "suppression" effect has been highlighted in every summary of the AIR evaluation. For each of a variety of highly correlated indicators (police contacts, arrests, court appearances, and violence-related offenses*), the report notes a profound reduction in delinquent behavior which has three important characteristics: (1) the drop is immediate and obvious, even in the first month after release; (2) the drop is drastic, reducing delinquency by as much as 74%; and (3) the drop interrupts a sharply accelerating (even exponential) rate of delinquent behavior for these youths. We will comment on these factors in turn.

Suppression and the Missing Delinquents

What disturbs us most about the controversy surrounding this report is that the various summaries ignore the problem of the "missing delinquents." Due to the timeliness of their evaluation, the AIR authors were forced to publish their findings even though 20% of the UDIS and DOC delinquents had not yet been released from programs or institutions. More important, at no point in their report do the AIR authors make it clear that over 30% of the delinquents were observed for less than one month in the post-intervention period. Published summaries of the AIR report have also failed to report this fact.

The finding that delinquent activity has been sharply and abruptly reduced is based largely on a series of plots which compare the number of "delinquent acts" (e.g., police contacts) each month per 100 youths before

* In fact, allegations of violence; none of these measures implies a conviction for the offense

intervention with the number per 100 youths following release. And the plots are visually compelling, showing a huge drop on each index. But while the data for each month before intervention include all 487 youths in the combined samples, the post-release data in successive months include only those youths who have been free for that period of time.

If the released youths were a fully representative subsample of all the youths in this study, we would have only the slight problem of increasingly unstable estimates due to decreasing sample sizes. But it is plausible to assume (and the data suggest) that the still incarcerated youths, including those on whom we have no post-intervention measures at all, were the "most delinquent" of all and that their inclusion in the before measure and omission in the after measure accounts for a large proportion of the "suppression" effect. To the extent that length of incarceration, seriousness of delinquency, and subsequent recidivism are positively related, this is akin to measuring everyone in a room, dismissing the tallest 25%, remeasuring the rest, and then concluding that everyone in the room has shrunk.

Careful scrutiny of the AIR report suggest that there was indeed a biased attrition of cases from pre-intervention to post-release. For the "most delinquent" youths, for example, those who had more than 16 prior police contacts, the attrition rate was 40%. For the "least delinquent" youths, those with only 1-8 prior police contacts, the attrition rate was only 25%. Clearly, the "most delinquent" youths were overrepresented in pre-intervention and underrepresented in post-release measures of delinquency.

If different proportions of UDIS and DOC youths are still incarcerated at the end of the observation period, it is also misleading to conclude that the two programs have had an equivalent impact on recidivism. (Moreover, the DOC youths remain on parole; if staying out of trouble is a condition of their release, they have an additional reason not to recidivate. In effect, the DOC "treatment" has not ended.)

Exponential Increases and the Regression Artifact

Whatever decrease in recidivism may remain after accounting for biased attrition rates, it is still an important decrease if it interrupts an exponentially increasing criminal career. But the possibility of an artifact in this increase is suggested by one's impression, surely misleading, that without the intervention, these few youths might be responsible for most of the crime in the United States within a few months!. In fact, this exponential increase is largely a statistical quirk, the mathematical result of accumulated rare events. The appearance of an exponentially increasing rate of delinquency in the pre-intervention period is a result of the same process that has generated a regression artifact in these data.

The meaning of regression in this context is that, other things equal, a post-release drop in delinquency occurs by chance alone when youths are selected for "treatment" on a basis of abnormally high pre-intervention rates. UDIS and DOC youths were selected for "treatment" largely on this basis. As a result, when the data are aggregated, an exponential increase appears in the aggregated data. And in post-release, a drop in the rate of delinquency is guaranteed.

As an analogy, consider the roll of dice. If an abnormally high number comes up (for example, eleven or twelve), the odds are as high as 35 to 1 that another roll of the dice will produce a lower number (for example, two or ten). Thus, when a high number appears, we can "treat" the dice, utter a magic phrase, and "cause" a smaller number to appear on the next roll.

The appearance of a smaller number has nothing to do with our magic phrase, of course, but is the result of a regression artifact. The

same principle applies to a reduction in the rate of delinquency for a selected cohort. Youths were selected for intervention because their rate of activity immediately prior to intervention was higher than normal. Conditions thus are ripe for a regression artifact. The AIR authors have noted this possibility but offer no estimate as to the magnitude of the regression artifact. Is it large enough to wipe out the "suppression" effect?

In our opinion, the rate of delinquent activity for some of these youths would have dropped without an intervention. (We discuss the acknowledged influence of maturation below.) A regression artifact is most certainly at work in these data. In our reanalysis, we are devoting the greatest effort to this problem.

REIFICATION OF PLACEMENT CATEGORIES AND THE "SOLUTION" TO DELINQUENCY

The authors of the AIR report reach conclusions about the comparative efficacy of the various program options within UDIS. Subdividing all program options into three categories, they claim that "Level III" placements, the most incarcerative (e.g., wilderness programs, intensive therapeutic residency), are the most effective. The authors then argue ex post facto that the Level III placements are more effective because they are more "convincing" to the youths and therefore are taken more seriously by the youths. Group homes, on the other hand; or so the authors argue, may incorporate the worst of two contrasting intervention models. Group homes are neither placements at home nor "drastic" enough to impact the youths. Before accepting these explanations as guides to future policy, however, several important issues must be raised.

Impacts of Regression and Maturation

The authors' claim for the superiority of Level III UDIS placements is based on a series of before/after comparisons in which two confounds are possible. First, youths have been selected for placement in Level III. Some of the before/after comparisons seem to fit the classic model of the regression artifact which we just discussed. A greater percentage drop in delinquent activity may occur precisely because the youths had higher pre-intervention rates of activity, and thus, had further to drop.

Moreover, a commonplace finding in the delinquency research is that youths tend to "mature" out of delinquency. The rate of delinquency for most youths will therefore drop over time whether there is an intervention or not. Few criminologists are surprised by the recent drop in national crime rates; the average age in the country is increasing, and as a population matures, the crime rate drops as a matter of course. When comparing the impact of various placements on recidivism then, it behooves one to make sure that youths placed in one program are neither younger nor older than youths placed in other programs. A greater drop in one program may reflect nothing more than a greater proportion of older youths in that program. Apparently no test on the relative ages of the UDIS youths across placements was made by the authors of the AIR evaluation.

Placement Labels

But a larger issue concerns the meaning of "Level III." In aggregating placements to Levels I, II, and III, the AIR authors were attempting to rank the various placements on the basis of how "drastic" an intervention each seemed to be. Level I placements were deemed the least "drastic," Level III placements the most "drastic." But in their aggregation, the authors grouped

together placements of widely varying types and placements which were attempting to impact youths in fundamentally different ways. Level III placements, for example, included both wilderness-stress programs and high security therapeutic hospitalization programs. Similarly, programs with vastly different reputations (but bearing the same descriptive labels, largely for purposes of remuneration; e.g., various stripes of "group homes") have been equated with one another. When apparent differences were found between the recidivism outcomes Level I, II, and III program categories, the authors assume, and all summaries stress, that their sorting criterion "explains" the difference.

Of course, it may be highly misleading to equate programs within such arbitrary categories. Wilderness-stress programs and therapeutic hospitalization programs are both labeled as "Level III" placements. While both of these programs are "intensive" they are quite different and their impacts cannot be easily compared. And for analytic purposes, it may be just as misleading to label a group of placements simply because they share a nominal category (e.g., "group home") as it is to label a group of delinquents simply because they share a nominal category (e.g., "chronic offender"). Sample sizes may have dictated some of these arbitrary decisions but this purely statistical issue should not misdirect policy.

The UDIS staff made important distinctions among the various placements in terms of "quality." One Level III placement was especially well regarded by the UDIS case managers and the dedicated staff of this program might well have been extraordinarily successful wherever they worked. Since the high proportion of youths in Level III UDIS placements were in that single highly regarded program, the "quality" of program personnel may be profoundly

intertwined with the "Level III" category. The AIR authors note the high ratings by UDIS case managers of the Level III placements, but in general, do not deal with the "quality" of the placement as it relates to recidivism outcomes. The assumption that program placements within a nominal category are interchangeable is thus not a wise assumption particularly among those placements that the AIR authors have labeled "Level III."

Recidivism as a Single Measure of Effectiveness

Finally, to the reduced extent that recidivism results may hold, can one conclude that the various programs of UDIS and DOC have been effectively evaluated? Even here, we think not.

While DOC may accept impact on recidivism as an appropriate single measure of effectiveness --- though we suspect that they would not --- it is a singularly misleading measure for characterizing what UDIS meant to accomplish. Recidivism has always been but one (and not necessarily the most salient) way in which UDIS has attempted to reform the juveniles in their care. Their emphasis has been on individualizing care, on close monitoring of the activities of their charges, and on modifying the UDIS program for individual youths when warranted by progress or the lack of it. While much of this is commented on in the AIR report, the published summaries tend to ignore these points in favor of the more spectacular claims about recidivism.

Moreover, UDIS emphasized not programs, but program packages where youths could be moved from placement to placement, often according to a prearranged plan. Concentrating solely on the first (and often very brief) placement, misses the emphasis UDIS promotes. In addition, UDIS staff were openly aware that some of their placements and some of their workers were better

than others. UDIS had hoped that an evaluation would illuminate these issues. But on this score, and on a variety of other characterizations of program quality, the AIR evaluation is nearly silent. It is particularly ironic that UDIS, which so stressed individualization and outcomes other than recidivism, should be subjected to an evaluation which collapses placements into broad categories and virtually ignores all but the most traditional measures of program success.

UDIS is not "Deinstitutionalization"

The AIR report has also been widely interpreted as a commentary on "deinstitutionalization," with the assumption that UDIS is an attempt to achieve that outcome. But this too is misleading. In fact, some of the placements within UDIS are fully as institutional as DOC. In part, these UDIS programs were a response to the political environment: juvenile court judges required UDIS to maintain secure detention facilities within its panoply of placements before the courts would take seriously the program as a whole. But these secure placements are (a) inappropriately thought of as alternatives to institutionalization, and (b) so expensive within the range of UDIS alternatives that they seriously inflate the costs of UDIS. UDIS costs as reported by the AIR authors are therefore not an adequate measure of the costs of "deinstitutionalized" care.

CONCLUSION

Because of these problems in analysis and construct validity, the AIR report may have left an inappropriate impression on the corrections community. The report has been widely distributed and in its present form is cited as a trenchant commentary on a variety of crucial issues in juvenile corrections.

The even more widely distributed summaries of the report lack the context the AIR authors provide, and thus, are particularly misleading.

Our opinion on the basis of a preliminary reanalysis of these data is that the finding of a "suppression" effect is overstated by the AIR report. But in addition, we believe that any study which characterizes such an apparently different program as UDIS solely in terms of routine measures such as recidivism so caricatures the program as to be irredemably misleading. We recognize the fact that "harder" measures (e.g., recidivism, dollar costs) are easier to handle with the currently available social science methods than are "softer" measures. But program quality is crucial to an evaluation, particularly when the goal of the evaluation is set policy guidelines. This point is much more a comment on the inadequacies of our discipline, and suggests that in this area as in so many others, social scientists and technicians should not be allowed to dominate the definition of problems or solutions.

This letter was sent to Editor
Criminal Justice Newsletter
National Council on Crime and Delinquency
411 Hackensack Avenue
Hackensack, NJ 07601

June 15, 1978

END