STATISTICAL ANALYSIS CENTER

72153

Illinois Law Enforcement Commission

A Methodological Review of

"The Impact of Manpower Services on

Illinois Offenders," by George W. Knox

October 18, 1978

Statistical Analysis Center
By Carolyn R. Block

CRIMINAL JUSTICE INFORMATION SYSTEMS
J. David Coldren, Director

ILLINOIS LAW ENFORCEMENT COMMISSION
Daniel W. Weil, Chairman
James B. Zagel, Executive Director

Printed by authority of the State of Illinois
June, 1980
Number of Copies: 150
Printing Order Number: 80-27

ABSTRACT

"The Impact of Manpower Services on Illinois Offenders" by George Knox is a report submitted to the Illinois Department of Corrections (DOC) in June, 1978 on the effect of the Comprehensive Offender Manpower Programs (COMP.)

This is a detailed review and criticism of the methodology of the Knox report, particularly its control or comparison groups, and its cost-benefit analysis. The review finds serious problems both in the quality and in the interpretation of the data.

The review concludes that the Knox study was an ambitious attempt at a controlled evaluation, but the attempt failed. Its cost-benefit analysis is based on an incorrect analysis of inadequate data. The conclusions of the Knox report should therefore be disregarded.

In a memorandum dated May 25, 1978, Statistical Analysis Center (SAC) staff member Karen P. Smith commented on the methodology of an evaluation design, "Evaluation Design of the Comprehensive Offender Manpower Program" (COMP) by Cybersystem Research, Inc. of Chicago (a copy of the memo is Appendix A). The memo pointed out several problems in the design. Cybersystem's final report on the evaluation, titled "The Impact of Manpower Services on Illinois Offenders," was published a month later, on June 30, 1978. George Knox was the author. The following is a discussion of whether the design problems mentioned in the memo are still present in the final report. This paper is not an evaluation of the entire report, but only of the four problems mentioned in the SAC memo.

Four Design Criticisms

The memo lists four criticisms. 1. A cost-benefit analysis is mentioned, but its design is not described. 2. The program's effectiveness cannot be determined without a control group. 3. There is also no control for studying the effectiveness of prison training. 4. The design says it will develop prediction equations for success and failure on parole, but the method for doing this is only vaguely described.

The final report does include a cost-benefit analysis, with a control group. This review will discuss these two components of the evaluation in detail, examine any inadequacies of design or of the way in which the design was carried out, and decide whether manpower services in Illinois can or cannot be evaluated with these data. Since the cost-benefit analysis and the method of choosing the control group were complex, and contain a number of methodological problems, the bulk of this review will only deal with these two aspects of the evaluation. Other aspects will be mentioned briefly in the next paragraphs and in the appendix.

The third problem mentioned in the memo is that the effectiveness of the training program cannot be evaluated without a control group. The final report had no control group for the training program. Knox, the author, tries to solve this problem by denying that he had ever attempted to do an evaluation. This denial is made somewhat tenuous by the presence of the word "impact" in the title and the lack, anywhere in the long report, of any specific statement of purpose. In the introduction to the chapter on training programs, for example, Knox states that the subsequent analysis was "not intended to test hypotheses ... but was rather designed as part of an evaluation" (page 336). This is a contradiction in terms. If one is to evaluate the impact of a program, one has, by definition, an hypothesis - the hypothesis that the program is better in some way than no program or an alternative program. To test this implicit hypothesis, a control group is necessary. Without a control group, there can be no evaluation.

The method used for the analysis of the effect of training programs was to track clients from training programs to placement programs. Knox reports that 150 clients were tracked. Since there was no control group of training clients who did not attend placement programs nothing can be said about the effect of placement programs from these data. However, the data could be used to give the reader a general description of training students who later go to placement programs,if the 151 clients could be considered to be representative of the larger group. This, unfortunately, is impossible to determine from the Knox report. We do know that there were 1423 clients in the five training programs in the sample (Table 47, page 258). If all of these were later in placement programs, then only 11 per cent were tracked. If there was some bias in which clients were tracked and which were not, these 11 per cent would not be representative.

There is evidence of some bias. For example, 53 per cent of all training clients were black (Table 21, page 209), but 65 per cent of the tracked sample were black; 20 per cent of all training clients but only 12 per cent of the tracked sample reported that they had less than nine years of school (Table 23, page 212.) This shows that the sample may be biased; therefore, the data are not only useless as a controlled evaluation, they also cannot be used as a description of training clients who later enter placement programs.

The fourth problem mentioned in the memo refers to Knox's intention to develop prediction equations for success and failure on parole. This was not attempted in the final report.

There are numerous other problems with the methods in this report, but the following review will consider only two in detail - the control group and the cost-benefit analysis. Appendix B and C contain discussions of two problems which are not related to the problems in the original memo, but which deserve some mention. Appendix B discusses Knox's creative use of the term, "internal validity." Appendix C analyzes a particularly interesting table, which Knox misinterprets.

Good Aspects of the Evaluation

The evaluation was quite ambitious and it used innovative methods which should have led to a very useful product-a competently done evaluation utilizing available data at low cost. Knox attempted to do three things that are especially worthy of note.

1. He attempted to use a control group. The design of the control group was good, and would have produced useful data if it had been carried out properly. The control group will be discussed in detail in the following section.

2. Knox attempted to measure cost-effectiveness instead of limiting the analysis to cost-efficiency. This was such an ambitious undertaking that it probably could not have been done correctly given the time allowed. The final section of this review discusses the cost-effectiveness analysis.

3. A very good aspect of the Knox evaluation was its use of already available computerized data. A wealth of such data exists, but is used all too seldom by evaluators and researchers. The only way to bring off such an ambitious evaluation given the time and money constraints would have been to use available computer files. Knox thus took the best approach to the problem of how COMP could be quickly and cheaply evaluated. Whether this approach succeeded is the subject of the remainder of this review.

The Control Group

Three of the four problems in the SAC memo dealt with the lack of control groups in the design. In the final report, however, there was a control group for one aspect of the evaluation - the cost-benefit analysis of the effect of COMP placement programs on state clients.

The ideal control group, of course, is randomly chosen from the same population as is the experimental group, in this case the COMP group. When a random assignment to the two groups is impossible, as it is in this case, one alternative is to choose the control group randomly from a population that is exactly like the COMP group in every respect, except in having been serviced by a COMP program.

This is what the Knox study attempted to do. If it had succeeded, it would have been possible to determine the extent to which COMP service was related to any other variable, for example, to recidivism.

The control group was drawn only for COMP placement clients who were from the state, not the Federal or a county system. Thus, conclusions can only be made about COMP placement clients who were in a state prison. There is no problem with this, however, since it is explicitly mentioned in the report.

For the control group to be adequate, COMP state clients and the controls should be drawn from the same population. Were they?

Although the report is quite unclear about its methodology, the control group seems to have been chosen in the following way. First, a comparison was made of COMP placement clients from the state system during FY1977 and all state system offenders since 1970, to determine if the two groups could be considered to be from the same population. It was decided that they could be. Therefore, the state offenders could be used as controls for state COMP offenders. Two controls were then chosen for each COMP client. In order to also control for "correctional history", the controls were matched for convicted offense and date of first incarceration, but were otherwise random. This was done by choosing as a control someone convicted of the same offense whose prison ID number was the next one before and after the ID number of each COMP client. These controls were again compared to the COMP client group, and negligible differences were found. Then the COMP and control groups were compared on the test variable - recidivism. This system of choosing a control group is quite good, and would have produced data adequate for a cost-benefit analysis if it had been more carefully carried out.

Unfortunately, there are a number of problems with it. These problems will be discussed chronologically.

The first step in developing the control group was to compare COMP placement state clients to the general DOC population to see if the COMP clients were systematically different. If so, DOC exoffenders would not be an adequate control group. There are the following problems with how this step was carried out:

1) The two groups were compared on variables which Knox had shown to have no relationship to either recidivism or getting a job.

2) The groups were <u>not</u> compared on other variables which Knox had shown to be related to recidivism or getting a job. These variables include sex and number of months in custody (see pages 293 and 297.)

3) The groups were also not compared on variables which Knox had not studied, but which have been shown by others to be highly related to recidivism. The most important of these is age at release.

4) In many of the comparisons, there was so much missing data that the results should be disregarded.

5) The most important difference which could disqualify DOC exoffenders as controls was not (and probably could not have been) directly measured by Knox. This is the COMP client's motivation to get a job. If clients of COMP programs are systematically screened so that only motivated exoffenders are accepted, the apparent success of the program could be due to the screening, not to the program itself.

Although systematic bias between the COMP and the control groups could not have been directly measured, there are indirect checks that could have been done on the existence of bias. For example, Knox's Table 4, on page 168, summarizes the screening decisions of each program, as reported by the program director.

(Knox does not check the accuracy of the program directors' claims. If records were kept of all applicants, those accepted could have been compared to those rejected.) The placement agencies reported between one and eight (a mean of 4.33) "absolute or desired" requirements for acceptance into the program. A mean of 1.8 were absolute requirements. Some of these requirements, such as being on parole or probation, would not affect control group comparisons. However, others might, such as being referred by a parole officer, a judge or an authorized agency (11 programs), being "employable" (9 programs), or being free of drugs (7 programs).

In addition to the initial intake screening, COMP clients who do not meet certain standards, such as showing up for appointments, may be dropped from the list of "official clients. This would inflate any percent success figures. It also would mean that COMP clients were highly selected, and therefore not comparable to the DOC group. Knox's Table 1 indicates that eleven programs may screen in this way. However, the survey question about screening was vaguely stated, and apparently the responses of the program directors were not clear enough to report (see Appendix A, page 460, question 4). The issue of screening is probably the single most important point for an evaluator of a social service to clarify. If there is no information about screening, there is no information about how the program defines its "client," and therefore, it is impossible to make any conclusions about the effectiveness of the program.

Knox attempted to overcome this difficulty after the fact by comparing COMP clients to DOC offenders on a number of variables. This may have been acceptable were it not for the problems one through four, discussed above.

Table 1 documents the first and the fourth problems on this list. Missing data for the state COMP placement clients ranges from 19 to 95 percent. For the DOC group, the minimum percent missing ranges from 19 to 91.[1] If we decide to ignore all findings where more than a third of either group is missing, we are left with only seven comparisons. Of these, only one, highest grade in school, could be expected to affect recidivism, according to Knox's data.[2] Even here, the degree of the expected relationship is so small as to be negligible, only two percent, and there are conflicting findings. There was a slight difference between the COMP and the DOC groups on this variable, however (see Table 2). Fewer of the DOC group than the COMP group went as far as tenth grade.

One variable showed a relatively large difference, but there was so much missing data that these results cannot be trusted. They are also presented in Table 2.

We may conclude, then, that there is reason to doubt that COMP clients and the average DOC offender are not systematically different. Knox has not examined differences in many variables that may be related to recidivism or getting a job. The one variable he did examine which had adequate data did show slight differences. If Knox proposes to use DOC offenders as a control group, the burden is on him to show that the COMP and DOC groups can be considered to be from the same population. He has not shown this.

---

[1] See footnote a to Table 1.

[2] Data for the "Expected Effect" column is taken from two tables in Knox's report, Table 72 page 293, which compares COMP clients with eleven sets of "background characteristics" on their recidivism, and Table 74, page 297, which presents the percent differences and Chi square significance of 175 2X2 contingency tables of fourteen "background variables" with "relapse rates" for fifteen placement programs. There are serious problems with the report's interpretation of Table 74. These problems are discussed in Appendix C.

TABLE I

Comparison of COMP state placement clients (Serviced) and General DOC 1970-1977 (Non-Serviced) Groups

Variable Description	COMP Client N	COMP Client % Missing[a]	DOC N	DOC Minimum % Missing[a]	Expected Effect on Recidivism (page source)	Differences Found
County of Origin	3,505	21%	25,205	25%	no information	Counties with programs have more COMP clients
Marital Status[b]	3,595	19%	33,437	a	no effect (297)	Can not be determined[b]
Number Illinois Commitments	3,416	23%	27,114	19%	no information	3% more COMP had no prior commitments
Number Other Commitments	3,343	25%	23,979	28%	no information	3% more COMP had no prior commitments
Military Status	3,370	24%	23,823	29%	no effect (293,297)	5% more DOC were veterans
Months Employed two years before Incarceration	230	95%	2,952	91%	no effect (293) possible effect of less than 12 months (297)	17% more COMP employed 11 or fewer months
Number of Children	2,939	34%	22,451	33%	no information	no difference
Alcohol Abuse	3,169	29%	23,320	30%	no effect (297)	no difference
Prior Arrests	707	84%	7,364	78%	no information	no difference
Age at First Arrest	1,835	59%	10,793	68%	no information	no difference
I.Q.	1,674	62%	9,094	73%	no information	no difference
Reading Score	1,462	67%	7,101	79%	no information	no difference
Highest Grade in School	3,473	22%	24,047	28%	2% more recidivism for 11 or fewer grades (293) no effect (297) (also, no effect on getting a job - p.311)	6% fewer COMP had 9 or fewer grades, but no difference at 11 or fewer grades

[a]The total N for these data (pages 372-392; Tables 107-119) is not given for the DOC, non-serviced group. It must be at least the N of Table 108, 33,437. Assuming there was no missing data for marital status, 33,437 is the N. The real N can be no less than 33,437 in any case. Minimum % missing was figured with this base. The actual % missing is probably, therefore, greater than this minimum. For COMP clients, the base used to figure missing data was 4462, the total state COMP clients, given in Table 47, page 258.

[b]The marital status table, Table 108 page 378, can not be interpreted, since there are categories which are not mutually exclusive, such as "married" and "married living with spouse."

TABLE 2

Grades of School and Employment before Incarceration for COMP group (serviced) and General DOC 1970-1977 (non-serviced)

	COMP	DOC
Grades of School[a]		
9 or fewer	877 (25%)	7,367 (31%)
11 or 12	1,458 (42%)	9,233 (38%)
12 or more	1,138 (33%)	7,447 (31%)
Total	3,473	24,047
N missing	989 (22%)	9,390 (28%)
Months Employed During two years before Incarceration[b]		
11 months or fewer	136 (59%)	1,244 (42%)
12 months or more	94 (41%)	1,708 (58%)
Total	230	2,952
N missing	4,232 (95%)	30,485 (91%)

[a]Source: Table 119, page 391.

[b]Source: Table 112, page 383.

Knox's second step in choosing a control group was to choose the DOC offender with the same criminal conviction, and the next higher or lower ID number, as described above. It was a good idea to use this mechanism to control for convicted offense, since five programs have "being a felon" as one of their absolute or desired requirements for acceptance, and it is reasonable to expect that conviction might affect what happens to the inmate in prison and after release. This produced the actual control group. However, there are the following problems with this control group, in addition to the problems already discussed.

1. Even though Knox checks the COMP client group and the control group for comparability, it is still doubtful that they can be considered to have been chosen from the same population. The comparison of the two groups is given in Table 121, page 398. The problems with the comparisons between COMP clients and DOC offenders, which were just discussed, also apply here. The groups are compared on ten variables, only two of which may be related to recidivism (months employed and grade in school), and variables, such as age, which affect recidivism, are not analyzed. Further, though the earlier analysis showed quite a problem with missing data, Table 121 does not mention the number of cases. Therefore, it is impossible to know the quality and trustworthiness of the comparisons.

2. Even though what Knox calls Cohort One and Cohort Two were supposed to have been randomly chosen in the same way, and should therefore be very similar, there is evidence that they are <u>not</u> similar. This means that they might not be considered to be a good control group, and that a cost-benefit analysis or other conclusions should not be based on that assumption.

Both "cohorts" were chosen by picking a case with the same criminal conviction as each COMP case, but Cohort One was composed of the case <u>before</u> the COMP case, and Cohort Two was composed of the case <u>after</u> each COMP case. In other words, there were two controls chosen for each COMP case. Aside from doubling the size of the control group, this should have had no affect on the analysis. For analysis the COMP group should have been compared to the control group - all of them. There is absolutely no theoretical or statistical reason for analyzing Cohort One and Two separately. Not only is it indefensible to do so, pretending to have two control groups is misleading. It can lead to false conclusions.

The effect that treating one control group as two has on conclusions of the cost-benefit analysis will be discussed later. At issue now is the representativeness of the one control group, the 4772 cases which were chosen for the 2386 COMP cases,[1] two controls for each COMP client. We have already discussed one major problem with this control group - that it might not be really comparable to the COMP group. Knox does not give adequate evidence that they are comparable on any variables which might affect recidivism, and the important issues of client motivation and the screening that placement programs give their clients are left unanalyzed. We cannot, therefore, conclude that the COMP and control groups are really comparable. However, let's assume for the moment that control cases chosen as they were for this study would be comparable, so that we can examine the 4772 control cases themselves. Were they actually chosen as stated, and do they therefore represent a sample of DOC exoffenders for criminal conviction and date of incarceration?

[1]This is the 1861 total in Table 120, page 397, plus the 525 missing COMP clients mentioned on page 396. See footnote d of Table 3.

No, they do not. That is, even if a group chosen as the Knox study design describes would have been an adequate control group, the group actually chosen did not meet the design's specifications, and therefore cannot be considered a control group. The reason for this is, again, the large number of missing cases. There were enough missing control cases to affect the recidivism rates of the controls. Further, there is evidence that recidivism was, in fact, affected. In other words, the missing control cases were probably systematically different from the non-missing control cases, the cases which were used for the cost-benefit analysis.

Evidence for this is found in Knox's Table 120, and the discussion on page 396. Table 3, of this report, is an attempt to untangle the origin of the COMP client and the control cases used for the cost-benefit analysis. Line 8 of Table 3 gives the number of COMP state placement clients for whom controls were sought. There were 2386 (see footnote d of Table 3 for the source of this figure.) This figure is only 53 per cent of all known COMP state placement clients. The other 47 percent is missing. In addition, there was a 21 percent loss of clients for whom state origin or recidivism were not known, so they could not be included (line 5). The COMP client sample, in other words, is 53 per cent of a group that is 79 per cent of the total. On top of this, another 22 per cent were lost before the cost-benefit analysis (line 9). Twenty-two percent of the COMP group for whom controls were sought are missing. This means that some matched controls were found for clients who never ended up in the client group for analysis. Overall, more than 58 percent of the client group is missing (58 percent in line 9, plus the additional missing in line 5.) If data on only 40 per cent were found for analysis, these must have been the 40 percent on which COMP agencies had records.

TABLE 3

Missing COMP Client and Control Data in Cost-Benefit Analysis

Data Description	Source	Number	Percent Missing	Base (line number)
1. Total sample of COMP clients	a) Table 47, p.258[a]	8685	-	-
	b) Table 1, p.157[b]	8037	-	-
2. COMP clients in placement programs	Table 47, p.258, and p.1	7262	-	-
3. COMP placement clients for whom criminal justice origin is known	Table 47, p.258	6430	11%	2
4. COMP placement clients for whom recidivism is known	Table 71, p.292	7186	1%	2
5. COMP placement clients for whom both 3 and 4 above are known	Table 72, p.293 (but see Table 73, p.295 and note c)	5725[c]	21%[c]	2
6. State criminal justice origin of all known placement clients	Table 47, p.258	4462	-	-
7. State criminal justice origin where recidivism is also known[c]	a) Table 72, p.293	2499[c]	44%	6
	b) Table 73, p.295	3828[c]	14%	6
8. State COMP clients for whom controls were sought	Table 120, p.397 and p.396[d]	2386	38%	7b
			47%	6
9. COMP clients used to determine recidivism in cost-benefit analysis	Table 120, p.397	1861	22%	8
			58%	6

Data Description	Source	Number	Percent Missing	Base (line number)
10. COMP placement clients for whom it is known if they were placed	Table 65, p.282	6945[e]	4%	2
11. COMP placement clients for whom both placement and criminal justice origin are known	Not given, but overall percent known is 89% (Table 47, p.258)	(6181[f])	-	-
12. COMP placement clients used to determine program costs and tax benefits in cost-benefit analysis	Table 123, p.401	3087	50% 31%	11 6
13. COMP placement clients used to determine averted correctional costs of cost-benefit analysis	Table 126, p.407[g]	3115	19% 30%	7b 6
14. Controls used to determine recidivism in cost-benefit analysis	Table 120, p.397	a) 1393[h] b) 1249[h] c) 2642[g]	42% 69% 48% 72% 45%[g] 70%[g]	8 6 8 6 8 6

[a]Agrees with the total in Tables 21, 24 and 52, and is the highest N of any general frequency distribution.

[b]"No. clients served in FY 1977" as self-reported by programs.

[c]The two figures in line 7, 2499 and 3828, should be the same. There is no explanation in the report of why they are not. (Perhaps the Table 72 figures are for those clients where all of the variables are known. If so, then the 5725 in line 5 is too low.)

[d]The total "serviced group" in Table 120 is 1861. This is defined as "the total serviced group during FY 1977 who were state system exoffenders," (page 396). But, "data on correctional status was not found for 525 of these cases." This 525 missing added to the 1861 equals 2386.

[e]Known data are the 7262 "total" in Table 65 minus the 317 "missing".

[f]Estimated from the overall percent known. Figure is 89% of 6945, line 9.

[g]These are apparently all state placement clients for whom recidivism is known, although the "expect relapse rate" was determined from the controls of another group - in Table 120. At any rate, since this total does not agree with the total in Table 72, it probably was taken from the data in Table 73.

[h]There were 2386 COMP placement clients for whom the study attempted to find controls. Two controls were sought for each COMP client. There should be 2386 controls in each half of the control group, but there are only 1393 and 1249.

[g]Cohorts One and Two combined. This is 55% of line 8 times 2 and 30% of line 6 times 2. They are multiplied by 2 because two controls were sought for each COMP state placement client.

It is reasonable to assume that these 40 percent were more likely to have been "successful" in the COMP program than the missing 60 percent. In other words, the COMP sample used for the analysis is probably not representative of all state COMP placement clients in general. It is probably biased towards having less recidivism than the average COMP client.

But what of the control group? The control group, naturally, has many missing cases because it is based on the COMP client group, which has about fifty percent of its cases missing (line 8.) On top of that, 525 (22 percent) of these COMP cases are missing. Because there could be missing data on two controls and on each COMP case, there are the following possibilities:

1. Cases could be represented in the control group (once or twice) but not in the COMP group.

2. Cases could be in the COMP group, and represented twice in the control group.

3. Cases could be in the COMP group, and represented only once in the control group.

4. Cases could be neither in the COMP group nor in the control group.

The last possibility includes over 60 percent of state placement COMP clients. We do not know how many cases there are in the first three possibilities. Knox's Table 120, however, gives us some idea of the number of missing controls. Table 5 of this report and line 14 of Table 3 summarize the Table 120 information. If there had been no missing cases, there would have been 4772 control cases, half chosen before and half after each COMP case. Knox calls these "Cohort One and Cohort Two". Actually, there are only 2642 controls, 1393 before plus 1249 after. This is a 45 percent missing rate (see Table 3, line 14). Thus, a missing data rate of over 60 percent for the COMP group becomes a rate of 70 percent for the control group (Table 3, line 14).

TABLE 4

Comparison of Cohort One and Cohort Two[a]

Variable	Means One	Means Two	Significance of Difference[b] (t)	Variances One	Variances Two	Significance of Differences (F)
Months employed	10.16	10.63	NS	53.29	54.32	NS
Number of arrests	6.73	7.18	NS	57.76	55.06	NS
Grade level	10.39	10.31	NS	3.76	3.72	NS
Age at first arrest	17.01	17.15	NS	19.62	22.47	NS
Number of Illinois commitments	.55	.57	NS	1.46	1.14	NS
Number of commitments in other jurisdictions	.09	.14	$p \leq .01$	0.212	0.270	NS
I.Q. score	98.33	98.00	NS	171.35	193.21	NS
Number of children	1.33	1.20	NS	19.45	2.34	$p \leq .001$
Reading score	71.00	70.70	NS	682.25	705.96	NS
Arithmetic score	59.04	61.77	$p \leq .01$	531.30	423.95	NS

[a]Sources: Knox, Table 121 page 398 and Table 120, page 397.

[b]The N for each variable is not given in Knox. These calculations assume an N of 1393 for Cohort One and 1249 for Cohort Two. Actually, there were probably missing data on these variables. This would make the N smaller, and would decrease the chance of finding significant differences in the means.

We know, then, that almost half of the control cases which were sought were not found. However, is there any evidence that these missing cases are systematically different from the control cases that were found? Yes, there is evidence that the control group is biased, and moreover, that it is biased in the variable under analysis - recidivism.

A standard way of measuring data reliability is a "split-half" reliability test.[1] In such a test, the sample is divided randomly into halves, and the two halves are compared. If any significant differences are found between the two halves, some systematic bias must be present in the study. The presence of Cohort One and Cohort Two allows us to do a variation of the split-half reliability test. The two halves were supposedly randomly chosen from the same population. If this is true, then they should not differ on recidivism, since recidivism is the variable being analyzed. Table 4 gives the means and standard deviations of the two "cohorts" on eleven variables.[2] Table 5 compares the two on the proportion who were recidivists.

If the control group had been randomly chosen, and missing cases also happened randomly, we would expect to find no difference between Cohort One and Two in the mean or the variance on any of the variables in Table 4. However, the means of number of commitments in other jurisdictions and of arithmetic score are significantly different, and so is the variance of number of children.

---

[1]See, for example, Fred N. Kerlinger, Foundations of Behavioral Research, New York: Holt, Rinehart and Winston, Inc. 1973, pages 445-455.

[2]As mentioned earlier, the first ten variables in Table 4 are not the best choice for comparison, since only months employed and grade level may affect recidivism, and then only if they are dichotomized as "12 months or less" for the former and "11 or fewer grades" for the latter (see Table 1). The variables in Knox's Table 121, from which Table 4 was calculated, were not dichotomized.

There is less than a one per cent chance that such different means would be found if the two parts of the control group were really randomly chosen.

The difference of the two cohorts in recidivism rate is even more interesting.[1] One cohort had a 16.7 per cent recidivism rate, and the other had 21.9 per cent. If the cohorts were randomly chosen, there should be no difference between the two. Table 5 gives the probability that a difference of .052 (.219 - .167) would be found if there were no real difference. It would be found fewer than five times out of 10,000. In other words, it is highly unlikely that the DOC cases Knox used to determine recidivism rates for the cost-benefit analysis are an unbiased control group.

The Cost-Benefit Analysis

Since it has been shown that "Cohort One" and "Cohort Two" are not an adequate control group, and since the cost-benefit analysis was based entirely on the assumption that they were a control group, the results of the cost-benefit analysis are meaningless. There are, however, a few additional problems with the cost-benefit analysis that deserve a brief mention.

1. The calculation of the benefit of averted correctional costs is based on the calculation of recidivism in the COMP group and the control group (Cohort One and Two). Recidivism is calculated by dividing the number of recidivists into the total N. However, the total N used for the control group calculation was the N not counting the missing cases, and the total N used for the COMP group calculation is the N counting

---

[1]Since recidivism is a dichotomy, a difference of proportions test was used.

Table 5

Difference in Cohort One and Cohort Two Recidivism

	Proportion of Recidivists
Cohort One	.167
Cohort Two	.219
Difference	.052
Significance of the difference	Z = -3.47 ($p < .0005$)

the missing cases. This results in an artificially low percent recidivism for COMP clients relative to the recidivism for the control group.

The correctly calculated figures appear in Table 6. Compare it to Knox's Table 120, page 397. Thirteen percent of the 1861 COMP state placement clients who were not missing were "in custody".[1] This should be compared to the 19 percent in custody of the control group 2642 Nonmissing exoffenders. This six percentage point difference is not large, especially when you consider that there is a five percentage point difference between the two halves of the control group, which are supposedly the same. (The total number of cases is so large that even a two percentage point difference would be statistically significant, if a significance test could be done. However, such tests cannot be done here, because of the bias present in the sample. This was discussed above.)

Knox starts with a very small difference, which is so unreliable that it cannot be considered a difference at all. Then he miscalculates the percent in custody for COMP clients, arriving at a figure of ten percent instead of thirteen percent. The percent in custody of the control group is also miscalculated, by dividing the control group into two halves, Cohort One and Two, and figuring the percentages separately for each half. This produces a 17 percent and a 22 percent figure, a difference of seven and twelve percent from the miscalculated COMP figure. These incorrect estimates of recidivism are later called the "lower bound" and the "upper bound" in the cost-benefit estimates. The actual figures should be 13 percent for COMP clients and 19 percent for the controls, and even this six percentage point difference is meaningless in light of the bias in both samples. Knox bases his entire cost-benefit analysis on these miscalculated and misinterpreted figures.

[1] Notice Knox's inconsistent definition of "in custody" discussed in footnote c of Table 4.

TABLE 6

COMP and Recidivism[a]

	COMP	Cohort One	Cohort Two	Combined Controls[b]
In Custody[c]	245(13%)	232(17%)	273(22%)	505(19%)
All Other	1616(87%)	1161(83%)	976(78%)	2137(81%)
Total	1861	1393	1249	2642
Missing	525(22%)	993(42%)	1137(48%)	2130(45%)
Total	2386	2386	2386	4772
Percent of Total COMP State Placement Clients[d]	58%	69%	72%	

[a]Source: Table 120, p.397 of Knox.

[b]Cohort One plus Cohort Two

[c]"In Custody" includes the following categories of Table 120: permanent assignment, full diagnostic, partial diagnostic, serving other sentence consecutively or concurrently, direct transfer in, declared parole violator, returned work release violator, returned parolee, escape, and escape from work release. All these are included in Knox's definition of recidivism on page 201, but the definition also includes other categories which are not included in Table 120. The reason they are not included is not given. They are the following: 90 day temporary, 60 day status, investigative detention, returned escapee, returned furlough escapee, mandatory release violator, bond violator, authorized absence defaulter, escape from institution, escape while on furlough, escape while on writ, escape while on bond.

[d]See Table 3, lines 9 and 14.

2. Another term of the cost-benefit formula is the benefit of taxes paid by employed COMP clients. This calculation is based on state COMP clients for whom placement is known. For over 50 per cent of them, it is not known (Table 3 line 12). It is reasonable to assume that the missing cases are more likely to have been unsuccessful in employment and recidivism than the others. In addition, knowledge of the number of dependents is necessary to figure taxes (see Knox's Appendix C). However, about 34 per cent of COMP clients are missing this information (see Table 1 in Knox). Therefore, the estimate of tax benefits from COMP clients is very unreliable.

3. Even if the estimate of tax benefits had been reasonably accurate, it was still not properly calculated. A proper calculation of benefits is the benefit of the program minus the benefit of no program. Some DOC exoffenders undoubtedly had jobs and contributed to the tax rolls, too. There is no tax benefit to the COMP program unless COMP clients are more likely to have jobs than non-COMP clients. This was not determined by the analysis.

4. The measurement of costs of the COMP program is based on the same N as the tax benefit estimates, with the consequent missing cases of over 50 per cent (Table 3, line 12). Also, the calculation of program cost is very strange. The formula on page 402 says that difference between the costs for placed clients and for all clients weighted by the proportion not placed, added to the cost per client served, equals the number of clients hired. This is not true. The first term in this equation, the weighted cost per client not placed, is claimed to be the "estimated placement cost" (page 402). Perhaps Knox means to weight the first term by the proportion of clients placed. At any rate, the origin of the measurement of "cost" in the cost-benefit analysis remains obscure.

5. Averted correctional costs were not calculated by using the same sample as was used to estimate recidivism.

The estimate of recidivism, based on data with over 60 per cent missing (Table 3, line 9), was used to calculate averted correctional costs of another group with over 30 per cent missing (Table 3, line 13). Whether the recidivism of the first group can be assumed to be true of the second group is questionable.

6. There are other problems with the cost-benefit analysis. We will not discuss them here, but many are mentioned in the footnotes to the tables in this report.

Conclusion

Any one of the above problems would have been enough to invalidate the conclusions derived from the cost-benefit analysis. The implication of this review is, then, the following: The cost-benefit analysis in the Knox report is based on incorrect analysis of inadequate data, and the conclusions from this analysis should be disregarded. The Knox study was an ambitious attempt to do a controlled evaluation, but the attempt failed.

Appendix A

MEMORANDUM

Illinois Law Enforcement Commission
120 South Riverside Plaza
Chicago, Illinois 60606

To: J. David Coldren Date: May 25, 1978

From: Karen P. Smith KPS Copies: Ruth Perrin

Subject: Cybersystem Research, Inc.'s Design for an Evaluation of the Comprehensive Offender Manpower Program

I have read the Evaluation Design of the Comprehensive Offender Manpower Program (COMP) and asked Ruth Perrin to do so as well. The comments which follow are a synthesis of Ruth's and my thoughts on the design. In brief, we both believe that this document fails as an evaluation design because it does not detail an effective way of judging the success or failure of the COMP program.

Specifically, the evaluation design suffers from the following problems:

1. Its first promise is to "provide a replication and refinement of comparative cost-benefit analysis and cross-sectional cost-effectiveness analysis," yet the design itself does not reveal the means by which this will be accomplished. The suspicion is that the evaluator had used cost-benefit analysis before, but had not figured out how to apply it in the COMP setting, and therefore glossed over this critical point in the design. The cost-effectiveness component, by the way, is the only one in the design which has true evaluative potential.

2. The design proposes to do "a comprehensive statistical analysis of client processing data," however, the analysis described will not serve to evaluate the COMP program's effectiveness. According to the design, statistical analysis will go only so far as to identify the "demographic characteristics of the serviced groups" to discover which ones may need "special attention" in order to get jobs. The problem is that no control group, i.e., non-"serviced" ex-offenders, is anticipated by the design. Because of this deficiency no conclusions can be drawn either as to the effect of "demographic factors" (undefined by the evaluator) or the effectiveness of the COMP program in securing jobs for ex-offenders.

3. The evaluation is to include "a prison training tracking component." That is, the design proposes to follow those ex-offenders who have received job training in prison and who have been "serviced" by COMP, to determine whether prison training is effective in securing employment after imprisonment. Once again, the absence of controls will make such analysis inconclusive: there is no way to determine whether prison training is effective unless its effect is isolated from COMP "servicing," and vice versa.

4. The final promise contained in this evaluation design is to develop "prediction equations and technical assistance" for the Parole Board to use in granting or denying Early Release to prisoners. This component, vaguely described as "using discriminant analysis" to predict "who will succeed and/or fail," is nonetheless given "priority in terms of completion." It will be interesting to see which components are completed when, and the forms that they take.

In fairness, it should be noted that the task of evaluating a 20-project, statewide ex-offender placement program is a massive one. I am not sure what a good design would have entailed, but the one put together by Cybersystem is flawed because it purports to do the massive job using undersized tools. A better approach would have been to acknowledge the problems, and then cut the evaluation's objectives down to manageable proportions.

Appendix B

Internal Validity

Since Knox mainly uses available computerized data for his evaluation, he devoted a lot of space in the report to his testing of the quality of the data. This is admirable. However, Knox also claims to have analyzed the internal validity of the data when he, in fact, did not. This may be quite misleading to the reader.

Internal validity refers to the extent to which error variance is minimized in the study, that is, the extent to which all possible sources of bias are reduced.[1] A discussion of internal validity should therefore cover such topics as whether controlled conditions were used, and the reliability of the measures used. A discussion of internal validity would have been very enlightening in this case, since, as this review shows, bias is a serious problem throughout the evaluation.

Instead, under the heading "Validity of the Data", Knox discusses whether or not there are coding errors (pages 197-199.) Coding errors are, of course, important, but they have very little to do with validity.

---

[1]Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1966, pages 5, 23-24.

Appendix C

Interpretation of Table 74

Table 74 on page 297 gives the Chi-square significance and percent differences of the relationships between "relapse rates" and 14 other variables for each of the 15 placement programs. The most interesting thing about this table is that hardly any of the differences are significant, and those relationships that are significant tend to be present in one placement program.

At the five percent level of significance, we would expect five out of 100 relationships to be significant, even if the actual association between the variables were zero. In other words, if you run enough tables, you are likely to find some significant relationships just by chance. Knox ran 175 tables for the results in Table 74 (14 variables times 15 programs minus 35 with no information.) Out of 175 tables, some will be significant just by chance. The number of significant findings for each variable is listed in Table C. Altogether, there were fifteen. Only two variables had as many as three significant relationships with recidivism. Knox, however, concludes that variables with even one significant finding "differentiate being returned to prison." A more likely conclusion would be that none of the variables makes a significant difference.

A more interesting finding in Table 74 is that eight of the fifteen programs had no significant findings, three had one, three had two, but the other one had six. Six significant findings is quite at odds with the results for the other programs, and seems very unusual. Instead of trusting these six findings, Knox should question whether there is some sort of bias in the data-collection practices of this program (Operation DARE-Chicago) that might be producing the strange results in Table 74. It is possible that this unusually high number of significant findings for Operation DARE could be due to DARE's higher N. Since Table 74 does not report the N's, this is impossible to determine.

Table C

Significance of Relationships between Recidivism and Other Variables

Variable	Number of significant differences	Programs with Available Data
Marital status	1	14
Military status	0	14
Education	1	15
Recent wage	0	15
Employment history	2	15
Previously fired	0	13
Employed at termination	2	15
Drug involvement	1	13
Criminal convictions	3	8
Drivers License	1	10
Alcohol Problem	0	7
Drug Problem	0	12
Never Placed	1	15
In Custody over 12 Months	3	9
Total	15	175

Source: Table 74, page 297.

END