Posted 12/11/10

FASTER, CHEAPER, WORSE

Rehabilitation doesn't lend itself to shortcuts.
Neither does research and evaluation.

By Julius (Jay) Wachtel. Is "corrections" a non-sequitur? No, insists NIJ. Its landmark 1997 report, "Preventing Crime: What Works, What Doesn't, What's Promising," argued that carefully designed and appropriately targeted programs of sufficient dosage and duration can indeed rehabilitate. Intensive, theory based "cognitive and behavioral treatments" were particularly recommended for high-risk populations.

That's exactly what Project Greenlight offered. Developed by the Vera Institute of Justice and conducted in New York between February 2003 and February 2004, it applied a "cognitive-behavioral" approach to mitigate personality traits associated with offending such as impulsivity, antisocial attitudes and drug use. Inmates would participate in therapeutic sessions, receive housing and employment assistance, and interact with parole agents and social workers before release. Ex-offenders would leave with detailed, step-by-step plans to help them successfully reintegrate into the community.

As usual, funding issues butted in. What was intended to be a three-year pilot project was cut back to one year. While that didn't affect participants, to increase their numbers treatment was slashed to eight weeks from a design length of four to six months. Class sizes were also increased three-fold, from the recommended eight to ten participants to twenty-six. Just like elsewhere in government, notions of "faster, better, cheaper" had clearly taken hold.

Experiments normally include an experimental group and one or more control groups that are virtually identical in all respects but receive no treatment or "intervention." Because the Department of Corrections intended to house the program in a male-only, minimum-security facility in New York City, Project Greenlight's experimental group (GL) was comprised of 344 low-risk inmates who originated from (and would be released to) New York City. There were two control groups. One, TSP, included 278 low-risk inmates, also from New York City, who would be housed at the same facility and treated with the department's five-week Transitional Services Program. A second control group, UPS, included 113 low-risk inmates from outside New York City who would be released from upstate prisons without benefit of a program.

To assure that any differences in outcomes between groups are not due to differences in their composition, experimental subjects are normally picked at random and assigned to groups one at a time. But that's not what happened with Greenlight. According to the program's published report correctional officials at first assigned inmates to GL and TSP in large batches, rather than one-by-one. While investigators eventually regained some control, in the end they conceded that the design was only quasi-experimental. However, they declared it was sufficiently robust to eliminate the possibility that the groups were systematically different from the start.

Mean survival time (weeks)	
Total Arrests	
Greenlight	70.62
TSP	79.28
Upstate	78.54
Felony Arrests	
Greenlight	84.55
TSP	90.93
Upstate	91.17
Parole Revocations	
Greenlight	79.51
TSP	85.83
Upstate	88.84

Outcomes were measured one year later. Surprisingly, GL participants seemed substantially worse off. Thirty-one percent of the experimental subjects had been rearrested, compared with 22 percent of TSP participants and 24 percent of those in the untreated UPS group. GL's also "survived" for substantially briefer periods before arrest.

It's well accepted that the best predictor of future offending is past offending. That's consistent with Greenlight data, which indicated that the more serious one's criminal record the greater the likelihood of arrest after release (coefficients with asterisks denote statistical significance, the more the greater.) But study group also seemed to matter, with Greenlight participants forty-one percent more likely to fail than those treated with TSP. (Similar though statistically non-significant results were reported when comparing GL to Upstate.)

Greenlight and TSP		
Age at release	-.04	.96***
Education	-.02	.98
Race/Ethnicity		
NH White/other	---	---
NH Black	.05	1.05
Hispanic	-.12	.89
Prior arrests	.05	1.05***
Primary offense		
Robbery	---	---
Violent	-.27	.77
Drugs	.49	1.63**
Property	.40	1.49*
Other	-.51	.60^
Substance abuse		
None	---	---
Alcohol only	.52	1.69*
Drugs	.15	1.16
Alcohol and drugs	.15	1.16
Age at first arrest	-.02	.98
Study Group	.34	1.41**

Assuming that the groups were equivalent as to all important characteristics before treatment (we'll come back to that later), investigators surmised that one or more aspects of Greenlight was making things worse. They speculated about a "mismatch" between the program, which was designed for high-risk offenders, and the low-risk nature of those actually treated. Other likely suspects include GL's highly abbreviated format, its departure from the original design, poor implementation, and subpar performance by case managers.

Fast-forward to November 2011 when a Project Greenlight update reported outcomes after thirty months. Participants were coded for risk of recidivism, an index comprised of criminal history and other measures. While members of the experimental (GL) group did more poorly overall than those in TSP and UPS, the gap between GL and TSP was statistically insignificant and far outweighed by the gap between both programs and UPS, whose participants fared well while receiving no treatment at all. Low and medium-risk inmates did exceptionally well in UPS, while those at medium and high-risk did especially poorly in GL. Actually, low-risk inmates tended to succeed in each program, with those assigned to GL actually doing considerably better than participants in TSP but falling somewhat short of the untreated Upstate group.

Exhibit 1. Percent of Participants Without a Rearrest After 30 Months			
	Greenlight	**TSP**	**UPS**
Risk Level	(N = 345)	(N = 278)	(N = 113)
Total Sample	47.5	51.8	66.4 **
Low-Risk	80.4	70.0 ^	86.4
Medium-Risk	44.0	51.7	69.0 **
High-Risk	23.7	33.8 *	32.1

Note: All comparisons of statistical significance are with the GL group. ^ p < .10; * p < .05; ** p < .01

Why did GL succeed with low-risk inmates? Researchers guessed that their personal characteristics (e.g., attention span, cognitive and social skills) were most compatible with the program's intensity and its compressed format. As for the relative success of the untreated UPS sample, it might reflect the advantage of not unduly upsetting inmates by coercively transferring and programming them shortly before setting them free.

Complex after-the-fact explanations are inherently untrustworthy. What if the presumed effects were artifacts of biased assignment? Indeed, the study's own data suggests that the groups were different from the start.

Variables	GL (N=339)	TSP (N=274)	Upstate (N=112)	Total (N=725)	Sig.
Total Arrests					ns
Mean	8.65	7.81	6.67	8.02	
Std. Deviation	9.1	8.1	9.7	8.8	
Total Convictions					ns
Mean	5.94	5.31	4.33	5.45	
Std. Deviation	7.6	6.8	8.5	7.5	
Felony Arrests					ns
Mean	4.57	4.12	3.57	4.25	
Std. Deviation	4.7	3.9	4.3	4.4	
Felony Convictions					ns
Mean	1.76	1.72	1.53	1.71	
Std. Deviation	1.6	1.6	1.6	1.6	
Misdemeanor Arrests					ns
Mean	3.11	2.63	2.09	2.77	
Std. Deviation	5.7	5.2	7.4	5.8	
Misdemeanor Convictions					ns
Mean	4.18	3.59	2.80	3.75	
Std. Deviation	6.8	6.1	7.6	6.7	

Each arrest and conviction variable was at its highest level in Greenlight and at its lowest in the untreated Upstate group, with TSP holding the middle ground. Some of the mean differences appear substantial. So the implications are clear: since the GL group had more hardheads, poor results were inevitable. On the other hand, as the authors pointed out, none of the differences between means reached significance (that's probably because sample sizes were so small and the fluctuations in scores, measured by standard deviation, so large.) In any event, when nonrandom methods are used to form groups, one cannot assume that participants come from the same population, so statistical significance is meaningless. A more parsimonious interpretation is that the GL group's bias in the direction of more serious criminal records increased recidivism. Greenlighters seemed least amenable to treatment because they were the most criminally inclined. Upstaters fared relatively well because they were the least. Speculation that Greenlight itself had a criminogenic effect remains just that.

Alas, the conceit that short-term rehabilitative attempts can influence post-release outcomes is nothing new. No matter how carefully designed a program might be, convicts who spend years in prison learning all the wrong lessons are unlikely to be transformed in two months. Still, in an era of shrinking budgets there is a lot of pressure to devise solutions that are better and cheaper than simply locking people up. In "Economical Crime Control," the lead article in the November/December 2011 ASC newsletter, Philip Cook and Jens Ludwig argue for reprogramming $12 billion a year from prisons to early childhood education and to initiatives that address the "social-cognitive skill deficits" of young persons in trouble with the law.

Effective community-based solutions, though, can be be very expensive. Deinstitutionalization left us with the worst of both worlds: mentally ill persons who are untreated *and* homeless. To do better with criminal offenders would require far heavier investments in research and evaluation than bean-counters would likely tolerate. "Corrections" may not be a non-sequitur, but "economical" crime control most certainly is.