CHAPTER SIX

Evaluation in Corrections: 'Nothing works' versus 'What works'

David Riley

The efficacy of human service provision in a variety of contexts has been a focus of debate for more than fifty years. Eysenck's attack on psychoanalysis in the 1950s and early 1960s prompted a raft of publications in which both sides of the debate presented 'evidence' in support of their respective positions.[1] However, despite the prestige of the protagonists, no consensus emerged.

What did occur, though, was greater scrutiny of other areas of service provision, such as the efficacy of social work intervention. Similarly, intervening with offenders has been extensively analysed, but in spite of this there is still sufficient ambiguity for even well-qualified practitioners to make claims in Court, for example, that intervening with child sex offenders is ineffective. Even more remarkable was the wide-ranging denunciation of treatment in Corrections put forward by David Farabee in his 2005 book *Rethinking Rehabilitation: Why Can't We Reform Our Criminals?*

While sociologists in particular had expressed doubts about the efficacy of Correctional treatment, it was Robert Martinson who led the criticism in 1974 with his article 'What works: Questions and answers about prison reform', published in the conservative journal *The Public Interest.* In this article he summarised 231 studies published between 1945 and 1967 and on the basis of his analysis concluded that 'with few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism'. As Smith, Gendreau, and Swartz noted, his pronouncement labelled

1 Eysenck 1953.

'nothing works' became an instant cliché and was eagerly embraced by many policy makers and academics, particularly in the United States, where it was used to challenge the role of prisons and rehabilitation in general.[2] This cliché struck a nerve in the conservative social climate of the time, and Martinson's 'second thoughts', which were published in a near retraction of his earlier pronouncement had little impact, received comparatively little attention and certainly failed to effect policy change.[3]

While the rehabilitative ideal was kept alive by such luminaries as Paul Gendreau and Robert Ross, who published their 'Bibliotherapy for cynics' in 1979 and their more in-depth narrative review of treatment outcome studies that appeared in the *Justice Quarterly* in 1987, they were almost lone voices in a pessimistic environment.

Although, perhaps predictably, no consensus developed regarding the efficacy of treating offenders, what the debate did achieve was an acceptance by practitioners that they needed to document carefully the outcomes of their efforts. Furthermore, the professional journals increasingly began to publish evaluations of offender programmes. Predictably, the results reported in the studies were anything but uniform, thereby fuelling rather than diminishing the increasingly acrimonious debate.

Such a state of affairs was almost inevitable, as differing approaches applied to highly variable populations, utilising differing criteria of success, are bound to throw up results as diverse as the methodologies themselves.

In addition, there is the potential for a much more sinister undermining of apparently promising approaches. This was eloquently described by Michael Gottfredson in his seminal paper 'Treatment destruction techniques', which appeared in 1979 in the *Journal of Research in Crime and Delinquency*. Essentially, Gottfredson made a scholarly examination of the basis on which research reviews had failed to support an effect for rehabilitative efforts. He was able to identify a number of strategies used by some authors that could be applied to any evaluation attesting to treatment efficacy, which no matter how

2 Smith, Gendreau & Swartz 2009.

3 Martinson 1979.

sound the research or unequivocal the outcome, called into doubt their success.

However, Andrews and Bonta point out that the strategies identified by Gottfredson as undermining favourable treatment reports may in fact do the opposite and support a positive outcome.[4] For example, stressing the criterion problem of unreliability in the measurement of outcome variables would be likely to diminish rather than improve treatment effects. Likewise, discounting the underlying theory on which the treatment was predicated would be to decrease, not increase, the degree of positive change observed.

The problem with such narrative reviews was, to a degree, overcome by applying the technique of meta-analysis to treatment outcome studies. Meta-analysis is a way of empirically evaluating the results of multiple studies and examining the effects of the various treatments with respect to other variables, such as treatment modality, provider characteristics and participant type.

The starting point for such a 'systematic review' is to develop firm, objective criteria for inclusion of studies during a specified period and a specification of the data sources that will be examined to derive the pool of evaluations to be analysed. Additionally, a number of variables are specified that are thought to have a bearing on the matter being investigated, and these are objectively defined so that they can be coded from the descriptions in the studies and entered into the data set for the analysis. The analysis itself is a way of statistically aggregating the results of all these studies and examining the impact that various variables have on the outcome.

Notable pioneering work in the area of meta-analysis was carried out by Mark Lipsey who in 1989 published a meta-analytic review of nearly 450 correctional outcome studies and on the basis of his analysis concluded that, on average, treatment reduced recidivism by approximately ten percent. Furthermore, he was able to show significant differences in treatment outcome depending on various methodological and procedural variables pertaining to these evaluations.

It seemed that now, for the first time, a value-free objective method of assessing treatment efficacy in Corrections was available. However, later developments were to dismiss that notion.

4 Andrews & Bonta 2010.

It soon became apparent that different meta-analyses drawing on the same set of source studies were capable of producing differing results. Whitehead and Lab conducted a systematic review of treatment outcomes and presented a rather qualified conclusion as to the effectiveness of treatment.[5] When Andrews, Zinger *et al* re-evaluated those studies, employing what they termed a 'clinically relevant and psychologically informed' methodology, their result revealed quite a different picture, indicating that certain approaches applied to higher risk offenders yielded good outcomes, whereas other approaches, which were less targeted, were considerably less successful.[6] It was this paper, published in the journal *Criminology*, that formed the basis of the risk, needs and responsivity principles which underpin effective interventions for offenders and form a cornerstone of the *Psychology of Criminal Conduct* authored by Andrews and Bonta, which is now in its fifth edition.[7]

The findings of the Andrews, Zinger *et al* meta-analysis have been reaffirmed and by the start of the new millennium, McGuire was able to report that there were now more than twenty major systematic reviews of Correctional treatment that together sampled a very large number of outcome studies.[8] This body of work attested to the value of intervening within Correctional environments and, further, that the gains achieved by modern psychological approaches targeting criminogenic factors in high risk groups yielded results comparable to such medical procedures as AZT in the treatment of Aids and coronary bypass surgery for myocardial infarction.[9]

McGuire's finding was further reaffirmed in 2008 in a 'Review of systematic reviews' (a meta-analysis of the meta-analyses) by Christopher Lowencamp and colleagues.

It is the opinion of this author that treatment efficacy in Corrections is no longer a matter of dispute, although it is acknowledged that the practice of evaluating treatment outcomes is beset with major problems, and I have yet to read an evaluation that is so methodologically robust as to defy criticism.

---

5 Whitehead & Lab 1989.
6 Andrews *et al* 1990.
7 Andrews & Bonta 2010.
8 McGuire 2002.
9 Marshall & McGuire 2003.

It is, in fact, impossible to evaluate any Correctional programme in the same terms as the highest standards demanded in the trial of some medical procedures or pharmaceuticals, for example. Such an evaluation would involve the random assignment of subjects to treatment and control conditions, and neither those receiving the treatment, nor those providing it, or even those who evaluated the outcome data, would be aware of who had received the treatment and who had not.

While some may deprecate the value of such a 'double blind' approach to treatment evaluation in Corrections, and others would question such a methodology on ethical grounds, one has to remember the powerful impact of the 'Hawthorne effect' discovered by Elton Mayo in his pioneering application of industrial psychology in the General Electric Corporation plant. Likewise, the 'demand characteristics' of the experimental situation, so eloquently described by Martin Orne *et al* recently,[10] and the 'Experimenter effect' of Robert Rosenthal must be taken into consideration.[11] Additionally, there is the intriguing finding in some meta-analyses that merely nailing up a sign 'TREATMENT UNIT' in some parts of a prison can contribute to a small but statistically significant reduction in recidivism.

It is now well accepted that treatment evaluations should involve some objectively defined outcome measure (ideally something like reconviction, re-arrest, or at least some robust intermediate measure of change) for both the treated and untreated groups over a uniform time period. Studies, however, differ in the criteria by which outcome is judged and may involve a wide variety of dependent variables including parole violation, re-arrest, reconviction, re-imprisonment, re-commission of a specified offence, seriousness, rate of reoffending, time to reoffend, and measures of 'before and after' offending history. Given such a wide array of potential outcome variables, deciding which items should be nominated in advance is not straightforward. For example, the evaluation of the intensive treatment programme at the Canada's Regional Psychiatric Centre (Prairies) in Saskatchewan for high risk, violent and personality-disordered inmates has indicated that, while blanket measures of re-arrest, reconviction and re-imprisonment failed to distinguish between the recipients of therapy and their

10 Orne 1962.
11 Rosenthal 1966.

non-treated controls, more detailed analysis of the type and seriousness of reoffending did suggest the presence of genuine treatment gains.[12]

A more basic problem relating to treatment evaluation pertains to exactly what is being measured. Typically, the assumption is, given treated and non-treated offenders are followed up for a specific period, that any differences (or lack of them) are attributable to a treatment programme. This may not be the case, as there is increasing evidence that pilot programmes, for example, in which a well-structured intervention is provided by well-trained and closely supervised practitioners to groups of offenders selected on the basis of risk and criminogenic factors, may achieve anticipated gains, yet when exactly the same programme is rolled out more widely, the anticipated benefits do not accrue. This was the case with the *Reasoning and Rehabilitation* cognitive skills programme in the English Corrections system, where its widespread implementation failed to deliver the anticipated benefits in terms of reductions in reoffending. This was despite highly promising results from the pilot, in which expected reductions were observed, and in which the principles of risk and need were clearly demonstrated to operate.[13]

While the exact reasons for the failure of this programme remain unclear, it is reasonable to speculate that critical programme integrity variables may not have been addressed, and it is also not clear whether the subsequent evaluation was examining the impact of the programme itself, or was inadvertently documenting deficiencies in implementation; an area described by Gendreau, Goggin and Smith as the 'forgotten issue' in Correctional programme evaluations.[14]

Many evaluations of offender treatment report results for those individuals who complete a treatment programme and sometimes (but not invariably) indicate the attrition rate. Few treatment evaluations take the more hard-nosed Cochrane approach, whereby the outcomes of all those assigned to the treatment group are incorporated into the analysis.[15] There is a growing body of evidence that treatment dropouts may in fact have poorer results than they might have had if they had never entered the treatment programme in the first place. Certainly

12 Wong 2007.
13 Falshaw, Friendship, Travis & Nugent 2003.
14 Gendreau, Goggin & Smith 1999.
15 Higgins & Green 2008.

those opting out of treatment tend to do more poorly than those who complete it. While there appears to be general acknowledgement of this issue, it is by no means routine to include such cases in the final results, and it could be argued that treatment studies that exclude dropouts from the analysis could be biasing the outcome in favour of a positive result for the treated group to a significant degree.

Finally, there remains the issue of what should be the benchmark against which treatment outcomes should be compared. This is usually achieved by way of a 'comparable' control group, which usually involves the selection of individuals who have not received treatment but who exhibit similar levels of risk as assessed by matching known risk factors.[16]

However, while using apparently equivalent control groups to benchmark treatment gains is superficially (and perhaps even seductively) persuasive, the devil may lie in the detail. Whenever treatment is voluntary, it is highly likely that significant (and largely unknown) processes are in operation that cause some individuals to enter programmes while others do not. This could conceivably result in highly anomalous findings, such as those reported in the treatment of the highly selected, high risk (and often psychopathic) population at the maximum-security mental health facility in Penetanguishene in Ontario, Canada. While the programme was one that would not in the modern era be usually anticipated to achieve positive treatment gains (since it was a therapeutic milieu where the rules and mores were largely established by the inmates themselves), later evaluation indicated that psychopathic individuals who completed or participated in this programme did significantly more poorly than those who did not.[17] While one may conclude that participating in this programme made these offenders worse, and the authors subsequently suggested some mechanisms that may have contributed to this, an equally plausible interpretation may be that certain individuals (and perhaps even the more deviant or entrenchedly criminal) do not enter the programme because of a desire to bring about change in their life, but as a form of impression management, seeking to give evidence to such bodies as the Parole Board of genuine change. This idea finds

16 Or in slightly more sophisticated studies, achieving equivalence on some form of risk measure such as the LSI-R or Static-99.

17 Quinsey, Harris, Rice & Cormier 2006.

support from the investigation undertaken by Seto and Barbaree, in which they evaluated the outcome for individuals who had taken part in a sex offender treatment programme and compared the results for those who were above and below the median on measures of treatment progress and those who were above and below the median on measures of psychopathy as assessed by the PCL-R.[18] Initial results from this investigation appeared to indicate that those who were above the average level on the PCL-R (but not necessarily reaching the cut-off for a diagnosis of psychopathy), and who rated above the average on the measure of treatment progress, had four times the sexual reoffending rate of the other three groups. While this finding was subsequently significantly qualified, if not totally retracted, in a further paper by Barbaree (where Seto apparently declined to be named as a co-author), the result does have worrying implications that extraneous factors may be impacting on treatment outcome and, further, that such factors may be extremely difficult to control.[19]

The eminent empiricist DT Campbell, has helpfully distinguished between internal and external validity in research and evaluation.[20] Internal validity pertains to the methodological purity and robustness of design in a scientific sense, whereas external validity represents the degree to which any result may be generalised to, and have significance for, phenomena external to the research setting. The conundrum is that internal and external validity are inversely related, and to the extent that one seeks a result that is relevant to and generalisable about events external to the research situation, the confidence one can have about the results in a methodological sense must decrease.

Despite these many complexities and conundrums, it remains this author's view that the debate about the big question 'does treatment work?' has genuinely been answered in the affirmative. This is because the plethora of evaluations, and burgeoning numbers of systematic reviews of those evaluations, consistently point to a significant degree of positive change in the direction of reduced recidivism. Additionally, increasingly sophisticated meta-analyses have cast a degree of light on what might work, under which conditions, with which offenders. In other words, the position taken by Martinson at the beginning of this

18 Seto & Barbaree 1999.
19 Barbaree 2005.
20 Campbell & Stanley 1963.

debate, and still taken by the more extreme reactionaries, that 'nothing works,' is no longer tenable.

It is recommended that practitioners heed the lessons learned from the meta-analyses about programme type, provider quality, programme integrity, matching of treatment modality to participant level and appropriate targeting of criminogenic factors, and cooperate with this research endeavour by maintaining records that permit detailed assessments of outcomes over time. Practitioners need to learn from the lessons of others, and their own experience, that within any global phenomenon there are a host of questions we may seek to answer. These include questions such as:

- Was the approach or programme implemented effectively and to a high standard?
- Was the programme being delivered as designed?
- Was the programme being delivered to a high standard of therapeutic integrity?
- Have the providers received adequate training, support and supervision?
- What do intermediate measures of change reveal about client progress?
- Has the programme itself engaged the recipients, and is it perceived as salient to their future functioning?
- What specific outcome measures may be sensitive enough to document treatment change?
- How might risk variables, motivational factors, and personality attributes be better controlled to ensure equivalence between treatment and control groups?
- Are some providers who show high levels of integrity in service provision achieving better results than others?
- What observed within-group variations would give a clue as to how treatment and eventual effectiveness may be better achieved?

- What contextual factors (such as reintegrative needs) are associated with positive outcome and how may these be better managed?

The above list is far from exhaustive, but it illustrates that a more profitable way forward for future research might be to more tightly control the variables in the evaluation settings, thereby enhancing the internal validity of the exercise. This in turn has the potential to contribute to a larger knowledge base, which can be incrementally enhanced, enabling organisations and jurisdictions to steadily improve the effectiveness of their interventions, rather than too quickly seeking to answer the ultimate question, 'does it work?' Instead of asking 'does it work?', they should first ask, 'how can we fine-tune our strategies to make them work?'

References

Andrews DA & Bonta J (2010). *The Psychology of Criminal Conduct* (5th edition), Cincinatti OH, Anderson.

Andrews DA, Zinger I, Hoge RD, Bonta J, Gendreau P & Cullen FT (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28(3):369–404.

Barbaree HE (2005). Psychopathy, treatment behaviour and recidivism; An extended follow up of Seto and Barbaree, *Journal of Interpersonal Violence*, 20, 1115–31.

Campbell DT & Stanley JC (1963). *Experimental and Quasi-experimental designs for Research*. Chicago, IL: Rand McNalley.

Eysenck HJ (1953). *Uses and Abuses of Psychology*, Middlesex, UK: Penguin.

Falshaw L, Friendship C, Travers R & Nugent F (2003). *Searching for what works: An evaluation of cognitive skills programmes.* Home Office Research Findings Report, London.

Farabee D (2005). *Rethinking Rehabilitation: Why Can't We Reform our Criminals?* Washington DC: AEI Press.

Gendreau P, Goggin C & Smith P (1999). The forgotten issue in effective correctional treatment: Programme implementation. *International journal of Offender Therapy and Comparative Criminology*, 43, 180–87.

Gendreau P & Ross RR (1979). Effective correctional treatment: Bibliotherapy for cynics, *Crime and Delinquency*, 25, 463–89.

—— (1987). Revivication of rehabilitation: Evidence from the 80s, *Justice Quarterly*, 4.

Gottfredson MR (1979). Treatment destruction techniques, *Journal of Research in Crime and Delinquency*, 16, 39–54.

Higgins JPT & Green S (eds) (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, England and Hoboken, NJ: Wiley-Blackwell.

Lipsey MW (1989). The efficacy of intervention for juvenile delinquency. Paper presented to the Conference of American Society of Criminology, Nevada.

Marshall WL & McGuire J (2003). Effect size in the treatment of sexual offenders. *International Journal of Offender Therapy and Comparative Criminology,* 46, 653–63.

Martinson R (1974). What works: Questions and answers about penal reform, *The Public Interest*, 35, 22–54.

—— (1979). New findings, new views: A note of caution regarding sentencing reform, *Hofstra Law Review* 7, 243–58.

McGuire J (Ed) (2002). *Offender Rehabilitation and Treatment: Effective Programs and Policies to Reduce Reoffending*. Chichester, UK: Wiley.

Orne MT (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–83.

Quinsey VL, Harris GT, Rice ME & Cormier CA (2006). *Violent Offenders, Appraising and Managing Risk* (2nd edition) Washington DC, American Psychological Association.

Rosenthal R (1966). *The Experimenter Effect in Behavioural Research*. New York: Appleton.

Seto MC & Barbaree HE (1999). Psychopathy, treatment behaviour and recidivism. *Journal of Interpersonal Violence*, 14, 1235–48.

Smith P, Gendreau P & Swartz K (2009). Validating the principles of Effective Intervention: A systematic review of the contribution of meta-analysis in the field of corrections. *Victims and Offenders*, 4, 148–69.

Whitehead J & Lab S (1989). A meta-analysis of juvenile correctional treatment, *Journal of Research in Crime and Delinquency*, 26, 276–95.

Wong S (2007). Treatment of violence-prone forensic patients: It works! Paper presented to the Third International congress on Psychology and Law, Adelaide.