

# The University of Chicago Law School Roundtable

---

Volume 2 | Issue 2

Article 6

---

1-1-1995

## Evaluating the Impact of Community Penalties

George Mair

Follow this and additional works at: <http://chicagounbound.uchicago.edu/roundtable>

---

### Recommended Citation

Mair, George (1995) "Evaluating the Impact of Community Penalties," *The University of Chicago Law School Roundtable*: Vol. 2: Iss. 2, Article 6.

Available at: <http://chicagounbound.uchicago.edu/roundtable/vol2/iss2/6>

This Article is brought to you for free and open access by Chicago Unbound. It has been accepted for inclusion in The University of Chicago Law School Roundtable by an authorized administrator of Chicago Unbound. For more information, please contact [unbound@law.uchicago.edu](mailto:unbound@law.uchicago.edu).

# Evaluating the Impact of Community Penalties

GEORGE MAIR

For much of the twentieth century, many criminal justice policymakers and criminologists have been obsessed with the search for viable alternatives to custodial sentences. In England and Wales, finding alternatives to custody has arguably been the single most important penal policy issue since the end of the Second World War. Since Parliament passed the 1948 Criminal Justice Act, penal policy has been driven by the need to respond to increases in the prison population that have led to serious problems of overcrowding and control. Policymakers and criminologists have laboured to discover an alternative to custody that actually works. But this search is fundamentally misconceived—a point that is only rarely noted and even then, often ignored.

This is not to deny that the key question is whether community penalties or intermediate punishments actually “work.” Despite (or perhaps because of) the many years of frenetic research during the 1970s and 1980s,<sup>1</sup> there are no easy answers to this question. This Article revisits the issue of community penalties and focuses on how best to evaluate them. It seems to me that this is a question central to the Symposium.

This Article is organized into five parts. Section I discusses the work of Robert Martinson, whose legacy, “nothing works,”<sup>2</sup> unfortunately still dominates much of the debate on the effectiveness of community sentences. Section

---

George Mair is the principal research officer for the Home Office Research & Planning Unit in London, England.

1. Stephen Stanley and Mary Baginsky, *Alternatives to Prison: An Examination of Non-Custodial Sentencing of Offenders* (Peter Owen, 1984); John Pointing, ed, *Alternatives to Custody* (Basil Blackwell, 1986); Antony A. Vass, *Alternatives to Prison: Punishment, Custody and the Community* (Sage, 1990); Helsinki Inst for Crime Prevention and Control, *Alternatives to Custodial Sanctions* (1988); Josine Junger-Tas, *Alternatives to Prison Sentences: Experiences and Developments* (Kugler, 1994); Norman Bishop, *Non-Custodial Alternatives in Europe* (Helsinki Institute for Crime Prevention and Control, 1988). For a worldwide perspective, see Ugljesa Zvekic, ed, *Alternatives to Imprisonment in Comparative Perspective* (Nelson-Hall, 1994); Ugljesa Zvekic and Anna Alvazzi del Frate, eds, *Alternatives to Imprisonment in Comparative Perspective: Bibliography* (Nelson-Hall, 1994).

2. Robert Martinson, *What Works?—Questions and Answers about Prison Reform*, 35 *Pub Interest* 22 (1974).

II examines the relatively recent interest in meta-analysis and the “nothing works” debate. Section III makes some general points relating to evaluation. Section IV describes a recent major evaluative study of an Intensive Probation (IP) initiative<sup>3</sup> in England and Wales. Section V sets out general lessons from the IP evaluation for the evaluation of community penalties in general.

It may be helpful to make three preliminary points. These points are only loosely connected to the topic of this Article, but they constitute much of the necessary background, and, to a considerable degree, they frame what follows.

First, my views are coloured by my position as principal research officer of the Home Office Research and Planning Unit. The Unit conducts research for the government and is indeed part of the state apparatus. My approach to evaluation therefore focuses on the needs and priorities of policymakers. Second, the idea of alternatives to prison is now officially dead in England and Wales. There was always uneasiness about an idea implying prison was the only appropriate sentence for a criminal to receive. The 1991 Criminal Justice Act eradicated the idea of alternatives to custody by introducing a “just desserts” approach to sentencing. Although this approach was somewhat diluted by section 66 of the 1993 Criminal Justice Act, it remains official policy. Third, since 1984 when the Statement of National Objectives and Priorities<sup>4</sup> was introduced, the probation service in England and Wales has been in a state of almost constant flux. A service that must react to new—and what tend to be seen as threatening—policy developments is unlikely to be at its best. Indeed, change continues: new proposals were published in a Green Paper in March 1995,<sup>5</sup> and they are likely to create further uneasiness about the role of probation.

### I. The “Nothing Works” Debate

Twenty-one years ago, Robert Martinson's article *What Works? Questions and Answers about Prison Reform* appeared in the journal *Public Interest*<sup>6</sup> and sparked a furious academic debate. Years after the term “nothing works” was first coined, it remains a wonderful sound bite. But like so many sound bites, it does not withstand close scrutiny.

It should not take more than a couple of minutes to realize that “nothing works” is a meaningless formulation. No one would maintain that everything works, and it is equally absurd to claim that “nothing works.” At the very least—and only partly frivolously—one can say that “nothing” has not been

---

3. For a discussion of the IP initiative, see George Mair, et al, *Intensive Probation in England and Wales: An Evaluation* (HMSO, 1994).

4. Home Office, *Probation Service in England and Wales: Statement of National Objectives and Priorities* (HMSO, 1984).

5. A Green Paper (so-called because of the color of the cover) is a government paper setting forth options for legislation and calling for consultation by policymakers. See Home Office, *Strengthening Punishment in the Community* (HMSO, 1995).

6. Martinson, 35 *Pub Interest* at 22 (cited in note 2).

tried and that the “works” part of the idea bases the effectiveness of sentences wholly on recidivism, a simplistic and problematic criterion. In fact, nowhere in Martinson's article does he state that “nothing works.” His own summary of the evidence was certainly qualified: “[w]ith few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism.”<sup>7</sup> It is questionable whether this statement can be reduced to the simple conclusion that “nothing works”; indeed it is worth noting that the statement itself contains several ambiguous terms—“few and isolated,” “reported,” “so far,” and, perhaps most important, “appreciable effect.” Further, given the evidence Martinson considers, his statement is too negative. His article is littered with phrases such as “impossible to interpret,”<sup>8</sup> “no clear evidence,”<sup>9</sup> “difficult to interpret,”<sup>10</sup> “ambiguous results,”<sup>11</sup> “suggestive,”<sup>12</sup> “equivocal,”<sup>13</sup> “problem in interpreting,”<sup>14</sup> and “important caveat.”<sup>15</sup> In these post-modern times when we are very aware that reading a text is not a simple activity, and when we know about post-structuralist theories of literature and the deconstruction of texts, a reading of the article should take full account of these cautionary noises and come to a suitably cautious conclusion.

For better or worse, Martinson's work was hijacked by the media. Ted Palmer's work has suggested that it was a result of the interaction between Martinson's article and the media that “nothing works” became an unequivocal finding.<sup>16</sup> But this is not an adequate reason to explain the rapid appropriation—or, more properly, misappropriation—of Martinson's work. Instead, timing was crucial. As Francis Cullen and Karen Gilbert demonstrated,<sup>17</sup> Martinson (quite unintentionally) found a remarkably receptive audience on both sides of the political spectrum: the Right saw Martinson's work as an attack on the rehabilitative ideal that they believed was soft on offenders; the Left saw Martinson's work as an attack on “treatment” they believed was unjust, intrusive, and an abuse of power.<sup>18</sup> In addition, the community's penological agenda was shifting; the justice model, decriminalization, diversion, and crime prevention were all growing in significance. Martinson's work could only encourage these developments, and it therefore fell on very fertile ground.

Equally important was the seductive clarity of “nothing works”—a neat, simple formula that cut through the caveats, equivocations, and convolutions

7. *Id.* at 25.

8. *Id.* at 28.

9. *Id.* at 27.

10. *Id.* at 45.

11. *Id.* at 30.

12. *Id.* at 31.

13. *Id.* at 32-34.

14. *Id.* at 41.

15. *Id.* at 47.

16. Ted Palmer, *Correctional Intervention and Research* xix, 16-17 (Lexington, 1978).

17. Francis T. Cullen and Karen E. Gilbert, *Reaffirming Rehabilitation* (Anderson, 1982).

18. *Id.* at 111-12.

of much social science research.<sup>19</sup> A top advertising agency would be hard-pressed to come up with a phrase so attractive.

In addition to the fact that the conclusion that “nothing works” is not borne out by the details of Martinson's original article, his argument itself is fundamentally flawed in two ways. First, he relies on recidivism as the sole criterion for judging the success of a sentence. Second, he fails to address the issue of how sentences are implemented. These weaknesses are evident both in Martinson's work and in the primary studies he relies upon.

Recidivism has played a central role in post-war criminology. But the usual measurement of recidivism—reconviction rates—has been treated with far more respect than it deserves. Too rarely is it recognized that reconviction rates are, for the most part, artifacts constructed from rather dubious data. Indeed, there is a real issue as to whether recidivism should be used as a measure of effectiveness for all sentences. Certainly, a major objective of community penalties is to reduce re-offending, but whether prisons or fines have this goal as their main objective is debatable. Even if one accepts that reconviction rates are an appropriate measure of effectiveness (and they *may* be more helpful as performance indicators), they are by no means a simple measure; they have various limitations that need to be fully understood before meaning is assigned to them.

First, reconviction rates cannot measure retribution, reparation, general deterrence, denunciation, or whether an offender has been diverted from custody—all of which could be counted as objectives of the sentencing process and, therefore, potential measures of success not captured by reconviction rates. Second, reconviction rates can be defined in a variety of ways. Michael Maltz, for example, identifies nine, each of which has more than one qualifying condition.<sup>20</sup> Third, reconviction is *not* re-offending: estimates from the British Crime Survey (BCS) suggest that for every one hundred offenses committed, only two result in a criminal conviction.<sup>21</sup> Fourth, it is unclear whether there is a “correct” follow-up period: twenty-four months tends to be the norm,<sup>22</sup> but some offenders are re-convicted quickly (e.g., car thieves), while others may take five or more years for reconviction (e.g., sex offenders).<sup>23</sup> Fifth, it is difficult to decide when to begin counting reconvictions: in the case of prison sentences, one normally begins counting at the time of release, but

---

19. Perhaps I should make clear that I do not believe that this is a result of social scientists' congenital inability to be plain-spoken but rather a reflection of the caution that must be shown in the face of complex realities that rarely have clear-cut solutions.

20. The nine factors are arrest, reconviction, incarceration, parole violation, parole suspension, parole revocation, offense, absconding, and probation. Michael D. Maltz, *Recidivism* 62 (Academic, 1984).

21. See Home Office, *Digest 2: Information on the Criminal Justice System in England and Wales* 29 (HMSO, 1993).

22. Charles Lloyd, George Mair, and Mike Hough, *Explaining Reconviction Rates: A Critical Analysis* 6 (HMSO, 1994).

23. *Id.*

for community penalties (e.g., probation or community service orders), counting begins on the date of sentencing.<sup>24</sup> Sixth, there is a time lag between the offense and the conviction. As a result, some offenders offend prior to their "target" appearance but are re-convicted after they have begun their sentence (false positives), while toward the end of the follow-up period others will re-offend but not be reconvicted until after the follow-up period has been completed (false negatives).<sup>25</sup> Seventh, reconviction cannot simply be equated with failure: if it were, the implication would be that all reconvictions are of equal severity, which is patently not the case.<sup>26</sup> Eighth, police and prosecution practices have an impact on reconviction rates (by way of cautioning and discontinuance) that is difficult to measure.<sup>27</sup> And finally, interpreting reconviction rates is complex, (e.g., can all probation day centres be condemned for a high national reconviction rate?).<sup>28</sup>

Martinson's study also fails to account for *how* a programme's success is influenced by the particular manner in which it is operated and organized. This is not surprising, since the primary studies themselves generally fail to address these issues. At most, researchers spend a paragraph or a page describing a programme; we are rarely given the full rationale for why it exists. Researchers only occasionally study how programmes are planned, how they are put into practice, and how and why they change over time. Yet these matters are crucial to determining the success or failure of a project. For example, if a large number of the projects Martinson considered were poorly planned, badly implemented, starved of resources, or were administered by untrained and uncommitted staff with a high turnover rate—would it be any surprise to find the programmes had failed?

By concentrating on outcome, and solely on recidivism, Martinson offered a very limited view of evaluation. For most of the last twenty years, however, the debate on alternatives to custody has followed the same, repetitious path. Opponents have taken the same studies investigated by Martinson, perhaps added a few more, carried out their own review, and proclaimed that Martinson was wrong; Martinson's defenders have counter-attacked.<sup>29</sup> Indeed, there

24. *Id.*

25. *Id.* at 7.

26. *Id.* at 8.

27. For example, an increase in police cautioning is likely to lead to offenses that previously would have resulted in a conviction ending in a caution, which would depress reconviction rates. Similarly, if prosecution discontinuance rates increase, reconviction rates would be depressed.

28. Lloyd, Mair, and Hough, *Explaining Reconviction Rates* at 10 (cited in note 22).

29. See, for example, John T. Whitehead and Steven P. Lab, *A Meta-Analysis of Juvenile Correctional Treatment*, 26 *J Res Crime & Delinq* 276 (1989); Don A. Andrews, et al, *Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis*, 28 *Criminology* 369 (1990); Steven P. Lab and John T. Whitehead, *From "Nothing Works" to "The Appropriate Works": The Latest Stop on the Search for the Secular Grail*, 28 *Criminology* 405 (1990); Don A. Andrews, *A Human Science Approach or More Punishment and Pessimism: A Rejoinder to Lab and White-*

has even been a debate about whether or not Martinson recanted (or more appropriately, modified) his views before his death.<sup>30</sup>

## II. Old Wine in New Bottles: Meta-Analysis

Recently, the participants in this debate discovered meta-analysis. One of the key figures in the development of meta-analysis describes the differences between primary, secondary, and meta-analysis as follows:

*PRIMARY ANALYSIS* is the original analysis of data in a research study. . . . *SECONDARY ANALYSIS* is the re-analysis of data for the purpose of answering the original research question with better statistical techniques, or answering new questions with old data. . . . Meta-analysis refers to the analysis of analyses. . . . [t]he statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.<sup>31</sup>

Meta-analysis, then, is held up as a major new approach that can synthesize large numbers of disparate studies. In this respect, it might be seen as a way of breaking out of the “nothing works” debate. It is a step forward but not the great leap forward that proponents claim.

Only a handful of relevant meta-analyses have been carried out so far.<sup>32</sup> All (with one possible exception)<sup>33</sup> conclude that rehabilitative treatment has a positive affect upon recidivism. This is indisputably good news, but there are several caveats that need to be taken into account.

---

head, 28 *Criminology* 419 (1990).

30. See Anthony Doob and Jean-Paul Brodeur, *Rehabilitating the Debate on Rehabilitation*, 31 *Can J Criminology* 179 (1989); Paul Gendreau, *Programs That Do Not Work: A Brief Comment on Brodeur and Doob*, 31 *Can J Criminol* 193 (1989).

31. Gene V. Glass, *Primary, Secondary, and Meta-Analysis of Research* 5 *Educ Res* 3 (Nov 1976).

32. The major meta-analysis works include the following: Carol J. Garrett, *Effects of Residential Treatment on Adjudicated Delinquents: A Meta-Analysis*, 22 *J Res Crime & Delinq* 287 (1985); Rand Gottschalk, et al, *Community-Based Interventions*, in Herbert C. Quay, ed, *Handbook of Juvenile Delinquency* 266 (John Wiley & Sons, 1987); Whitehead and Lab, 26 *J Res Crime & Delinq* at 276 (cited in note 29); Don A. Andrews, et al, *Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis*, 28 *Criminol* 369 (1990); Mark W. Lipsey, *Juvenile Delinquency Treatment: A Meta-Analytic Inquiry into the Variability of Effects*, in Thomas D. Cook, et al, eds, *Meta-Analysis for Explanation* 83 (Russell Sage, 1992).

33. Whitehead and Lab, 26 *J Res Crime & Delinq* at 291 (cited in note 29), concluded that “correctional treatment has little effect on recidivism.” However, the authors did not calculate an overall effect size. See Friedrich Losel, *The Effectiveness of Treatment in Institutional and Community Settings*, 3 *Crim Beh & Mental Health* 416, 419 (1993), which calculated the mean effect size for the Whitehead and Lab study and found it to be positive.

First, for all their seeming statistical sophistication, meta-analyses simply gather a wide variety of very disparate studies and reduce the findings to a fairly crude set of categories (most notably, in the case of recidivism, findings are reduced to a simple “yes” or “no”). Given that primary studies often rely on somewhat questionable methods, it is difficult to see how these problems are somehow ironed out by further statistical manipulation. I have already noted Maltz's point about the varying definitions of recidivism used by researchers,<sup>34</sup> and I am not convinced that meta-analysis can make such definitions consistent with each other.

Second, the meta-analyses conducted thus far have concentrated almost exclusively on juvenile delinquents in the United States. It is assumed that if treatment works for juveniles, it will work for adults—but this may be a naive assumption. Adults are more likely to be married, have children, be unemployed, or have drug or alcohol problems. In addition, it should be remembered that the U.S. criminal justice system casts its net more widely than do other systems (e.g., the U.K. system), so that on average the offenders included in meta-analyses are not likely to be at a particularly high risk for recidivism.

Third, the meta-analyses continue to rely on primary studies conducted in the 1950s and 1960s. This reliance on the past is not especially helpful. No one would doubt that there have been many radical changes during the past thirty or forty years in the kinds of offenses committed, levels of offending, approaches to dealing with offenders, and the entire cultural context in which offenders live their lives. Thus, to try to apply lessons that might have worked in the past may not be a sensible approach.

Fourth, as soon as one begins to look in detail at a meta-analysis, problematic issues begin to emerge that could have a bearing on the studies' findings and how they are interpreted. Take for example, Mark Lipsey's epic study,<sup>35</sup> partly because it is the most ambitious carried out so far, and partly because Lipsey very clearly sets out the categories for the variables studied. Lipsey cites 443 studies. For the primary delinquency measure (which is not always recidivism) half of the follow-up studies have follow-up periods of less than six months.<sup>36</sup> Almost half of the studies were coded as having low treatment integrity (43.8 percent), and approximately 75 percent were either low or moderate on this rating (79.5 percent).<sup>37</sup> Twenty percent of the studies covered institutionalized juveniles and 50 percent covered non-juvenile justice interventions.<sup>38</sup> Almost two-thirds of the programmes were less than two years old, which raises questions about the lasting impact of an initially enthusiastic response to a program.<sup>39</sup> Finally, one-quarter of the programmes

---

34. Maltz, *Recidivism* at 68-87 (cited in note 20).

35. Lipsey, *Juvenile Delinquency Treatment* at 83 (cited in note 32).

36. *Id.* at 104 (29.6% of the primary studies had a follow-up period between fourteen and twenty-six weeks and 13.5% had a follow-up period between one and thirteen weeks).

37. *Id.* at 108.

38. *Id.* at 109.

39. *Id.* at 110.



were administered by criminal justice personnel, one-quarter by mental health personnel, and 20 percent by lay persons.<sup>40</sup>

Fifth, researchers using meta-analysis do not fully agree about its advantages, and they themselves have identified several technical issues that need to be addressed.<sup>41</sup> Indeed, some have suggested a set of criteria for evaluating meta-analysis.<sup>42</sup>

Finally, meta-analysis—despite the claims of many of those who use its findings—can really only tell us that treatment in the general sense can have a positive affect upon recidivism. It cannot tell us precisely what kind of treatment should be used, when it should be used, in what doses, with what offenders, or within what kind of programmes. Nor does it tell us about the kinds of structures, the kinds of staff required, or the resources necessary to deliver effective programs.

It is too easy to claim that meta-analyses demonstrate which principles need to be followed to guarantee effective programmes—for example, targeting high-risk offenders, ensuring high-treatment integrity, focusing upon criminogenic factors, or using structured programs appropriate to the abilities and needs of offenders. Such claims have often been made in the U.K. during the last two years. But it is important to emphasize that meta-analysis as presently structured cannot offer such principles with any confidence (and, indeed, the meta-analysts themselves tend to be cautious with their claims). The primary sources rarely say anything detailed about the kinds of programmes studied, how they were organised and delivered, whether they changed over time, and so forth. And the reductive focus of meta-analysis, whereby aspects of studies are coded down into a few crude categories, means that even if such details were available, they would be categorised in such a way as to render them unhelpful.<sup>43</sup>

Most of the principles that are claimed to have emerged from meta-analysis are, in fact, common-sense formulations that any rationally planned programme should address—for example, targeting appropriate offenders, using a multi-modal approach, focusing upon criminogenic needs, and basing the programme upon sound theoretical foundations. Where meta-analysis can be useful is in *suggesting* what appear to be the most promising ways to move forward; it certainly does not offer cast-iron recipes for successful treatment programmes. For example, Lipsey's work has been used in the U.K. to claim that employment is the key to success in reducing re-offending among young offenders, but careful scrutiny of his findings shows that only four studies fall

---

40. Id at 101.

41. See, for example, Gene V. Glass, Barry McGaw, and Mary Lee Smith, *Meta-Analyses in Social Research* 217-31 (Sage, 1981); Robert Rosenthal, *Meta-Analytic Procedures for Social Research* 127-36 (Sage, 1991).

42. R. J. Bullock and Daniel J. Svyantek, *Analyzing Meta-Analysis: Potential Problems, an Unsuccessful Replication, and Evaluation Criteria*, 70 *J Applied Psych* 108 (1985).

43. It is worth noting that coding in general is an absolutely critical factor in meta-analysis that is usually ignored.

into the category of juvenile justice programmes that focus upon employment.<sup>44</sup> This is not a very large number, and in any event there is no information about how such knowledge might be operationalised.

Meta-analysis, then, is by no means the final word on the "nothing works" discussion, but it does offer us a chance to break away from a sterile and moribund debate. This is where evaluation may be helpful.

### III. Evaluation

Perhaps the key reason for the longstanding debate over "what works" is the lack of high-quality evaluation. The vast majority of studies that have evaluated community penalties have been badly conceived, have been limited in focus, have relied upon simplistic research designs, and have tended to make claims that are not supported by the data. Indeed, there is considerable confusion about what "evaluation" is and whether it differs from "research."

Every now and then, certain words and phrases acquire a power over and above their normal everyday meaning. A few years ago, one such word in criminology was the term "community." There was discussion of community policing, community crime prevention, community probation, community prisons, and so on.<sup>45</sup> It seemed that anyone who wanted to set up a new initiative had only to use the word "community" to be noticed favourably. But David Nelken noted a fatal weakness behind this when he asked, "[i]f community is the answer, what is the question!"<sup>46</sup> "Community" began to bear such a weighty meaning that it ultimately became meaningless, simply a word to be used to show that one understood the issues.

There are signs that the same process may be at work with the term "evaluation." "Evaluation" is being used more and more by those who, only a few years ago, would have used the term "research." Now, "evaluation" is used as a synonym for "research" and as a key to unlock the doors to funding. It is, however, essential to differentiate "research" from "evaluation" no matter how much the processes may overlap in practice. There is no doubt that research can be (and often is) evaluative and that an evaluation will include (if not consist wholly of) research, but the differences between the two processes should not be glossed over or ignored. I do not propose to offer any definitive answer to the question of what is meant by "evaluation," but I do offer some thoughts that should be addressed both by those who commission

44. Lipsey, *Juvenile Delinquency Treatment* at 124 (cited in note 32).

45. Paul Henderson and Sarah del Tufo, *Community Work and the Probation Service* (HMSO, 1991); Tim Hope and M. Shaw, eds, *Communities and Crime Reduction* (HMSO, 1988); David Brown and Susan Iles, *Community Constables: A Study of a Policing Initiative* (HMSO, 1985); Prison Service, *Community Prisons: A Consultation Paper* (London, 1992).

46. David Nelken, Address to the Howard League Conference (1985), quoted in Cedric Fullwood, *Probation, Community and Inter-Agency Dimensions: A Future Look*, in Robert Shaw and Kevin Haines, *The Criminal Justice System: A Central Role for the Probation Service* 119 (Inst of Criminology, Cambridge, 1989).

evaluations and by those who carry them out.

“Evaluation” implies some assessment of success or failure, some kind of measure of how well a programme or an initiative is working, and how much of its aims and objectives are being achieved. Evaluation is carried out for someone or some organisation and is connected with the policy process.<sup>47</sup> “Research,” on the other hand, is more concerned with the establishment of new facts. While it is connected to the scientific process of the acquisition of knowledge, it is not necessarily linked to the policy process. Thus, “the purpose of evaluation is to provide guidance to program staffs and policy makers rather than to contribute to theory development. Accordingly, the intent of evaluation studies is to document the utility and implementation of purposeful actions rather than to temporarily find ‘truth’ by rejecting a null hypothesis.”<sup>48</sup>

Evaluation is, by its very nature, “political” in a way that research is not; it is complex, problematic, and riddled with uncertainties.<sup>49</sup> Various models of evaluation can be found, but the two main types of interest here are “process evaluation” and “outcome evaluation,” or “impact evaluation.” Process evaluation concentrates on assessing how a project was put into practice, what actually happened at the ground level of the project, what kind of changes took place, and why they occurred. Outcome evaluation, on the other hand, is concerned with the impact of the project—how far it moved toward achieving what it set out to achieve, and whether there are any other consequences. The point to emphasise is that both are necessary for a full evaluation. Indeed, I would argue that a process evaluation is a prerequisite for any outcome evaluation (and that a great deal can be learned from it in its own right). Process evaluation provides the explanatory framework within which outcomes can be interpreted; without a focus on process, outcome measures are left in a vacuum and it becomes impossible to learn from success or failure.

With respect to this Symposium, what kind of issues should we consider to evaluate properly “viable alternatives to prison?” First, it is crucial to clarify the meaning of “viable”; should we read it in a purely political sense, a financial sense, an organisational sense, a structural sense, a humane sense, an operational sense, in terms of public or judicial satisfaction, or in some combination of these ways? To be slightly flippant, we know that alternatives to prison are viable in the crudest sense of the word simply because so many of them exist: the United States has spent vast sums of money devising and

---

47. This may happen at the macro- or micro-level, or it may be that the evaluation is carried out for those who support or oppose the policy or strategy in question.

48. Leigh Burnstein and Howard E. Freeman, *Perspectives on Data Collection in Evaluations*, in Leigh Burnstein, Howard E. Freeman, and Peter H. Rossi, *Collecting Evaluation Data* 15, 16-17 (Sage, 1985).

49. Paul Ekblom, *Evaluation: The Management of Uncertainty*, Paper Presented to the British Criminology Conference (1989) (on file with the *University of Chicago Law School Roundtable*).

researching alternatives to prison (or intermediate punishments) over the last twenty years, and, as I noted at the outset, a great deal of English penal policy since 1948 has been driven by the need to provide alternatives to prison. The same can be said for many other European countries.<sup>50</sup>

In terms of evaluation, it is necessary to look at *all* the possible interpretations listed above, perhaps considering some to be more important than others, as well as keeping an eye open for any unintended or unanticipated consequences that have not been considered. This, of course, makes life much more complicated, since it is unlikely that an alternative to prison would be equally viable in all the senses listed. For example, an alternative to prison might be effective in political and financial terms, although operationally (in terms of actually diverting offenders from custody) it might be a failure. Indeed, success measured by one criterion might even lead to perceived failure as measured by another: research into probation centres in England and Wales suggested that they were fairly successful as an alternative to a custodial sentence, although a subsequent reconviction study showed very high levels of reconviction.<sup>51</sup> By focusing on offenders with a long criminal history and successfully selecting them for a community penalty, probation officers were setting themselves up for failure in terms of recidivism. Two measures of the effectiveness of probation centres (diversion from custody and recidivism), therefore, cut across each other; success in one almost precludes success in the other.

No court sentence has but one objective: sentences aim to do a variety of things—among them, reduce offending, retribute, repair, denunciate, and generally deter. In evaluating the effectiveness of sentences, therefore, to rely upon one measure (such as recidivism) is, to say the least, misleading. A variety of measures of effectiveness are required to evaluate a sentence. To go down such a path raises almost as many issues as it resolves, and I will discuss some of these issues later. First, however, I wish to provide an example of the approach to evaluation that I have outlined here.

#### IV. Intensive Probation in England and Wales

The IP initiative that began in April 1990 may well have been the last in a series of official “alternative to custody” initiatives.<sup>52</sup> These initiatives have officially ceased since the introduction of the 1991 Criminal Justice Act.

---

50. See, for example, Norman Bishop, *Non-Custodial Alternatives in Europe* (Helsinki Inst for Crime Prevention and Control, 1988).

51. George Mair and Claire Nee, *Day Centre Reconviction Rates*, 32 *Brit J Criminology* 329 (Summer, 1992).

52. These are attendance centres, day centres, specified activities schemes, suspended sentence, community service, and bail information schemes. See George Mair, *Intensive Probation in England and Wales: Origins and Outlook*, Paper Presented at the British Criminology Conference, Bristol (July 17-20, 1989); George Mair, *Part Time Punishment: The Origins and Development of Serious Attendance Centres* (HMSO, 1988); A. E. Bottoms, *The Suspended Sentence in England, 1967-1978*, 21 *Brit J Criminology* 1, 26 (1981); Charles Lloyd, *Bail Information Systems: Practice and Effect* (HMSO, 1992).

Early in 1989, the Home Office invited ten probation areas to establish IP programmes that would run for two years.<sup>53</sup> This policy was motivated by a variety of factors: prison overcrowding, financial constraints, a perceived need for effective punishment, the need to regain the confidence of sentencers, the success of efforts to reduce juvenile involvement in court proceedings, and the U.S. example of intensive probation sentencing. This is not unusual, since social policy in general is driven by many disparate factors, and penal policy is no different in this respect.<sup>54</sup> Perhaps more important, however, is the point that IP was thereby expected to satisfy—implicitly if not explicitly—very different demands. If IP would have proven able to reduce offending, reduce the prison population, provide effective punishment, keep sentencers satisfied, and reduce costs all at the same time, the Holy Grail would indeed have been found.

IP was intended as an experimental, developmental initiative, and as a result, there were few guidelines about what it should look like and how it should be put into practice. It was intended to target high risk offenders, seventeen to twenty-five years old, who were likely to receive a prison sentence.<sup>55</sup> IP schemes were expected to be rigorous and demanding—they have to include strict referral and selection procedures, comprehensive individualised programmes worked out with the offender and approved by the court, frequent contact with probation officers, a focus on confronting offending behavior, a multi-agency approach, and services for ethnic minority and female offenders.<sup>56</sup>

These guidelines left considerable discretion to the initiative and imagination of probation officers so that they could establish innovative projects. But even before IP officially began, problems emerged. Only seven of the ten areas participated fully. One did not wish to get involved,<sup>57</sup> one failed to develop an IP programme during the course of the twenty-four-month experiment,<sup>58</sup> and one ran for less than twelve months.<sup>59</sup> Even among those that did take part, there was evidence that they feared Home Office disapproval if their programmes were unsuccessful.<sup>60</sup> Not all IP schemes were at the same level of development—some were underway before the official start date, some were planning to implement IP programmes before the official start date, and others were waiting for the official start to begin the process.<sup>61</sup> Innovative IP projects were rare—probation officers did not seem to have the space, time, or enthusiasm to grasp the opportunity offered them, and local politics and

---

53. Mair, et al, *Intensive Probation Evaluation* at ix (cited in note 3).

54. Mair, *Origins and Outlook* (cited in note 52).

55. Mair, et al, *Intensive Probation Evaluation* at ix (cited in note 3).

56. *Id.* at ix-x.

57. *Id.* at 7.

58. *Id.* at 20.

59. *Id.* at 9, 18.

60. *Id.* at 9.

61. *Id.*

bureaucratic procedures also proved to be obstacles to success.

Given the nature of the IP initiative, a simple evaluation design focusing solely upon recidivism would have been singularly inappropriate. The study was built around the need for both a process and an outcome evaluation. The process evaluation was vital not only because of the importance of providing a context within which outcome measures could be interpreted, but also because of the experimental nature of IP. It would have been unfair to condemn IP as a failure without trying to discern the reasons for failure, and such data could only be supplied by a process evaluation. For the process evaluation, three IP programmes were chosen for detailed study; they were not intended to be representative of IP but rather were chosen on the basis of their different approaches.<sup>62</sup>

For the outcome evaluation, various measures of effectiveness were devised. The primary measures were: (1) reconviction rates,<sup>63</sup> (2) diversion from custody, (3) the financial costs of IP, (4) the views of sentencers, and (5) the views of offenders.<sup>64</sup> The outcome evaluation also used secondary measures to cover factors that were specific to individual IP schemes or to the individualised plans put together for offenders. These included help with accommodation, employment, drug misuse, relationships, and use of leisure time.

The research was carried out from 1990 to 1992, and a full report on the results of the evaluation was published recently. In the following subsection, I will briefly discuss some of these results, and I will then examine some issues relating to the central policy development of IP and its implementation.

#### A. THE RESULTS OF IP

During the twenty-four months of the experiment, 1,677 offenders were referred for intensive probation.<sup>65</sup> The majority were male (95 percent),<sup>66</sup> and more than three-quarters were between seventeen and twenty-five years old.<sup>67</sup> Around 6 percent were members of ethnic minority groups,<sup>68</sup> and 83 percent of referrals were unemployed.<sup>69</sup> On the whole, IP schemes were successful in catering to ethnic minority offenders (although at least one area had no ethnic minority referrals),<sup>70</sup> but almost half of the female offenders came from only

---

62. Id at 8.

63. Reconviction rates were measured during the course of the IP project as well as during the subsequent period of supervision, including the time until reconviction, the reconviction offence, and the sentence.

64. Mair, et al, *Intensive Probation Evaluation* at 9 (cited in note 3).

65. Id at 11.

66. Id.

67. Id.

68. Id.

69. Id.

70. Id at 12.

one IP scheme, which had set up special programmes for female offenders.<sup>71</sup>

The most common offence was burglary (42 percent of the cases studied),<sup>72</sup> with theft (16 percent) and violent offenses (14 percent) also being fairly common.<sup>73</sup> Only 6 percent were first offenders,<sup>74</sup> while 51 percent had six or more previous convictions.<sup>75</sup> Fifty-four percent had served a previous custodial sentence,<sup>76</sup> and almost two-thirds were sentenced at the Crown Court.<sup>77</sup> These figures suggest that IP was successful at targeting offenders who had a high risk of serving a custodial sentence, but they hide wide variations among the areas, as table 1 shows.

Table 1  
Characteristics of Referrals to Intensive Probation Schemes by Area

IP Area	Total Number of Offenders	17-20 (%)	Male (%)	Was Sentenced at Crown Court (%)	Has Six or More Previous Convictions (%)	Had Previous Custodial Sentence (%)
Durham	95	52	98	50	61	67
Leeds	532	47	97	73	50	58
West Midlands	187	86	96	47	29	41
Berkshire	94	88	94	60	35	32
Gwent	186	36	99	55	68	68
Hampshire	351	40	98	85	63	62
Manchester	17	82	88	35	59	65
Northumbria	215	35	81	39	43	38
Total	1,677	50	95	64	51	54

It is clear from the table that some areas were much more successful at targeting high-risk offenders than others; Hampshire, Gwent, Leeds, and Durham were markedly better in this regard than the West Midlands, Berkshire, and Northumbria.<sup>78</sup> The reasons for such differences lie in the design and objectives of the schemes. Some focused on the Crown Court while others did not;<sup>79</sup> some aimed for seventeen- to twenty-year-olds only, while others did not;<sup>80</sup> some simply re-designated an existing facility as an IP scheme;<sup>81</sup> one designed and set up a voluntary programme.<sup>82</sup>

71. Id at 19-20.

72. Id at 11.

73. Id.

74. Id.

75. Id.

76. Id.

77. Id.

78. Manchester has been ignored because it only had seventeen offenders.

79. Mair, et al, *Intensive Probation Evaluation* at 23 (cited in note 3).

80. See, for example, id at 12-20.

81. Id.

82. Id at 19.

Sentencer satisfaction with IP was high.<sup>83</sup> But widely varying views of IP were put forward, not all of which could be easily reconciled: sentencers saw IP as an alternative to custody as well as expecting it to reduce re-offending;<sup>84</sup> they did not agree about the position of IP among the sentencing options available to the courts;<sup>85</sup> and they often mentioned relatively minor offences as being suitable for IP.<sup>86</sup> In general, sentencers saw the advantages of IP as either system-oriented (i.e., leading to a reduction in the numbers being sentenced to custody) or welfare-oriented (i.e., rehabilitative).<sup>87</sup> They rarely made any mention of punitive aspects of IP.<sup>88</sup> By far, sentencers saw the major potential drawback of IP as the possibility that offenders—or the general public—might see it as a “soft” option.<sup>89</sup>

Probation officers who worked on IP schemes were enthusiastic about them (with the exception of those in the West Midlands, one case study area), and they considered themselves to be doing positive work with high-risk offenders.<sup>90</sup> Non-IP staff were not quite so enthusiastic, although their reactions were dependent upon the length of time the program had been in effect and the degree of consultation that had taken place in the local probation service. Negative comments centered on the elitist nature of IP, the claim that more resources had been made available to IP when the same ends could have been carried out by ordinary probation orders, and the use of voluntary organisations where staff were regarded as unqualified.<sup>91</sup> In addition, questions about the enforcement of orders were raised.<sup>92</sup> It was notable that there were few direct criticisms of IP as overly controlling.

Offenders who participated in IP schemes viewed the programmes positively.<sup>93</sup> Probation officers who had opposed IP expected offenders to rebel against the rigorous requirements expected of them, but, if anything, they appeared to enjoy and appreciate the attention that was given to them by their project workers.<sup>94</sup> For the first time they felt that someone was taking a real interest in them and what they did. This was reflected in the remarkably positive comments made about their project workers, who were, for the most part, untrained probation officers.<sup>95</sup> Unfortunately, because of the dominance of the project workers in their lives, the offenders perceived the role of probation officer as less

---

83. Magistrates in the case study areas were asked to complete a mailed questionnaire, and a few judges were interviewed personally. *Id.* at 118.

84. *Id.*

85. *Id.*

86. *Id.*

87. *Id.*

88. *Id.*

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.* at 118-19.

93. *Id.* at 119.

94. *Id.*

95. *Id.*



helpful.<sup>96</sup>

As required by the guidelines, all IP schemes used voluntary organisations to assist in running the IP programmes to a greater or lesser extent, but there was considerable uneasiness among probation officers about utilizing such organisations (and such uneasiness seemed on occasion to result in anxieties among workers in voluntary organisations about their role in IP).<sup>97</sup> Whether or not there was suspicion about the motives of voluntary organisations was dictated by the history of working with these groups: if this had been a feature of probation for some time in a given IP area, there was less unhappiness, but if it was a new development, suspicions were aroused (with consequent repercussions for referrals to voluntary organizations). Unfortunately, suspicions were fuelled by the publication of a government discussion paper in April 1990.<sup>98</sup> This paper, which coincided nicely with the beginning of the IP initiative, was widely perceived as a move by the government to take certain tasks away from the probation service and move them to voluntary organisations.<sup>99</sup> IP demonstrated that partnership schemes with voluntary bodies could work, although tension between the two has not yet completely disappeared.

While it should be clear that the IP initiative arose in a political context, it is important to discuss this further. Over the last ten years, government policy with regard to probation has focused on efforts to make it tougher and more accountable,<sup>100</sup> and IP was intended to be another step in this process. From this perspective, IP's symbolic function cannot be discounted: it was planned to provide rigorous and demanding punishment of offenders in the community. The probation services involved responded accordingly. However, it should be noted that in practice IP was rarely about controlling offenders. A great deal of social work was carried out under the aegis of IP, and even where control/surveillance was considerable, offenders were not unhappy about the demands made of them. The politics of IP, therefore, are not clear. For some probation officers, IP was a sensible approach for dealing with serious offenders in the community who might have received a custodial sentence in the community. For others, it was a further betrayal of traditional social work values and another nail in the coffin of social work with offenders. All had to make some kind of accommodation to IP, and in this sense its contribution to culture change in probation should not be underestimated.

## B. POLICY DEVELOPMENT AND IMPLEMENTATION

The results of IP were not simply outcomes of the individual schemes; they were also a result of the way in which IP was developed as a policy, the way that

---

96. It is possible that this may have an affect on post-IP supervision. See *id.*

97. *Id.*

98. Home Office, *Partnership in Dealing with Offenders in the Community* 119 (HMSO, 1990).

99. *Id.*

100. *Id.* at 121.

policy was transmitted to the probation service generally and to practitioners, and the approaches taken at the local level to put IP into practice. These issues are all too often assumed to be unproblematic and unrelated to the outcomes of policy initiatives. In fact, they can have a critical impact upon the success or failure of policies.

With regard to the development of IP at the central government level, there were several relatively minor, though not insignificant, matters that suggest that IP was not developed as well as it might have been. In the first place, the IP initiative was initially incorporated in another development, the "Tackling Offending" initiative.<sup>101</sup> This led to a good deal of confusion among probation services about how IP differed from Tackling Offending—was it a separate development or a refinement? The confusion was compounded by the fact that the government introduced the IP initiative only six months after the Tackling Offending initiative was proposed. If areas had been approached at the same time or twelve months later, initial responses might have been more welcoming. IP areas were even more confused by an announcement soon after IP had formally begun that a new community penalty was to be introduced in October 1992 (the "combination order"); the relationship between IP and the new order was never satisfactorily explained.

There were also several assumptions that appeared to be embedded in IP policy and that were not borne out in practice. Having a formal start date for IP was a bureaucratic nicety but relatively meaningless; those areas that began IP on the formal start date had been running a scheme (under another name) for some time—and the length of time for which schemes had been running was an important factor in their success.<sup>102</sup> The government all too easily assumed that all of the probation areas approached were functioning at the same level of development and were willing and able to respond equally effectively to the IP initiative. In practice, some areas took a considerable amount of time to set up an IP scheme, and this again had consequences for success.

Home Office guidelines regarding the form of IP were vague, which had the advantage of allowing areas the discretion to develop their own ideas but also the drawback of uncertainty about the precise goals of IP. Some IP schemes had initially been responses to the Tackling Offending initiative and had simply changed names.<sup>103</sup> As officials struggled to define IP in their own minds, one area was included, then excluded, and finally included again in the IP initiative.<sup>104</sup> Plans to develop an additional ten IP schemes twelve months after the first group had begun to operate simply crumbled.<sup>105</sup> There was little evidence of innovative and imaginative IP schemes, perhaps because of fear of failure, cynicism, lack of time, or lack of imagination. The most innovative scheme was

---

101. Home Office, *Tackling Offending: An Action Plan* (HMSO, 1988).

102. Home Office, *Partnership in Dealing with Offenders* at 121 (cited in note 94).

103. *Id.*

104. *Id.*

105. *Id.*

a voluntary one that departed from Home Office guidelines.<sup>106</sup>

The issue of clarity versus ambiguity in policymaking has no easy answers; it is difficult to strike the right balance between stating policy in general—and inevitably ambiguous—terms and in issuing clear and specific but often inappropriate and rigid, instructions. At the very least, it would be helpful for the government to be clear about where ambiguity exists and why it is utilised. If one desires to encourage experimental initiatives, then ambiguity and vagueness are helpful, though again it should be noted that there may be a price to pay in terms of trying to evaluate such initiatives. Setting the criteria for success may not be easy in such cases.

### V. “Crawling from the Wreckage”: Concluding Thoughts

What has the evaluation of the IP initiative taught us about using this approach to evaluate the impact of community penalties? First, although the Home Office collected a great deal of data, not all of it points to a uniform conclusion. On the whole, IP appears to have been successful, bearing in mind that success is a matter of degree and not an all-or-nothing judgment.<sup>107</sup> Not only did IP target high-risk offenders, but sentencers thought it was useful, and offenders appreciated it. Reconviction data are not yet available, but work has begun on a reconviction study that should be completed before the end of the year.

IP was not equally successful in all areas—it worked better in some probation areas than in others, and it worked better in some manifestations than in others.<sup>108</sup> The research demonstrated that the process by which policy is developed centrally, transmitted to probation managers, and then translated into practice is considerably significant not only in its own right but in terms of its impact upon the workings—and thus the outcomes—of IP. It has helped to address the point made some years ago by Alan Harland and Philip Harris<sup>109</sup> about the lack of process studies in criminal justice research and how this cripples new developments.<sup>110</sup> It also confirms the findings of Michael Musheno and his colleagues<sup>111</sup> about the importance of organisational conditions as a prerequisite to successful policy. Without a focus on process evaluation, many of the important conclusions and insights from this analysis would have remained hidden.

The study was not as successful in collecting data on the costs of IP or at

---

106. *Id.*

107. Dennis J. Palumbo, Steven Maynard-Moody, and Paula Wright, *Measuring Degrees of Successful Implementation*, 8 *Evaluation Rev* 45, 47 (1984).

108. Mair, et al, *Intensive Probation Evaluation* at 124 (cited in note 3).

109. Alan T. Harland and Philip W. Harris, *Structuring the Development of Alternatives to Incarceration*, in Stephen D. Gottfredson and Sean McConville, eds, *America's Correctional Crisis: Prison Populations and Public Policy* 179 (Greenwood, 1987).

110. *Id.* at 180-81.

111. Michael C. Musheno, et al, *Community Corrections as an Organizational Innovation: What Works and Why* 26 *J Res Crime & Delinq* 136 (1989).

utilising secondary measures of outcome. Few probation areas are able to assess the costs of specific parts of their work, although implementation of a Resource Management Information System may change this. But even if IP were several times more expensive than ordinary probation, it would still be less costly than custody in a straightforward comparison.<sup>112</sup> Our plans to collect information about the extent to which IP successfully tackled problems faced by offenders (such as accommodation, employment, and drug or alcohol misuse) were hindered by the relative inaccessibility of such information. There is no doubt that probation officers collect and hold such information in the course of their work, but it is kept in files as part of manual case notes; considerable resources would have been required to collect such information solely for research purposes. Again, this situation should change in the future. An ambitious Case Record Management System is under development, and a research project that will test various approaches toward a Problem Assessment Scale is about to commence and become available for probation officers to use.

On a more general level, it is possible to summarize some pointers for setting up alternatives to prison which are related to the approach to evaluation that has been outlined in this Article.<sup>113</sup>

First, careful analysis of the perceived problem is vital, and planning should be based upon this assessment as much as possible. The situation may be temporary; it may be a result of remands to custody rather than sentencing; it may be because of the numbers of criminals sentenced to custody; it may be that sentencing lengths were increased; or, it may be that new sentences were over-using custody. Thus, different approaches may be required in each of these situations, and none of them may necessitate a new or revamped community penalty. For example, if the goal is to reduce the prison population, there are other ways to do so without new non-custodial sentences. Therefore we should not necessarily concentrate our attention on this solution.<sup>114</sup>

Second, the new penalty should have clear achievable objectives that can be prioritised. One of the curses of alternatives to custody has been the stream of researchers who have argued that alternatives only serve to complement prison sentences rather than substitute for them.<sup>115</sup> Maeve McMahon has demonstrated that some of the studies most frequently cited as providing evidence of net-widening were methodologically flawed, and she has cautioned against the

---

112. Sophisticated analyses of sentences on a cost-effectiveness basis are rare, but exceptions have been found. See, for example, Martin Knapp, Eileen Robertson, and Gill McIvor, *The Comparative Costs of Community Service and Custody in Scotland*, 31 *How J Crim Just* 8 (1992).

113. Of course, the most thorough evaluation must consider other factors not within the scope of this Article (e.g., adequate funding, the need to mobilize agency support, and the recognition that programs have a shelf life and are not immutable).

114. David Smith, et al, *Reducing the Prison Population: An Exploratory Study in Hampshire* (HMSO, 1984).

115. See Maeve McMahon, "Net-Widening": *Vagaries in the Use of a Concept*, 30 *Brit J Criminology* 121 (1990).

negativity and pessimism which can ensue if these conclusions are accepted at face value.<sup>116</sup>

Third, there should be a structured and coherent approach to implementation at all levels. The implementation process is usually taken for granted when a new initiative is established; there is a naive assumption that what has been planned on paper will be easily put into practice. This is rarely the case. In the first place, there is often a long, hierarchical chain of individuals who have to implement the programme. The greater the number of steps to be climbed before the programme can be put into practice, the greater the chance that changes will take place. This may be due to misunderstanding, reinterpretation due to local circumstances, or opposition.

Fourth, the initiative should include a basic monitoring system. Monitoring involves the collection of simple, basic information that should help to identify the existence of a problem. While monitoring is unlikely to identify the precise nature of the problem, that is not its objective. Those overseeing the programme should agree on a monitoring system, and that system should be relevant—and be perceived by others as relevant—to their work. It should be regularly analyzed and interpreted and, whenever necessary, its findings should be promptly acted upon. Of course, monitoring has its limitations, and we need to be clear about these. It should not be confused with research or evaluation. Rather, it should be treated as an integral part of a new initiative. It can play a vital role as an early warning system.

Fifth, a full, independent evaluation should be planned from the beginning, although outcomes from the first twelve months should be treated with caution. Such an evaluation is time-consuming and resource-intensive; it can affect the programme under study and can be difficult to interpret; it may make life more, rather than less, complicated for policymakers, as it is unlikely to provide simple answers. But it much more accurately reflects the complex nature of the phenomenon being evaluated. It should also offer some understanding of the mechanisms that produced outcomes (which allows for more effective replication of the program, or at least a better chance to learn from failure).

As I wrote this Article, two quotations kept running through my mind. One was from Hegel: “[W]hat experience and history teach is this—that nations and governments have never learned anything from history, or acted upon any lessons they might have drawn from it.”<sup>117</sup> The other was from George Santayana, who said, “[t]hose who cannot remember the past are condemned to repeat it.”<sup>118</sup> The history of alternatives to custody is largely one of re-invention of the wheel; the time to move on is long overdue, and one way to do so is to pay more attention to the role of evaluation.

---

116. Id at 132.

117. Georg Wilhelm Friedrich Hegel, *Lectures on the Philosophy of World History* 21 (H. B. Nisbet trans, Cambridge, 1975).

118. George Santayana, *The Life of Reason* 284 (Charles Scribner's Sons, 1905).