RISK ASSESSMENT IN PAROLE DECISIONS: A STUDY OF

LIFE SENTENCE PRISONERS IN ENGLAND AND WALES

Robert Anthony Forde

**A thesis submitted to the College of Life and Environmental Sciences of
the University of Birmingham
for the Doctorate in Forensic Psychology Practice**

**Centre for Forensic and Criminological Psychology
School of Psychology
University of Birmingham
United Kingdom**

Acknowledgements

I would like to thank my supervisor, Prof Anthony Beech, for his help in preparing this thesis. This has extended beyond advice about the conduct of research. I have considerable experience of writing psychological reports, journalistic articles, and fiction, but almost none of academic writing, and his help in deciding how to present this work has therefore been invaluable.

I would also like to thank my wife, Pamela, who agreed to marry me without any idea that she would spend the first three years of married life as a doctoral thesis widow. Her support, in this as in so many other things, has been second to none.

Most importantly, I would like to thank the 100 prisoners who gave me unrestricted access to often highly personal details of their lives, offending histories, and parole applications. I hope they feel that the result makes a fair and objective contribution to the debate over parole which periodically arises.

Abstract

This thesis examines the relationships between risk assessment and parole decisions. Chapter 1 introduces the problem. Chapter 2 systematically reviews 29 papers involving 20,568 participants, concluding that practices vary widely, but subjective rather than evidence-based risk assessment predominates. Chapter 3 reports a study of how parole decisions related to three widely-used risk assessment instruments (the PCL-R, the HCR-20, and the SVR-20), and recommendations of professionals (psychologists and probation officers) on 100 life sentence prisoners in England and Wales, 84 of whom were eligible for parole. The study found that parole decisions were related to the recommendations of professionals, especially that of the offender manager (external probation officer). Professional recommendations themselves were related to the more subjective subscales of the risk assessment instruments. Chapter 4 considers an instrument used in the research, the PCL-R psychopathy assessment, concluding that the PCL-R, although it may have been successful in academic research, lacks reliability when used as a risk assessment instrument "in the field". Chapter 5 discusses the findings, concluding that the present system of risk assessment for parole in England and Wales is not evidence-based and that as a result many low-risk prisoners are likely to undergo prolonged detention unnecessarily.

CONTENTS

LIST OF APPENDICES

LIST OF TABLES

FIGURE

CHAPTER 1: INTRODUCTION TO THE PROBLEM

From the earliest days of prison systems there have been schemes to release prisoners into the community before their entire sentence was served (Petersilia, 2003). Sometimes this took the form of day releases for various purposes, and sometimes it amounted to a form of early release. Usually release was dependent on good behaviour in prison, and often continued freedom depended upon continued good behaviour in the community. Such schemes were often operated at the discretion of the prison governor, or in the case of transported prisoners the governor of the prison colony. Petersilia (2003) describes a classic scheme developed by Alexander Maconochie (1787-1860), governor of the English penal colony at Norfolk Island near Australia. Maconochie's system, introduced in 1840, had five stages ranging from strict imprisonment to full freedom in the community. Prisoners earned points for good behaviour which enabled them to progress from one stage to the next. Some graduates of the system became successful members of society, earning the nickname "Maconochie's gentlemen". Maconochie is regarded as a pioneer in the field of offender rehabilitation.

Although Maconochie continued advocating his methods after returning to the UK to be governor of the new prison at Birmingham in 1849, he was dismissed in 1851 because it was felt that his methods displayed undue leniency towards serious offenders. He had an important influence on the development of parole schemes in the UK and Ireland, but there was little in the way of risk assessment in his scheme. Prisoners progressed to a higher stage because they had shown an ability to cope with the previous one, and not because of any consideration of the risks which they might pose to the public.

In the UK, a system of parole was formalised in the 1920s, and parole boards established for England and Wales, Scotland, and Northern Ireland. It was the job of the parole boards to assess the degree of risk posed by the prisoner and decide whether the level of risk was sufficiently offset by potential rehabilitative gains to make release desirable. Initially, parole was applied to prisoners with both determinate and indeterminate sentences ("lifers"), the latter only becoming eligible to apply for parole after the completion of a minimum term after the abolition of capital punishment. Previously there had been no stated minimum. The test which parole boards were to apply was whether the prisoner posed no more than a minimal risk to life and limb (Parole Board for England and Wales, 2012). However, although these terms of reference put risk to the public at the forefront of parole boards' considerations, there was no systematic scheme for assessing that risk. Parole boards were free to consider the offender's behaviour in prison, attitudes, work skills, education and family support, and anything else they wished, but there was no indication of the relationship between these attributes and criminal risk. For many of these factors, and many others widely assumed to be risk indicators, no relationship with risk has ever been demonstrated (Coid, Yang et al., 2011; Hanson & Morton-Bourgon, 2009).

Since the early 2000s parole in England and Wales has undergone important changes. In the first place, determinate-sentence prisoners have been taken out of the parole system. It is now the practice that determinate-sentence prisoners are automatically released halfway through their sentence, but remain under probation supervision and subject to recall to prison until the two-thirds point. Unless they are recalled, and subsequently apply for re-release, parole boards do not deal with them. Instead, UK parole boards are free to concentrate on indeterminate-sentence prisoners.

These include mandatory lifers (those convicted of murder, for whom the only penalty available is a life sentence), and discretionary lifers (those convicted of other serious offences, usually rape but sometimes violent robbery) for whom the judge may award either a determinate sentence or life sentence. In the last couple of decades, various laws have been passed requiring an indeterminate sentence if an offender is convicted of a second serious violent or sexual offence (Ministry of Justice, 2013). These have variously been termed automatic life sentences and indeterminate sentences for public protection. Essentially, these are life sentences with a minimum term determined by a sentencing formula which usually results in a short minimum, although there is no guarantee of release at the expiry of that minimum.

Indeterminate-sentence prisoners, and especially mandatory and discretionary lifers, may end up serving very long terms in prison (Ministry of Justice, 2013). As will be seen in Chapter 3, the sample of 100 lifers used for the research reported in this thesis contained some men who had served over 40 years. Clearly, this is extremely expensive for the taxpayer, regardless of any humanitarian concern for the prisoner. It would therefore be useful if it could be demonstrated that the decision to release prisoners or to retain them in custody was related to known risk factors. However, a search for evidence on this point proved unrewarding. There appeared to be no systematic review of the evidence relating parole decisions to risk assessment, although a number of published studies were found. The candidate found only one review of the literature (Caplan, 2007); this appeared not to be systematic, and restricted itself to North American literature. There was very little research on the relationship between recognised risk assessment instruments and parole decisions.

Another problem was that many of the published studies only concerned determinate-sentence prisoners, and others did not state whether they included indeterminate-sentence prisoners or not. This appeared to indicate a gap in the research which might be relevant to parole decisions in England and Wales, which are now almost exclusively taken in the cases of indeterminate-sentence prisoners. Furthermore, those studies which were found did not examine the relationship between formal risk assessment instruments and parole decisions. Again, in lifer parole hearings it is routine that professionals, such as psychologists and probation officers, make recommendations as to whether a lifer should get a progressive move. However, there were few if any studies which examined the relationship between formal risk assessment instruments and the recommendations made by professionals. This was particularly of concern because several studies (e.g., Bradford & Cowell, 2012; Heinz, Heinz, Senderowitz & Vance, 1976; Morgan & Smith, 2005a) had been found indicating an influence of professional recommendations on parole board decisions. If those professional recommendations themselves did not relate to known risk factors, this would suggest a major shortcoming in the whole process.

From an initial consideration of the published research available it was decided it was necessary to carry out a systematic literature review concentrating on empirical studies of the parole decision-making process. Initially, it was intended to include studies published in the last 20 years. However, during the search it became apparent that the number of empirical studies (as opposed to discussion, historical or policy papers) was very small. Accordingly, the time period was extended back to the mid-1960s, a period of approximately 50 years. This review of the literature is reported in Chapter 2 of this thesis, and is believed to be the only one of its kind.

When it came to investigating the parole process for lifers in the UK, the candidate had a potential source of data available from his day-to-day work. A large part of this work consisted of writing risk assessment reports on life sentence prisoners who were applying for parole. In a typical year approximately 50 of these would be undertaken, suggesting that a sample of 100 could reasonably be collected within the time limit imposed by a doctoral programme. This proved to be the case, and a study was undertaken relating the scores obtained on formal risk assessment instruments to both the parole recommendations of professionals, and the actual parole decisions which resulted. Again, it is believed that this is the only such research study ever undertaken in the UK. There have been some more specific studies, relating parole decisions to such things as psychopathy, as measured by PCL-R scores (Porter, ten Brinke, & Wilson, 2009) but these were not conducted in the UK and did not relate specifically to lifers. This study is reported in Chapter 3.

Chapter 4 considers one of the risk assessment instruments most frequently used in lifer parole assessments. Indeed, if it has not been carried out such an assessment is often requested by the Parole Board. This is the Psychopathy Checklist-Revised, or PCL-R (Hare, 2003b), currently one of the most widely used measures of psychopathy (Parent, Guay, & Knight, 2011) . This chapter examines the characteristics of the PCL-R and considers whether its pre-eminent position in psychopathy assessment is justified, and whether it is a useful measure of risk. In particular, consideration is given to the influence of clinical judgement upon PCL-R scores. This is undertaken in the light of some of the findings reported in Chapters 2 and 3.

Chapter 5 is a general discussion of the findings obtained, including a consideration of the utility of clinical judgement in risk assessment. It also considers whether the risk

assessments obtained by professionals are a useful component of the parole decision-making process, and illustrates some of the issues by reference to the critique of the PCL-R undertaken in Chapter 4. Finally, it considers implications for current practice, including whether the functions of the Parole Board should be modified. It also considers the ethical implications for psychologists involved in making recommendations based on risk assessment, and the development of a possible risk assessment scheme aimed purely at lifers.

CHAPTER 2: WHAT DETERMINES PAROLE DECISIONS? A SYSTEMATIC REVIEW OF 29 PAPERS INVOLVING 20,568 CASES.

Abstract

This chapter reports a systematic literature review of research into the factors associated with parole decisions. Between 1966 and the present 29 published papers were identified, reporting data from 20,568 cases. Online discussion groups, the Parole Board, the Home Office, the Ministry of Justice and prominent academics were asked to provide unpublished studies, but none were found. Most studies used correlations and regression analysis, and were therefore cross-sectional in nature. Parole applicants' criminal histories, index offence characteristics, current sentence characteristics, personal characteristics, community-related factors, staff reports, risk rating scales, and individual parole board member characteristics were all found to correlate with parole decisions. Parole decisions were largely predictable from a few variables, often characteristics of prisoners' offences or recommendations by professionals. However, different correlates were found by different studies in different jurisdictions, which limited the generalisability of findings. All jurisdictions emphasised the importance of taking parole decisions so as to minimise risk to the public. However, risk assessment was subjective and not evidence-based. Known risk factors were frequently not related to parole decisions, which often depended on factors known to be unrelated to reconviction risk, or risk of failing parole. Parole boards did not always state which factors formed the basis of their decision-making. Where they did, these factors did not

always relate to the actual decisions. The probable effect of this is that many low risk individuals are retained in custody unnecessarily. This is consistent with findings from earlier research on determinate-sentence prisoners.

Introduction

There is a good deal of evidence that parole can be a contentious issue. For prisoners serving indeterminate sentences it is their only route out of prison, and although the stakes are lower for prisoners serving determinate sentences parole can still shorten their sentence considerably. Concerns have been expressed about the criteria being used for granting parole, and whether there are criteria which are not recognised; for example, in Canada a bias against aboriginal ethnic groups has been alleged (Huebner & Bynum, 2008). There has also been a tendency towards increasing punitiveness in the Western world, which has reduced the willingness to grant parole (Griffin & O'Donnell, 2012; Zinger, 2012). This is despite the fact that parole applicants with the most serious offences are serving the longest sentences, are generally older and have the lowest reconviction rates (Liem, 2013). It is therefore legitimate to ask which criteria determine the decisions taken by parole boards, and whether they are aware of them. With this in mind, a systematic literature review was undertaken with the aim of clarifying which factors have been found to influence parole decisions in empirical studies, and what the implications are for policy, especially in England and Wales. Previously, only one review appeared to have been published in this field (Caplan, 2007). However, this was not a systematic review and restricted itself to North American research. It was helpful mainly in that it cited earlier studies which provided an introduction to the field. These were duplicated in the library searches reported below, and the original publications were read for inclusion in this review.

The following problems were found in conducting a systematic review of this topic:

1. Parole systems differ between jurisdictions. There are 50 different state parole systems in the United States, for example, as well as a federal system. Jurisdictions differ widely in the typical amount of time devoted to considering each case, which itself makes the process fundamentally different.

2. Parole systems change over time, even within the same jurisdiction. This may be partly because of political pressures and partly because of legislative changes. For example, since 2008 conditional release has been automatic for most prisoners serving determinate sentences in England and Wales (Ministry of Justice, 2013b). Parole decisions are therefore mainly required for those serving indeterminate sentences.

3. There are fundamental methodological differences between different studies: some have excluded particular categories of prisoner, some have focused on the decision-making process rather than the prisoner, and yet others have studied simulated rather than real-life parole decisions.

4. Even when the same basic concept is studied, there are differences in the measures used. For example, several studies have examined previous criminal history, but have operationalised it in different ways.

5. Details of measures have not always been reported. For example, when statistical prediction methods have been used the method of calculating these has not been given.

These differences made it difficult to conduct a statistical analysis of the evidence, such as a meta-analysis. However, it was still possible to conduct a systematic literature search, to find those factors which have been shown to influence parole decision-

making, and to discuss the applicability of these findings to current practice in England and Wales.

Method

The University of Birmingham Library has a search engine called *FindIt@Bham,* which searches all of the databases to which the Library has access. These include *BioMed Central, EMBASE, Europe PubMed Central, MEDLINE, Proquest Social Science Abstracts, PsycARTICLES, PsycINFO,* and *SwetsWise.* A search of *FindIt@Bham* was conducted using the search terms "parol*" and "prison*", the asterisks being "wildcards" which would identify any variants of those terms. This returned 6,659 results. From a search result *FindIt@Bham* derives a list of further terms and presents them to the researcher who can rule them in or out. The following terms were defined as admissible: prison, prisoners, parole, life imprisonment, prisoners parole England reports, and England life-sentence prisoners. The following terms were defined as inadmissible: any terms referring to French literature ("parole" being the French for "word"), state correctional institution, terms relating to political movements, substance use disorder, mortality, epidemiology, Public Health, methadone, women, mental disorders, substance abuse, substance-related disorders. This reduced the list to 296, for which abstracts were obtained. These abstracts were read and any studies excluded which did not deal specifically with the determinants of parole decisions. If the abstract did not make the design of the study unambiguously clear, the full text was retrieved and a decision taken after consulting that. This reduced the list of items to 25.

A similar search was carried out using the American Psychological Association's *APA PsycNet® Gold* service, which gives access to *PsycARTICLES, PsycBOOKS, PsycEXTRA, PsycINFO, and PsycCRITIQUES*. This returned 974 results; many of these

were duplicates of the University of Birmingham Library search. Search results returned by *APA PsycNet® Gold* were filtered by further search terms ("parol*" and "prison*"). This reduced the list to 172 studies all of which were determined by EndNote software to be duplicates of the University of Birmingham Library search. A further search was carried out using the British Psychological Society's *EBSCO* search portal, which gives members access to the *EBSCO* Behavioural Sciences and Law Collection. Using the search terms "parol*" and "prison*", the search returned 272 items; most of the suggested filter terms were irrelevant as a basis for excluding studies. However, selecting peer-reviewed studies only and including all crime-oriented terms except "books-reviews", "substance abuse treatment", "drug abuse" and "therapeutics" reduced the number to 171. Most of these were duplicates, but a further two articles were found.

Several prominent academic researchers, including Kevin Douglas, David Farrington, Louise Bowers and Laura Guy, were contacted, asking if they knew of any relevant studies, published or unpublished (see Appendix 3). Although they replied suggesting various published studies, none of these were additional to those already found, and no unpublished studies were identified. Several expressed the opinion that the available literature was very small. A similar request was sent to the Parole Board for England and Wales. The Board's Director of Business Development suggested two studies, both of which had already been discovered, and also stated that there was little published research in this field.

A similar request was made on two Internet-based discussion groups for forensic practitioners: the International Discussion on Psychology and Law (PsyLaw), hosted by the University of Nebraska, whose membership is global, though mainly North American, and the British Psychological Society's Forensic Practitioners' Forum,

whose membership is mainly UK-based. Neither approach yielded any new material. Finally, a search was made of the websites of the English Home Office (http://www.homeoffice.gov.uk/) and Ministry of Justice (http://www.justice.gov.uk/). Two further research studies were found which had not been listed elsewhere. The final set of studies was 29 documents, all concerned with the predictors and determinants of parole board decisions. A further three comment or discussion papers were found which were judged helpful for their background information, though they did not contribute data to the review. A flowchart illustrating the selection process is shown in Figure 1 (page 13). Example printouts demonstrating electronic searches and their results are shown in Appendix 1.

Quality assessment

Many reviews create a numerical rating of the quality of each study, often the sum of a series of ratings of different aspects of quality, such as the representativeness of the sample and the rigour of experimental design. However, such ratings are subjective and may serve to introduce a spurious conformity between studies (Greenland, 1994; Greenland & O'Rourke, 2001; Jüni, Witschi, Bloch, & Egger, 1999), while a widely recommended reference work on systematic reviews (University of York Centre for Reviews and Dissemination, 2009) describes the practice as "questionable and not recommended" (page 43), because of the poor validity of such scales. Nonetheless, quality clearly cannot be ignored: one criticism of many reviews is that they fail to exclude poor quality studies and produce results which artificially favour treatment interventions more than they should (Khan, Daya, & Jadad, 1996). Although this review did not concern treatments, the same general point applies. A further point is that quality assessment ratings should be easy to communicate to scholars and policymakers

Figure 1: Flowchart illustrating the study selection process

FindIt@Bham initial search using parol* and prison* search terms (6,659 results)

APA PsycNet® Gold additional search using parol* and prison* search terms (974 results)

BPS-EBSCO additional search using parol* and prison* search terms (275 results)

Additional requests made to prominent professionals, Home Office online archive, Parole Board, and online lists (4 results)

Filter using database-suggested filter terms (296 results)

Filter using database-suggested filter terms (172 results)

Filter using database-suggested filter terms (272 results)

Are these duplicates of existing results?

NO

YES

Examine abstracts: Are studies empirical, as opposed to discussion or policy documents? (302 results)

NO

YES

Retain (29 results)

Exclude (273 results)

who may wish to act on the results (Farrington, Gottfredson, Sherman, & Welsh, 2006).

They cannot readily do this if they do not understand them.

The studies reviewed here varied considerably. Some were simply correlational studies (Heinz, Heinz, Senderowitz, & Vance, 1976; Scott, 1974), while others included regression analysis (Holland, Holt, & Brewer, 1978), or discriminant analysis (Bonham, Janeksela, & Bardo, 1986; Parsonage, Bernat, & Helfgott, 1994; Winfree, Sellers, Ballard, & Roberg, 1990) and others performed group comparisons (Hoffman, 1994; Stone-Meierhoefer & Hoffman, 1982). These were all "real world" studies, but others were simulations (Turpin-Petrosino, 1999). This range of approaches did not lend itself to a single method of scoring for quality. Instead, a method of placing studies in a series of categories, graded for methodological rigour, was sought. With this in mind it was decided to adopt the Maryland Scientific Methods Scale (Farrington et al., 2006), which distinguishes five levels of methodological rigour:

1. **Correlational.** This can establish a statistical relationship between dependent and independent variables, but not a causal relationship.

2. **Before and after measures in a quasi-experimental design without a control group.** This can establish that a change occurred, but not why.

3. **Before and after measures in an experimental design with a control group.** Farrington et al. suggested that this was the minimum interpretable standard, although it was still subject to selection effects if the experimental and control groups were not truly comparable.

4. **Similar to 3, but conducting studies in multiple sites and controlling for other known influences on the dependent variables.**

5. **Randomised controlled trials (RCTs).** Farrington et al. considered these to offer the highest level of methodological rigour, while noting that they were relatively uncommon in studies of criminal behaviour. They also suggested that RCTs could be downgraded to level 4 if there was reason to suspect the adequacy of the randomisation.

Of the 29 studies covered by this review, 15 were correlational, reporting either correlations or regression analyses. The others used Cox proportional hazard techniques, statistical modelling, discriminant analysis, purely descriptive statistics or qualitative techniques. All of these, like correlational techniques, would be rated at level 1 in the Maryland Scale. Only three studies (Hoffman, 1994; Porter et al., 2009; Stone-Meierhoefer & Hoffman, 1982) utilised group comparisons and might reasonably be rated at level 3. Therefore, the majority of studies reviewed would be considered uninterpretable by the Farrington et al. (2006) criteria.

The position taken here is slightly less pessimistic. It is true that correlation does not necessarily imply causality, but causality certainly implies correlation. For example, if parole decisions were found to correlate with a risk assessment measure this would not necessarily prove that the parole board was using it as the basis for their decisions. However, if they did base their decisions on it, then a correlation would have to occur. Failure to find such a correlation would seriously undermine any claim to be using that measure, and would thus provide useful information, albeit of a negative kind[1]. That said, it is clear that the level of methodological rigour found in this field has been low.

---

[1] An exception to this would occur if there existed a mediating factor, so that parole decisions were influenced by factor X, which was influenced in turn by factor Y. In that case one would not expect parole boards to *claim* factor Y as a basis for their decisions, but research might demonstrate that it was.

The studies which were reviewed are described in Appendix 2, which includes a description of each study, the sample used, and comments on methodology.

Analysis strategy

Initial analysis indicated that the different studies had examined different possible influences on parole board decisions. However, in practice there was a limited range of variables because most studies used information gathered from official records, and jurisdictions did not differ greatly in the type of information which was recorded. For example, most jurisdictions record information about a prisoner's previous criminal history and this usually consists of the number of offences in each of several commonly used categories (i.e., sexual, violent, burglary and robbery). Similarly, variables relating to the current sentence generally consisted of the length, the amount of time served, and the type of offence for which it was imposed. Consultation of these variables led to the conclusion that the potential determinants of parole decisions could be grouped into eight sets of factors. These were:

1. Parole applicants' criminal histories (largely, types of previous offence)

2. Current sentence variables (sentence length, and type of offence)

3. Applicants' personal characteristics (educational and social variables)

4. Community-related factors (notoriety, victim statements)

5. Professional and correctional staff reports (e.g., prison officers, governors, probation officers and psychologists)

6. Risk rating scales (statistical predictors of reconviction or parole failure)

7. Individual parole board member characteristics (gender)

8. Cognitive processes in parole decision makers (mental depletion)

The strategy adopted was to consider each of these in turn, recognising that a number of studies have considered more than one set of factors. Tables below list the studies and the factors examined in them. Each of the items listed above will now be examined in detail.

1. Parole applicants' criminal histories

Table 1 shows the studies which examined parole applicants' previous criminal histories and those which examined the current (index) offence. Offences differed in type, severity, frequency, total number, and were often classified differently in different jurisdictions. Several studies either combined criminal history data into one measure, or used a specific aspect, such as seriousness or number of offences.

Table 1: Studies which examined parole applicants' criminal histories	
Variable	**Studies**
Previous criminal history	Bonham, Janeksela & Bardo (1986), Heinz, Heinz, Senderowitz & Vance (1976), Hood & Shute (2000), Morgan & Smith (2005a), Nuttall (1977), Scott (1974)
Nature of the index offence	Bradford & Cowell (2012), Heinz et al (1976), Holland, Holt & Brewer (1978), Huebner & Bynum (2006), Morgan & Smith (2005a), Scott (1974), Turpin-Petrosino (1999), Welsh & Ogloff (2000)

1.1 Previous criminal history

An early attempt to combine criminal history into a single variable was made by Scott (1974), working in an un-named "Midwestern state". The sample consisted of 359 male

17

and female prisoners. Scott developed a scale based on weighting previous numbers of arrests and convictions for misdemeanours or felonies and previous prison sentences. Four alternative scales were tested, but the all intercorrelated highly. The final choice had a small relationship with subsequent parole decisions; surprisingly, the more serious the offenders' histories, the *more* likely they were to obtain parole. Scott considered this to be an artefact of the American legal system, because offenders were often paroled when wanted in a neighbouring state. This solved the problem for the paroling state, but ensured the offender would not actually be freed. Scott presented field observations of parole board discussions indicating that the expense of housing offenders was one consideration.

Heinz et al. (1976) found a highly significant but modest correlation ($\tau_c=.28$) between the number of previous convictions and parole decision. The sample consisted of 294 prisoners in the state of Illinois, 95% of whom were male. Other measures of previous criminal involvement did not relate to the parole decision, but of 32% without previous convictions, more than 90% were paroled. However, even those with more than three previous convictions had a 58.8% success rate in obtaining parole, suggesting that previous record was an important, but not critical, factor.

In a UK study of 1682 parole decisions, Nuttall (1977) found that 68% of those with fewer than three previous convictions were paroled, but only 25% of those with three or more. Nuttall did not investigate the influence of the type or seriousness of previous offences, or previous adult prison sentences. Previous juvenile custody was found to reduce the likelihood of parole.

Bonham et al. (1986) studied a sample of 532 parole applicants in Kansas. It was Kansas policy at the time that the parole board should consider prior criminal record as

one of its criteria for granting parole. However, the authors found that the parole board made no systematic attempt to evaluate criminal history, and the parole decision was not related to the applicant's history of violent offending or the number of prior offences.

Hood and Shute (2000) studied the procedure then current in England and Wales. At that time, a three-member parole panel would consider parole applications after a preliminary interview by a single panel member. This "lead" member would make a recommendation, which the full panel followed in over 80% of cases. The parole decision was correlated with the total number of previous convictions, previous sexual or violent convictions specifically, previous custodial sentences, and previous breaches of supervision or bail. Although the Home Office had developed a statistical prediction scheme for recidivism panels did not have access to it. However, their decisions were significantly correlated with it. Sex offenders, whatever their level of risk, were paroled at a much lower rate than other prisoners. Morgan and Smith (2005a), in a study of 762 violent male offenders in Alabama, used a logistic regression technique to relate various criminal history measures to parole decisions, but no relationship was found.

1.2 Nature of the Index Offence

The nature of the index offence (the one which led to the current sentence) is an obvious possible determinant of parole decisions. Reference has already been made to studies showing that sex offenders were unlikely to gain parole whatever their history (Hood & Shute, 1999, 2000). The same study also found that the Parole Board greatly overestimated the risk of reconviction for incest offenders specifically, despite their being the sex offender group with the lowest reconviction rate (Hanson, 2001; Hanson

& Bussière, 1998). However, it is not just the category of offence which may be important, but other characteristics, such as the severity of injury in a violence offence.

Scott (1974) found that the principal determinant of parole decisions was the seriousness of the index offence, as measured on a scale he developed himself (see previous section). However, Heinz et al. (1976) found that the seriousness of the index offence, as rated subjectively by the researchers themselves, was weakly related to parole decisions. Their subjective rating may be doubtful because it was negatively correlated with sentence length, which presumably reflects the court's opinion of seriousness. Holland et al. (1978) found that parole boards focused mainly on the seriousness of the index offence, while stating that they did not. This was one of the earlier studies to suggest that board members rationalised decisions they had taken, without being aware of the true reasons why they had taken them.

Turpin-Petrosino (1999) conducted a simulation of parole decisions, in which 300 case scenarios were presented to each of 10 experienced New Jersey Parole Board members. The study followed a change in the law, which required prisoners to be paroled as soon as they became eligible, unless there were good grounds to suspect a high risk of recidivism. The experimental scenarios were designed not to give any legally-defensible grounds for refusing parole. Turpin-Petrosino found that significantly more parole applications were refused if the index offence contained aggravating circumstances (a reason specifically excluded under the new law), and if the index offence was violent or sexual. Parole Board members were also more likely to refuse parole if they regarded the original sentence as too lenient (another reason excluded by law). The study found that when board members refused parole they would emphasise the importance of certain items of information, not objectively related to risk, whereas

when they granted it they rarely cited any special reason. As in the Holland et al. (1978) study, the board members were taking a decision about parole on subjective grounds, but rationalising it by reference to objective information. The change in the law appeared to have had no effect.

In a study of 2,479 parole-eligible male prisoners in Canada, Welsh and Ogloff (2000) found that parole was more often declined if the index offence was robbery, or a sex offence. In a sample of American sex offenders, Huebner and Bynum (2006) found that applicants were less likely to be paroled if the victim of their index offence was a child under 10 years old.

Bradford and Cowell (2012) reported an initial study of 255 UK prisoners who had received "indeterminate sentences for public protection". These are indeterminate sentences imposed for a second sexual or violent offence (Ministry of Justice, 2013b). Although all of these offenders had committed serious offences, release rates for (nonsexual) violent offenders were more than eight times as high as those for sexual offenders.

1.3 Discussion: Parole applicants' criminal histories

Taken together, these results suggest that criminal history is often, but not always, a factor in parole decision-making, and that part of the variation in results is due to differences in measurement. Studies of history prior to the current offence have used the total number of prior offences, the number of serious prior offences (defined in various ways), or previous parole performance. Furthermore, the studies have almost invariably been correlational. However, characteristics of the index offence have often correlated with the parole decision. In particular, even when samples consist solely of people with

serious offences, there was a distinct bias against granting parole to those who had committed sex offences, despite considerable evidence that their reconviction rates for further sexual offending are generally low (Hanson & Bussière, 1998; Hanson, Morton, & Harris, 2003). It is not clear from these studies where the correlation between offence type and parole decision arises. There may be a causal relationship, but if so it is not clear whether parole boards are reacting directly to the index offence, or to some intermediate factor such as prison or probation reports.

2. Current Sentence Variables

Table 2 shows studies which examined whether parole decisions are related to characteristics of the current *sentence*, as opposed to the offence. In particular, studies have looked at the sentence length (or the proportion of it which has already been

Table 2: Studies which examined current sentence variables	
Variable	**Studies**
Current sentence length	Heinz et al. (1976), Morgan & Smith (2005a), Nuttall (1977), Scott (1974), Welsh & Ogloff (2000)
Time already served	Bonham, Janeksela & Bardo (1986), Caplan (2007), Morgan & Smith (2005a)
Institutional misconduct reports	Bonham et al (1986), Heinz, Heinz, Senderowitz & Vance (1976), Holland, Holt & Brewer (1978), Huebner & Bynum (2006), Huebner & Bynum (2008), Morgan & Smith (2005b), Welsh & Ogloff (2000)
Completion of institutional programmes	Bonham et al (1986), Bradford & Cowell (2012), Heinz et al (1976), Hood & Shute (1999, 2000), Welsh & Ogloff (2000)
Denial of offence	Hood, Shute, Feilzer & Wilcox (2002)

served), and the institutional conduct of the offender, as evidenced by such things as disciplinary reports and completion of offending behaviour programmes.

2.1 Current sentence length

Scott (1974) found that those with a history of disciplinary infringements served more time before gaining parole. The study did not ascertain whether the parole decision was taken because of the disciplinary infringements, or some other factor which might have been a correlate (e.g., age or criminal history). However, little time was spent on considering the various factors, since more than 100 parole cases were heard in a day and the median time spent considering each case was eight seconds. Parole board members had studied the case papers in advance, and agreed a decision instantly in many cases, discussion taking place where there was disagreement; the longest case took 55 minutes.

Nuttall (1977) in the UK and Welsh and Ogloff (2000) in Canada found that those serving shorter sentences were more likely to gain parole. In Nuttall's study parole was twice as likely to be granted if the current sentence was less than three years. Morgan and Smith (2005a) found that the length of the current sentence was the third most influential factor out of four. Of course, given that courts impose longer sentences for offences seen as more serious, sentence length could be a measure of perceived seriousness. However, in Illinois Heinz et al. (1976) found no relationship between sentence length and parole decisions.

2.2 Time already served

Bonham et al. (1986) found that Kansas prisoners were more likely to be granted parole the closer they were to completing their sentence. Morgan and Smith (2005a) found the

same correlation in Alabama, but it disappeared in their logistic regression analysis. In a review of North American studies, Caplan (2007) reported that this correlation was a common finding, which makes sense if parole boards are trying to limit the time during which parolees are at risk for reconviction.

2.3 Institutional misconduct reports

Heinz et al. (1976) found that parole was less likely if applicants had committed serious disciplinary infractions, as did Bonham et al. (1986). Holland et al. (1978) found that prison disciplinary reports contributed strongly to the recommendations of prison caseworkers, but not to the parole decision, nor to reconviction among those who were subsequently paroled. Morgan and Smith (2005a) found that the time served since the last disciplinary report correlated with the parole decision whereas the number of reports did not. However, in a logistic regression analysis, neither of these related to the parole decision.

Prison disciplinary reports were also considered by Welsh and Ogloff (2000), although their primary concern was possible ethnic bias against aboriginal Canadians. Theirs was the only study to break down institutional offences into categories, only one of which (fighting) was associated with the parole decision. The other categories were: (1) escape attempts, (2) drug use, (3) contraband, (4) damaging property, (5) causing a disturbance, and (6) other infringements. Huebner and Bynum (2008), examining possible ethnic bias against African-Americans, also found that a history of disciplinary infringements reduced a prisoner's chances of parole (by about 13%). This confirmed their finding in an earlier study of sex offenders (Huebner & Bynum, 2006). Neither study broke institutional infringements down by type.

2.4 Completion of institutional programmes

Heinz et al. (1976) found that parole was more likely if prisoners had enrolled in educational programmes; offending behaviour programmes were not available at that time. Bonham et al. (1986) found that successful completion of offending behaviour programmes rendered parole more likely. Hood and Shute (1999, 2000) found that prisoners were more likely to be paroled if they had completed the offending behaviour programmes recommended by their prison-based probation officer. It was not clear whether this was thought to indicate compliance or reduced risk. Surprisingly, Welsh and Ogloff (2000) found that sex offenders were *less* likely to be paroled if they had completed a sex offender treatment programme. The authors speculated that this was because further treatment needs had been identified during the programme. In their qualitative study of parole proceedings for indeterminate sentence prisoners, Bradford and Cowell (2012) underlined the importance of offending behaviour programmes, which were seen by parole board members as a means of attempting to change offending behaviour. Board members complained that suitable programmes were not always available, and stressed the importance of spending time in open prison conditions prior to release.

2.5 Denial of the index offence

Denial of the index offence may cause difficulties for parole boards, because denial often precludes participation in offending behaviour programmes, giving rise to the perception that risk has not been reduced (Hood, Shute, Feilzer, & Wilcox, 2002). Even among forensic psychologists, attitudes to denial vary widely, some regarding denial as a major barrier to early release (Freeman, Palk, & Davey, 2010). This is despite the

repeated finding that denial of one's offence is not associated in any simple way with increased risk of reconviction, at least for sex offenders, the group most likely to occasion anxiety, and may even be protective for some offenders (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Harkins, Beech, & Goodwill, 2010). Only one study specifically addressed the issue of denial and parole, and this was in a sample of 162 sex offenders in England and Wales (Hood et al., 2002). Risk was overestimated by the Parole Board for those who denied their offence. Indeed, the Board overestimated risk for the whole group: of those who did not reconvict with a sexual offence within four years, the Board had estimated that 92% would do so, a very high false positive rate. For a subgroup followed up for six years the false positive rate was 78%. The authors noted that the Static-99 risk assessment instrument was more accurate than the Parole Board; however, the latter, in overestimating risk, also prevented release for more true positives.

2.6 Discussion: Current sentence variables

These studies suggest that parole boards believe behaviour in prison is predictive of post-release behaviour, and that offending behaviour programmes reduce the risk of reconviction. It is questionable whether the evidence supports either of these beliefs. For example, (Austin, Coleman, Peyton, & Johnson, 2003), in a study of the LSI-R risk assessment instrument, discovered that none of its measures of institutional behaviour related to subsequent risk. This is not a new observation: Scott (1974) cited research with the same outcome 40 years ago.

Evidence for the effectiveness of offending behaviour programmes has also been widely questioned. There has been considerable criticism of the methodology used in

programme evaluations (Rice & Harris, 2003), and methodologically strict evaluations have failed to find clear evidence of effectiveness (Falshaw, Friendship, Travers, & Nugent, 2003; Hanson, Bourgon, Helmus, & Hodgson, 2009; Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005; Quinsey, Harris, Rice, & Cormier, 2006; Seager, Jellicoe, & Dhaliwal, 2004). It is also not clear that parole boards act in accordance with a belief in effectiveness. For example, reluctance to parole sex offenders seems incompatible with a belief in the effectiveness of the programmes available, and one study found that programme completion made parole *less* likely.

3. Applicants' personal characteristics

A number of studies have examined the relationship between obtaining parole and background characteristics, such as IQ, educational level, personality, ethnicity and age

Table 3: Studies which examined applicants' personal characteristics	
Variable	**Studies**
Intellectual and social factors, and gender	Heinz, Heinz, Senderowitz & Vance (1976), Morgan & Smith (2005a), Scott (1974)
Ethnicity	Huebner & Bynum (2008), Welsh & Ogloff (2000)
Mental health issues	Bonham, Janeksela & Bardo (1986), Huebner & Bynum (2008), Matjekowksi, Caplan & Cullen (2010), Matjekowski, Draine, Solomon & Salzer (2011), Welsh & Ogloff (2000)
Psychopathy	Porter, ten Brinke & Wilson (2009)
Age	Heinz et al (1976), Huebner & Bynum (2006)
Offender statement/ presence at hearing	Smith, Watkins & Morgan (1997)

(see Table 3). Two have also examined whether the offender's written statement or presence at a parole hearing makes any difference to the outcome. Table 3 shows the studies which examined these personal characteristics of parole applicants.

3.1 Intellectual and social factors, and gender

Scott (1974) found that prisoners served less time the higher their socio-economic status, educational level and IQ. Married men were also released earlier, and so were female prisoners. Initially he found that non-white prisoners served longer, but after controlling for criminological variables this effect disappeared. Heinz et al. (1976) also reported that educational achievement favoured parole, and that parole was more likely to be granted if men were married, had dependants, and had employment plans. However, in interviews the parole board members said that marital status did not influence their decision, and expressed scepticism about employment plans. As with findings on the seriousness of the index offence (Holland et al., 1978), parole board members apparently did not always realise what influenced them. Morgan and Smith (2005a) also found that educational level correlated with the parole decision, but this relationship disappeared in their logistic regression analysis.

3.2 Ethnicity

In Canada, Welsh and Ogloff (2000) found that non-white ethnic groups suffered an apparent disadvantage in parole decisions, but that this disappeared when criminological differences were controlled for. Huebner and Bynum (2008) reported a similar finding in the US, but found that African-American prisoners remained at a disadvantage despite controlling for other factors. By contrast, Hispanic prisoners received parole *sooner* than either black or white men. The authors speculated that Hispanic prisoners

were perceived as less threatening than African-Americans, but did not consider why they should have appeared less threatening than white offenders.

3.3 Mental health issues

Personality and mental health issues have also been examined in several papers. Bonham et al. (1986) found that problems with either physical or mental health generally diminished the prospect of parole. A range of studies has established a positive relationship between alcohol abuse and violent offending (Quinsey et al., 2006, pp 108-113), so it was surprising that alcohol usage rendered parole *more* likely. Drug abuse, on the other hand, had the opposite effect. Unfortunately, it was not clear how these were measured: use of both substances was placed by prison staff in one of five categories (no use, addicted, habitual excessive use, episodic excessive use, and no history of use but was using at the time of the index offence). The authors did not say how they converted this information into a numerical scale. Welsh and Ogloff (2000) also found that personal or emotional needs rated on a Canadian prison service instrument called the Case Needs Identification Analysis were associated with a lower probability of parole in their logistic regression analysis. However, there was no information concerning the reliability or validity of this instrument, or its relevance to recognised mental disorders. This study was aimed at identifying any effect of ethnicity, and although non-white prisoners were more often assessed as having personal or emotional needs, they were also more likely to have successfully completed programmes intended to meet them. Ethnicity was also the primary focus of the American study by Huebner and Bynum (2008), but they additionally noted that mental health problems delayed the granting of parole.

In a sample of 403 New Jersey prisoners, about half of whom had been diagnosed with a mental disorder of some kind, Matejkowski, Draine, Solomon, and Salzer (2011) found that this had no effect on parole decisions. They noted that mental illness was associated with substance abuse disorders, antisocial personality disorder, and violence within the prison, but none of these affected the parole decision. However, on the same sample, Matejkowski, Caplan, and Cullen (2010) found that *severely* mentally ill prisoners were discriminated against indirectly, because they committed more disciplinary offences within prison, and this in turn led to a lower rate of parole. Severe mental illness was defined as "schizophrenia, schizoaffective disorder, psychotic disorder NOS (not otherwise specified), major depression, mania, or bipolar or mood disorder NOS" (Matejkowski, 2014, personal communication).

3.4 Psychopathy

Although psychopathy has been found to be a risk factor for reconviction (Edens, Hart, Johnson, Johnson, & Olver, 2000; Hare, 2003b; Quinsey et al., 2006), only one study of 310 adult male offenders was found which related it to the parole process (Porter & ten Brinke, 2009). Psychopaths, defined as those scoring 30 or more on the Psychopathy Checklist-Revised (PCL-R) (Hare, 2003b), had committed more violent and nonsexual offences than non-psychopaths. Despite these indicators of risk (confirmed by their poor post-release performance) they were 2.5 times more likely to obtain parole than non-psychopaths. The PCL-R provides scores on two factors: Factor 1, intended to reflect the core personality characteristics of the psychopath, such as the remorseless manipulation of others, and Factor 2, a socially deviant and criminal lifestyle. The finding that psychopaths are more likely to obtain parole is consistent with the ratings of

charm and manipulativeness which contribute to the Factor 1 score. However, parole performance actually correlated significantly with Factor 2, and Factor 1 scores were not significantly associated with the parole decision. This is counterintuitive, though consistent with recent findings that Factor 1 assessed "in the field" has no reliability, although better results have often been obtained by academic research groups (Boccaccini, Johnson, & Janke, 2008; Edens, Boccaccini, & Johnson, 2010; Miller, Rufino, Boccaccini, Jackson, & Murrie, 2011).

3.5 Age

Most studies have examined the offender's age at the time of applying for parole, but few have found a significant influence of age on the parole decision. Heinz et al. (1976) found that older prisoners were less likely to be granted parole. However, the relationship was not simple: those in the oldest group (over 35 years) had an increased chance, as did the youngest. The authors speculated that the youngest prisoners were thought to be more malleable, while the oldest might have reached the point of desistance. In their sex offender sample, Huebner and Bynum (2006) found that older men were less likely to gain parole, especially if they had young victims.

3.6 Offender statement/presence at hearing

Smith, Watkins, and Morgan (1997) found that in Alabama their sample of violent male offenders were more likely to obtain parole if they attended their parole hearing in person. They did not investigate the reasons why this might be so, but this finding is consistent with that of Porter et al. (2009), who found that psychopathic offenders influenced parole outcomes in their favour. The relationship between violent offending

and psychopathy is, of course, long established. It is not clear whether this finding would generalise to non-violent offenders.

3.7 Discussion: Applicants' personal characteristics

Individual characteristics of parole applicants seem not to have been extensively studied, and when they have the results have been contradictory. For example, mental health needs have been found to correlate with parole decisions in some studies but not others. The situation is complicated by the fact that mental health needs, and personal and emotional needs, are often poorly defined, and may be rated by prison staff without mental health training. Definitive findings in this area will require much more robust measures. Results for ethnicity have also been contradictory, and complicated by the fact that some factors such as institutional behaviour may correlate with both ethnicity and the parole decision independently. Results for age have likewise been contradictory, which is surprising, given the widespread finding that reconviction risk reduces with age (Barbaree, Blanchard, & Langton, 2003; Hanson, 2001; Porter, Birt, & Boer, 2001; Quinsey et al., 2006; Sampson & Laub, 2003).

The finding that psychopaths are more likely to be granted parole seems to be a clear result. Nonetheless, it is contradictory in that the apparent ability to talk one's way into parole was related to the antisocial behaviour of Factor 2 of the PCL-R rather than the core psychopathic personality of Factor 1. However, recent research indicates that ratings of Factor 1 "in the field" have low or zero reliability (Boccaccini, Johnson, et al., 2008; Edens et al., 2010; Miller et al., 2011; Mokros, Stadtland, Osterheider, & Nedopil, 2010). This would prevent a significant correlation with any other variable. One possible interpretation is that this effective presentation is not a personality

characteristic (as the PCL-R assumes) but a technique learned through repeated encounters with the criminal justice system. The study did not establish whether the superior performance of psychopaths in applying for parole was due to impression management during hearings, or during interviews with professionals who reported to parole boards, or both. However, impression management would be consistent with the finding of Morgan and Smith (2005b) that the offender's chances of gaining parole improved if he actually attended the hearing, rather than submitting a case in writing.

4. Community-related factors

This term denotes those factors not associated with the offender, the character of his offence, or his sentence. These are such things as the assessment made by parole boards of the offender's ability to reintegrate into society, public reaction to the offence, and victim impact statements. Table 4 shows the studies which examined these factors.

4.1 Employment prospects

Since earning an income legitimately is the desired alternative to many offences, and a risk factor for others, an applicant's employability is an obvious concern for parole

Table 4: Studies which examined community-related variables	
Variable	**Studies**
Employment prospects	Heinz et al (1976)
Perceived community attitudes	Bonham et al (1986)
Victim participation in parole hearing	Caplan (2007), Morgan & Smith (2005b), Parsonage, Bernat & Helfgott (1994), Smith et al (1997)

boards. However, it was only considered by one study (Heinz et al., 1976), which found the Illinois Parole Board's decisions to be correlated with the offender's family situation and future employment plans, although board members denied being influenced by these and expressed scepticism about parole applicants' employment plans.

4.2 Perceived community attitudes

Bonham et al. (1986) considered the parole board's understanding of community attitudes to the offender. These were rated during the parole hearing itself by one parole board member while another interviewed the prisoner. The ratings were done on an official form and points allotted for the prisoner's score on each of 12 short (3 to 5-point) rating scales. The four-point scale for community attitudes ranged from "3-much community support" to "0-community antagonistic to offender's return". Clearly, this left considerable room for subjective judgement on the part of the rater, and there was no attempt to establish interrater reliability. Nonetheless, this rating contributed significantly to the parole board's eventual decision.

4.3 Victim participation in parole hearing

Of all members of the community the one most concerned with the case would clearly be the victim of the offence. In a study in Pennsylvania, where victims could attend parole hearings, send written statements, and give evidence in person, Parsonage et al. (1994) found that victim statements were the most significant determinant of parole decisions. In cases where victim statements were presented parole was refused to 43% of applicants, but only 7% where they were not. The study did not consider whether there was any difference between the impacts of written statements and personal appearance at the hearing. Furthermore, it is possible that there were confounds between

victims' eagerness to make statements and other factors. For example, it might be that only victims of the most serious offences would wish to oppose parole, or conversely that seriously traumatised victims would not wish to face their attackers.

In an Alabama sample of violent offenders, Smith et al. (1997) found that only victim and offender variables affected the parole decision: parole was more likely to be granted if the offender attended the hearing, but less likely if the victim did. Furthermore, the personal presence of the victim was more effective than written representations. However, the authors pointed out that there was an initial screening procedure which preceded the parole hearing, and which they had not considered, so it was possible that other variables might have an impact at that stage. Two of the authors (Morgan & Smith, 2005b), later took the screening process into account as well, confirming the original findings with respect to parole hearings, but finding that other factors determined the decision at the initial screening stage.

4.4 Discussion: Community-related factors

Given the importance of employment as a means of avoiding crime it is surprising that only one study considered it directly as a possible factor in parole decisions. A number of studies have shown legitimate employment to be protective against future reconviction (Feder & Dugan, 2002), and as an item in the HCR-20 it is predictive, albeit of theft and reoffending generally rather than violence (Coid et al., 2007). It may have been taken into account indirectly via the reports of professionals such as probation officers, but no study examined this possibility. Community attitudes were explicitly taken into account in only one study, using a highly subjective measurement. It is also clear that in at least two American jurisdictions victim statements had a

measurable impact on the parole process. This may not generalise to England and Wales, where victim impact statements are usually made at the trial rather than the parole hearing, although there is provision for probation officers to liaise with victims before parole hearings. Given that written representations from victims had less impact than personal appearance, it seems likely that second-hand representations of the victims' views would be even less influential.

5. Professional and correctional staff reports

In the Republic of Ireland the Irish Parole Board recommends whether to grant parole to life sentence prisoners, and the Ministry of Justice takes the final decision, accepting the Parole Board's recommendation in 87% of cases, (Griffin & O'Donnell, 2012). The authors did not examine which factors determined the Parole Board's decision itself, such as reports from professional and correctional staff, or single Board members who "pre-interview" offenders before the hearing. Table 5 shows the studies which examined these influences.

Table 5: Studies which examined professional and correctional staff reports	
Variable	**Studies**
Probation reports	Bradford & Cowell (2012), Heinz et al (1976), Morgan & Smith (2005a)
Senior prison staff reports	Morgan & Smith (2005a), Morgan & Smith (2005b), Winfree, Sellers, Ballard & Roberg (1990), Zinger (2012)
Parole readiness assessment	Bonham et al (1986), Bradford & Cowell (2012)
Lead parole board member's opinion	Hood & Shute (1999, 2000)

5.1 Probation reports

Heinz et al. (1976) found that the Illinois Parole Board's decision correlated (τ_c=.38) with the recommendation of the "correctional sociologist" (seconded probation officer), granting parole to 25% of cases where the recommendation was unfavourable, but 96% where it was favourable. The researchers questioned the wisdom of effectively allowing the parole board decision to be determined by these officers, whom they believed to be insufficiently trained and experienced.

Hood and Shute (2000) found that probation officers' recommendations correlated strongly with parole board decisions, but nonetheless 40% of those refused had been recommended for parole by both external and seconded probation officers. Bradford and Cowell (2012), in their study of 255 indeterminate-sentence prisoners, found that the recommendation of the offender manager (home probation officer) appeared to carry considerable weight. Of those who were moved from closed prison to open conditions 88% had been recommended for this by the offender manager, and for those who remained in closed conditions the offender manager had recommended this in 76% of cases. The researchers did not investigate the situation where opposing recommendations were made by different professionals for the same prisoner. Despite mostly following the offender manager's recommendation, in a qualitative section of the study parole board members complained that offender managers often based their reports largely on second-hand information culled from the reports of other professionals and telephone interviews with the prisoner. There is a clear contradiction in this.

5.2 Reports from senior prison staff

Several studies have examined the influence of reports from prison staff. Winfree et al. (1990) examined a special class of parole applicants in Texas. These were 208 prisoners who were reapplying for parole after being paroled once and having their parole revoked. This followed a change in the law intended to reduce overcrowding in Texas prisons. Under the new law, prisoners whose parole had been revoked for "technical" reasons rather than a new conviction were supposed to be paroled again at the earliest opportunity. Both before and after the legislation, the principal determinant of parole decisions was a recommendation received from the Hearings Section of the Department of Justice. After the change in the law the Hearings Section recommended more prisoners for parole but the Bureau of Paroles and Pardons took less notice, the overall result being about the same.

Morgan and Smith (2005a, 2005b) found that in Alabama professional reports had little influence on parole decisions. Prison staff reports, on the other hand, had a considerable influence. The authors constructed several different logistic regression models to predict parole decisions, the best fit being one which incorporated just four variables. The most significant of these was the recommendation of a senior corrections officer. The second most significant was the recommendation of the warden (prison governor). The third and fourth were sentence length, and number of felonies committed in the index offence.

Zinger (2012) noted that Canadian parole decisions were strongly influenced by the recommendations of correctional authorities. In Canada the procedure is that the Correctional Service of Canada (CSC) provides information to the Parole Board Canada (PBC) on the basis of which the Board makes its decision. Zinger noted that the CSC

almost always provides a recommendation as to what that decision should be, and the PBC nearly always follows that recommendation. The CSC is not legally required to make a recommendation, and Zinger questioned whether it should. He suggested that, under the influence of political trends, Canada was moving away from an evidence-based programme of conditional release towards harsher penalties for offenders.

5.3 Parole readiness assessment

Bradford and Cowell (2012) found that parole was more likely to be granted if an offender had no drug or accommodation problems and there was a "robust" risk management plan. Unfortunately, there was no definition of what constituted "robustness", and it appears to have been a subjective assessment by the Parole Board. This was similar to an earlier finding in Kansas (Bonham et al., 1986) where the quality of the parole plan was said to be an important determinant of successful parole applications. This was rated in a similar way to the ratings of community attitudes described above (see section 4.2). There was a four-point scale ranging from "3-Strong parole plan" to "0-no skills, no plan, no interest". As with the community attitudes scale, there was no indication of the reliability or validity of the scale.

5.4 Lead parole board member's opinion

One UK study (Hood & Shute, 1999, 2000) examined the influence of a "lead" parole board member who interviewed the applicant some time ahead of the hearing and reported to the full board. The lead member almost always gave a recommendation, which was confirmed by the full board in 80% of cases. Although the board members had available an objective assessment of the risk of serious reconviction during the parole period, they overestimated risk, particularly where the index offence was sexual.

A study in Kansas (Bonham et al., 1986) found that a board member would make a subjective overall assessment of recidivism risk. This was related to the board's eventual decision, but there was little indication as to how the board member arrived at this assessment.

5.5 Discussion: Professional and correctional staff reports

With respect to probation and prison staff reports there is evidence that parole boards have been influenced by them, but it is not clear what influenced those reports in the first place. This is concerning when one considers the earlier evidence cited above, showing that reports were often influenced by factors that have little or no relationship with parole success or recidivism, such as institutional disciplinary infractions. Generally it appears that parole boards are expected to give considerable weight to the risk of recidivism, but the risk estimates available to them are largely clinical, not to say subjective, in nature. The literature is clear that these are not likely to be accurate (Hanson & Morton-Bourgon, 2009; Meehl, 1954, 1973; Philipse, Koeter, van der Staak, & van den Brink, 2006; Quinsey et al., 2006; Quinsey & Maguire, 1986).

6. Risk rating scales

Two kinds of scale have been used to help parole boards make their decisions. One concerns the risk of parole failure, and the other the risk of reconviction. The distinction is important because most parole failures occur for technical reasons rather than reconviction (Vito, Higgins, & Tewksbury, 2012; Winfree et al., 1990). These reasons include failure to keep appointments with the parole officer, or behaviour which is thought to indicate risk. The latter might include associating with known criminals, or heavy drinking where the original offence involved alcohol. Table 6 shows the studies

which examined the use of risk rating scales by parole boards.

Table 6: Studies which examined risk rating scales	
Variable	**Studies**
Risk of parole failure	Hoffman (1994), Hood & Shute (1999, 2000), Pogrebin, Poole & Regoli (1986)
Risk of recidivism	Bradford & Cowell (2012), Bonham et al (1986), Holland et al (1978), Hood et al (2002), Padfield & Liebling (2000)

6.1 Risk of parole failure scales

Pogrebin, Poole, and Regoli (1986), using a sample of 292 paroled offenders in Colorado, developed a statistical predictor of successful parole completion. Prior to the development of the statistical predictor, the Colorado Parole Guideline Matrix, decisions were supposedly based on a detailed consideration of a large amount of information concerning the applicant's previous criminal behaviour and other factors. However, the authors observed that in practice the Parole Board nearly always accepted the recommendation made by prison staff. Following the development of the Matrix it was incorporated into the discussions of the Parole Board at its monthly meetings and a member of the research team attended each meeting as an adviser. After a year the Parole Board was given the option of incorporating the Matrix permanently into its deliberations, but refused. They apparently wanted to maintain complete discretion over the decision-making process. In other words, the objective instrument was not introduced because of policy considerations which were unrelated to its accuracy and potential utility.

Similarly, Hoffman (1994) reported that the Salient Factor Score (SFS), an objective predictor used in the United States Federal Parole System, had maintained its accuracy very well over the 20 years since its last validation. Nonetheless, it was superseded by an essentially similar instrument called the Criminal History Score, based on a revised set of scoring criteria which dropped some factors included in the SFS. These factors were excluded, not for any technical reason, but because of legal decisions which decided that their inclusion was unfair to the offender. This again demonstrates that practical utility is not the only consideration governing the use of objective measures in public service decision-making.

6.2 Reconviction risk scales

The other kind of statistical risk assessment instrument sometimes available to parole boards predicts reconviction rather than parole failure. Technical parole violations and minor offences are often of little interest to parole boards. For example, in England and Wales the Parole Board is supposed to be satisfied before granting parole that the risk to "life and limb" posed by the offender is "minimal" (Parole Board for England and Wales, 2012, chapter 4, p. 2). This is partly because the Parole Board now only considers those serving indeterminate sentences, who have mostly committed serious offences. Therefore, an objective instrument which estimates the likelihood of the offender committing a further serious offence should, in principle, be very helpful.

Hood and Shute (2000) studied the relationship between parole decisions and an objective risk indicator, the Risk of Reconviction Score (ROR). In the case of non-sex offenders only there was a strong relationship between this score and the probability of parole (chi-square=73.7, $p<.0001$). However, the Parole Board was still very cautious,

as half of these offenders had a ROR of 7% or less but 40% were refused parole. With sex offenders the Board exercised even more caution, refusing 78% of those with a similar objective risk, and there was no relationship with the ROR score in this group. Amongst those who had completed the Sex Offender Treatment Programme the parole rate was much the same whatever their risk category. Again, this suggests that parole panels are not necessarily very responsive to objective information about risk, preferring to make their own individualised judgements.

Holland et al. (1978) found that the California Parole Board did not use reconviction risk measures well. Such a measure was available, but the Board concentrated on the severity of the index offence whilst reporting that their decision was based on other factors. The decision of the Board was not related to subsequent success or failure after release. Bonham et al. (1986) found in contrast that an objective measure of recidivism risk did influence the Kansas Parole Board's decisions, although there were several other important determinants. It would therefore not be fair to conclude that objective risk estimates never influence parole boards.

Padfield and Liebling (2000) reported a study of the Parole Board of England and Wales relating to their decisions on discretionary lifers. A life sentence is mandatory for murder, but so-called "discretionary" life sentences can be imposed for other serious offences such as attempted murder, manslaughter, robbery, and rape. This is usually because the offender has committed several offences of the same type, or because the judge feels that there are mental health considerations which need to be adequately managed before release is considered. The study involved direct observation of discussions by parole panels. The authors considered that there was very little systematic risk assessment and indeed that risk was rarely discussed. They also reported

that some known and validated risk factors were never discussed, and others only "haphazardly". They concluded by questioning whether this really amounted to risk assessment at all and described the process as "more clinical than actuarial". This suggested a lack of appreciation of the value of objective risk indicators, and (as in other jurisdictions) an emphasis on an individualised assessment of each case. Bradford and Cowell (2012) reported similar findings from a study of prisoners given indeterminate sentences for public protection. There was no correlation between the statistical risk measure and likelihood of parole, and the bulk of those retained in closed prison conditions had low risk scores.

6.3 Discussion: Risk rating scales

From the above evidence it is clear that parole boards do not always make good use of risk rating scales. When reasons have been ascertained, they have reflected the boards' belief that individualised risk assessments need to be made, and a wish by the boards to maintain their discretion. However, when details of parolees' performance have been examined, risk scales have predicted more accurately than parole boards, which have often not been aware of the criteria on which their decisions were taken. This is consistent with an extensive literature on clinical judgements of risk (Hanson & Morton-Bourgon, 2009; Meehl, 1954, 1973; Quinsey et al., 2006; Quinsey & Maguire, 1986). It is also clear that many prisoners obtaining low scores on risk assessment instruments are nonetheless retained in custody. However, no studies have been carried out on mandatory life sentence prisoners (murderers).

7. Individual parole board member characteristics

Although it might seem an obvious subject for study, few researchers have examined the possibility that parole decisions may be influenced by the characteristics of the people who make them, rather than those about whom they are made. For example, although several studies have considered ethnicity of the offender (Huebner & Bynum, 2008; Morgan & Smith, 2008; Welsh & Ogloff, 2000), none has examined attitudes to ethnic minorities among parole board members. Table 7 shows studies which considered individual parole board member characteristics.

Table 7: Studies which examined individual parole board member characteristics	
Variable	**Studies**
Gender	Pogrebin et al (1986)
Individual variability	Gottfredson & Ballard (1966)

7.1 Gender

Pogrebin et al. (1986) did not set out to study whether there were gender effects in parole board decisions, but they did report relevant observations on the Colorado Parole Board. These concerned sexual offences. There was a single female member of the Parole Board, and they noted that when parole applications from sexual offenders were being considered she had a disproportionate input into discussions. They attributed this to a change in her manner when sexual offences were being considered, stating that she became much more assertive, and even aggressive, in discussing those cases. This appeared to influence other board members and result in fewer such cases being

paroled. It is easy to see why a female board member might identify with the victims of sexual offences more than others, since many such offences are committed against females (Ministry of Justice, 2012). Conversely, it would be easy to over-interpret a set of observations about a single person. However, it suggests an interesting possibility for further study, considering that many of the professionals who report on prisoners are female (Annison, 2013).

7.2 Individual variability

The earliest study reviewed (Gottfredson & Ballard, 1966) examined whether different parole board members judged cases differently. This was a study of 2,053 male parole applicants in California. Specific attributes of board members were not measured, but as parole decisions were taken by a single board member rather than a panel there was a possibility of bias. To eliminate the effects of sampling error (because board members might not all receive comparable cases) the researchers developed two prediction schemes, one for men who had been imprisoned before and one for those on their first sentence. These provided an expected length of prison time based on the time served by prisoners with the same characteristics. The authors initially found that parole board members did grant parole at different rates, but when the expected length of prison time was taken into account these differences disappeared. The authors concluded that there was no evidence of any significant differences in the criteria being applied by different board members.

7.3 Discussion: Individual parole board member characteristics

Just two studies were found which investigated the characteristics of individual parole board members. They raised more questions than they answered, for example, about the

possible influence of the gender of parole board members. This may be a politically sensitive topic, considering that there is a need to portray the parole system as fair and that prisoners may take legal action if it is not, but it is an unfortunate gap in the available research. Bridging that gap would require not only consideration of political sensitivity, but also of the fact that people are often not aware of the criteria upon which their judgements are based (Kahneman, 2011; Tversky & Kahneman, 1974).

8. Cognitive processes in parole decisions

Another feature of parole decision-making which has received relatively little attention is the ability of people to weigh and combine the many different sources of information typically presented to parole boards. This is a complex task, but it has generally been assumed that human beings can conduct it fairly and properly. Even where bias has been suggested, this has supposedly been due to the social attitudes of those making the judgements, such as those towards minority ethnic groups, rather than any inherent limitations of the human brain. In recent years cognitive studies have suggested that such limitations might in fact be crucial in complex decision-making (Kahneman, 2011; Tversky & Kahneman, 1974). One study examined this problem, and two papers were found which challenged and discussed it. These are shown in Table 8.

Table 8: Studies which examined cognitive processes in parole decisions	
Variable	**Studies**
Mental depletion	Danziger, Levav, & Avnaim-Pesso (2011a, 2011b), Weinshall-Margel & Shapard (2011)

8.1 Mental depletion

Danziger, Levav, and Avnaim-Pesso (2011a) studied parole decision-making in Israel. The practice there was for a number of cases to be decided in a day by the same three-person parole panel, consisting of a judge advised by a criminologist and a social worker. The mean number of cases decided per day was 22.58, taking a mean time of 5.98 minutes each. This suggests a very heavy cognitive load, and the authors cited earlier research suggesting that concentrated decision-making of this kind induced an effect termed "mental depletion". The relevant feature of mental depletion is that under heavy cognitive load people have been shown in a variety of situations to give up the decision-making and revert to the default or status quo position. Danziger at al. cited research which demonstrated this process in a variety of situations (Gailliot & Baumeister, 2007; Levav, Heitmann, Herrmann, & Iyengar, 2010; Muraven M. & Baumeister, 2000). In the parole context, the default position would be to leave the applicant in prison and refuse parole. Danziger et al. reported that judges tended to refuse parole more frequently as the session progressed. This was consistent with the findings of previous research. However, the research also suggested that after a rest (especially one which included food to restore blood sugar level) mental depletion would be ameliorated and the process would start again. Danziger et al. showed that this was true, and that after a mid-morning break and lunch break there was an immediate increase in the number of approvals, with a gradual reduction again as the session progressed.

These findings were questioned by Weinshall-Margel and Shapard (2011). They suggested various reasons why certain cases might have been sequenced in a way that would produce these results, and that the results were an artefact of case administration.

Danziger, Levav, and Avnaim-Pesso (2011b) rebutted this in a further analysis of their data, supplemented by some new information collected specifically to investigate these hypotheses. They were able to show that the sequencing hypothesised had rarely taken place, and that this involved too few cases to have affected the final result.

8.2 Discussion: Cognitive processes in parole decisions

What these findings suggest is that bias may be present in judicial and quasi-judicial proceedings for reasons which are nothing to do with the task or the social attitudes held by those making the judgements. Indeed, there are likely to be biases which are inherent in the way that human judgements are made, and therefore unavoidable. The Danziger et al. study may at first sight appear applicable only to a situation where large numbers of parole decisions are being taken in one session, but arguably the much longer parole hearings typical of UK practice (often a whole working day, in this candidate's experience) actually involve a great many decisions taken serially about the value of different reports, institutional behaviour, criminal history, and so forth. There is no guarantee that the difference in procedure will avoid mental depletion. If it does not, the legal and human rights implications are considerable.

Discussion

Whatever they may lack in quality, one firm conclusion which can be drawn from the above studies is that, whatever the jurisdiction and whatever the precise details of the measurements used, most parole decisions are relatively predictable on the basis of only a few variables, although these are not always the same and any causal relationship is often unclear. In some jurisdictions, only one variable was crucial, often a report by a probation officer or senior prison official (Bradford & Cowell, 2012; Heinz et al., 1976;

Hood & Shute, 2000; Morgan & Smith, 2005a, 2005b; Winfree et al., 1990; Zinger, 2012). This simply moves the quest for an explanation one stage further back in the process, since it is not known how those report writers themselves reached their recommendation.

Some of the variables on which parole boards have based their decisions are not related to risk to the public (the basis usually claimed), for example, prison disciplinary reports (Bonham et al., 1986; Caplan, 2007; Heinz et al., 1976; Holland et al., 1978; Huebner & Bynum, 2006, 2008; Morgan & Smith, 2005b; Scott, 1974; Welsh & Ogloff, 2000). When more objective statistical indicators of risk have been available, they have not always been used (Bradford & Cowell, 2012; Padfield & Liebling, 2000), or have even been rejected outright (Pogrebin et al., 1986). This appears to reflect a desire on the part of at least some parole boards to decide cases on an individual basis. However, risk is statistical by its nature, and there is no evidence that it can be effectively assessed using individualised clinical methods. Indeed, there is a substantial literature to the contrary (Hanson & Morton-Bourgon, 2009; Meehl, 1954, 1973; Philipse et al., 2006; Singh & Fazel, 2010; Wakeling, Freemantle, Beech, & Elliott, 2011). Nor is there any evidence that "adjusting" a statistical predictor by the addition of clinical judgement does anything but dilute its accuracy (Campbell & DeClue, 2010; DeClue, 2013; Krauss, 2004; Nugent & Zamble, 2001).

It may be argued that parole boards need to consider more than just risk when taking their decisions. For example, in many jurisdictions it is assumed that parole may contribute to the reintegration of the offender into society, and the Parole Board for England and Wales has the option of recommending progression to open prison conditions as part of this process (Ministry of Justice, 2013a). However, the same

document lists only three aims of the Parole Board. The first is to conduct risk assessments on parole applicants. The other two relate to business practice and public relations. Thus, although rehabilitative efforts are not discouraged, the paramount objective is clearly to protect the public from high-risk offenders. This is particularly the case with indeterminate-sentence prisoners, who are supposed to be released after their minimum term unless continuing risk indicates otherwise. Genuine indicators of risk can only be determined through research, but there is little evidence that parole boards make use of this.

The lack of use of research evidence is well illustrated by the fact that the age of parole applicants is typically ignored by parole boards. There is a wealth of evidence showing that risk decreases with age in all kinds of offender, albeit not necessarily at the same rate or from the same point. This has been demonstrated for sex offenders (Barbaree et al., 2003; Hanson & Morton-Bourgon, 2005; Wakeling et al., 2011), for violent offenders (Mokros et al., 2010; Sampson & Laub, 2003), and even for psychopathic offenders (Porter et al., 2001). Despite its obvious relevance to public risk, only two studies found age to be a factor in parole decisions (Heinz et al., 1976; Huebner & Bynum, 2006). In the first of these, older men were more likely to gain parole, but so were the youngest. In the second study, older offenders were *less* likely to be paroled. The relevant research was not available in 1976, but it was in 2006.

A similar point may be made with respect to index offence. Several studies found that sexual offenders were less likely to be paroled than nonsexual offenders (Bradford & Cowell, 2012; Hood & Shute, 2000; Pogrebin et al., 1986; Welsh & Ogloff, 2000). This applied to both real-life and simulated decisions (Turpin-Petrosino, 1999). The studies suggest widespread nervousness among parole boards when it comes to paroling

sex offenders — even incest offenders, whose reconviction rates are known to be relatively low (Hanson & Bussière, 1998; Hanson et al., 2003; Hood et al., 2002). It might be argued that this reflects the seriousness of a possible re-offence, rather than the probability, but this does not explain why violent offenders were more likely to be paroled than sex offenders (Bradford & Cowell, 2012).

Some concepts used by parole boards appeared vague and poorly defined, despite influencing decision making. For example, two studies found that the perceived quality of the release plan had an important impact on parole decisions (Bonham et al., 1986; Bradford & Cowell, 2012), but in neither case was there a clear definition of a "good plan". Many of these judgements appeared very subjective, a point underlined by Padfield and Liebling (2000), who described the Parole Board for England and Wales as conducting little risk estimation, none of which was objectively measured. Bradford and Cowell (2012) also noted that the Parole Board generally followed offender managers' recommendations, but simultaneously complained that their reports relied too much on second-hand information. Relying on a report which one complains is insufficiently evidence-based suggests considerable confusion. A related point is that parole boards were not always aware of the factors actually influencing their decisions (Bonham et al., 1986; Padfield & Liebling, 2000; Pogrebin et al., 1986). This calls to mind the work of Kahneman and Tversky (Kahneman, 2011; Tversky & Kahneman, 1974), who demonstrated that professionals in a variety of fields often take decisions on the basis of emotional factors or rules of thumb, but rationalise them afterwards by claiming to use a more logical and evidence-based approach. Similarly, some of the more recent studies have presented evidence that the task demanded of parole boards, namely the understanding and integration of very large amounts of verbal and written information,

simply places too great a cognitive load on the decision makers. This shows in the process of mental depletion, whereby decision-makers revert to the status quo when too many decisions have to be taken (Danziger et al., 2011a).

A variety of professionals, including psychologists, probation officers and psychiatrists, submit reports to parole boards. The boards themselves include some of these professionals among their number, and they are required to keep up-to-date with the research literature in their field. The evidence suggests that this is not enough to ensure rational decision-making.

Implications

Does this imply that parole boards are redundant? Scott (1974) questioned the utility of indeterminate sentences, and of parole boards, but that need not be the only conclusion. Hood and Shute (2000) found with determinate-sentence prisoners that the parole board could not estimate risk accurately, and consequently large numbers of low risk prisoners were being detained unnecessarily. After that study the law was changed so that determinate-sentence prisoners were automatically released on licence halfway through their sentence, although they could be recalled if their probation officer thought it necessary.

Currently in the United Kingdom life sentence prisoners are supposed to be released upon expiry of their minimum term unless there is evidence of continuing risk. However, parole boards do not seem to be able to assess risk any more accurately in the case of indeterminate-sentence prisoners than they could with determinate-sentence prisoners (Padfield & Liebling, 2000). A similar solution may be appropriate to that which followed the report by Hood and Shute. Perhaps paroling all life sentence prisoners on tariff expiry is too radical a step, but creating a firmer presumption in

favour of parole might assist the Parole Board. That is, parole would have to be granted unless a case could be made for continuing high risk on the part of the prisoner. Indicators of risk would need to be more objective than is currently the case if any real change were to be ensured. One study (Stone-Meierhoefer & Hoffman, 1982) examined a change in the US federal parole system and found that assigning a presumptive parole date did not lead to worse institutional behaviour (which had been suggested) but did lead to fewer enrolments in educational programmes. Offending behaviour programmes had not been introduced at that time. That study was conducted on determinate-sentence prisoners, however, and those serving indeterminate sentences might have a greater incentive to enrol in programmes in order to demonstrate a reduction in risk.

Alternatively, indeterminate sentences could simply be abolished and replaced with long determinate sentences. That is the case in some other countries, notably Norway, where the maximum sentence available is 21 years (Townsend, 2012), although this can be extended (for five years at a time) if a case can be made for genuine evidence of continuing risk.

Strengths and limitations

Perhaps the main strength of this review is that it made use of a wide range of sources, including several high-quality libraries, official sources, and a range of professional colleagues, contacted both personally and by means of the Internet. For example, a number of relatively obscure papers were supplied by colleagues in North America who had access to them through their employment. An additional strength is the period of time covered by the review, which covered almost 50 years. A weakness is that the candidate did not obtain an indication of interrater reliability for the exclusion and inclusion criteria by having a colleague replicate the selection procedure on a subsample

of studies. It might be possible to supply this if preparing the review for publication. A further weakness, compared with many contemporary systematic reviews, is that the sheer variety of studies, choice of variables, and methods made it difficult to synthesise the findings. In particular, it was judged not possible to conduct a meta-analysis.

Conclusion

Despite the technical difficulties in synthesising the data, it has been possible to conclude that in most jurisdictions parole decisions can be largely predicted from relatively few variables, and often only one. A further conclusion is that the decisions taken do not usually relate in any systematic way to known risk factors for reconviction or parole failure. Indeed, in some cases there is evidence that parole boards have preferred to maintain their discretion to use less systematic methods. These methods are known to be an unreliable means of assessing risk to the public, which is usually the main basis claimed for parole decisions.

FOREWORD TO CHAPTER 3

From the literature review presented in Chapter 2 it is clear that parole board decisions are often predictable from just a few variables. However, different variables may predict parole decisions in different jurisdictions, and even in the same jurisdiction in different time periods. The particular focus of interest in this thesis is what determines parole decisions for male life-sentence prisoners today in England and Wales, and the literature review did little to clarify this.

Another consideration is the type of sentence being served by prisoners in these studies. In most cases, there was no information about whether the sentences being served were determinate or indeterminate, or a mixture. Since the parole system in England and Wales now considers indeterminate-sentence prisoners almost exclusively, this could be an important factor. Many indeterminate-sentence prisoners serve very long periods of time, and this is especially true for those who have committed homicide or very serious sexual offences. Likewise, a number of the studies included female prisoners without differentiating them in any way from their male counterparts. For some studies, this could have affected the findings, and again this would lessen their applicability to male life-sentence prisoners today. For example, Scott (1974) found that female prisoners were more likely to be paroled, but this possibility was not examined in other studies. On the other hand, Petersilia (2003) suggested that, at least in the United States, parole for women offenders had been disproportionately restricted by increasing harshness towards drug offences, for which many of them had been convicted (pp.24-26). Also, some studies have shown that predictors of sentence

severity may differ between men and women (Crew, 1991) and if similar differences existed in respect of parole this could obscure parts of the process.

Most jurisdictions emphasised the importance of risk to the public when taking parole decisions. Nonetheless, the quality of risk assessment was generally poor. In England and Wales today, professional reports to parole boards may include a range of risk assessments made using instruments which have been developed since many of the studies which featured in the literature review were conducted. It would be instructive to examine whether these assessments influence present-day parole decisions, and how.

Finally, the literature review reported a widespread finding that parole boards took advice from professional report-writers. These reporters usually made a recommendation with respect to parole, and this recommendation often strongly predicted the parole decision. However, this did nothing to clarify how the professional report-writers themselves came to their recommendation. For example, there was usually no indication as to whether they had used a systematic form of risk assessment, and there was little previous research available to clarify this. Consultation with well-known criminological researchers, and the Parole Board for England and Wales, had confirmed this conclusion, suggesting that clarification of modern practice in this jurisdiction would be helpful.

The candidate's daily work included the preparation of reports on male life-sentence prisoners, and this gave him access to a great deal of data concerning them. Most of these reports were for parole purposes, although some were conducted for other reasons, such as recommending changes in security category. All of this work was conducted in England and Wales, and it appeared that this data could provide a means of researching these questions. An experimental study was therefore planned to examine parole

decisions in the years 2011-2013, using a convenience sample of cases referred to the candidate by the legal representatives of prisoners. The focus of the study was to be the use of risk assessments in parole proceedings. This would involve risk assessment instruments currently in regular use for this purpose, and would examine the relationships between the risk assessment instruments themselves, as well as their relationship with professional recommendations and the final parole decision.

CHAPTER 3: A STUDY OF RISK FACTORS AND PAROLE DECISIONS FOR 84 LIFE SENTENCE PRISONERS

Abstract

This experimental study related a series of criminological measures and risk assessment scores (HCR-20, SVR-20, and PCL-R) to the parole decisions taken about 84 life sentence prisoners. No relationship was found between parole decisions and those measures most predictive of reconviction. There were correlations between the Parole Board decision and those risk assessment subscales most dependent upon clinical or subjective impressions of the parole applicant. The Parole Board decision was predictable in 68% of cases from the recommendation given by the offender manager (external probation officer) alone, and predictable in 84% of cases from a combined measure of all three professional recommendations (prison psychologist, seconded and external probation officers). There was considerable agreement between professionals. This raised the question of what determined those recommendations. Exploratory regression analyses showed that 39% of the variance in the combined professional recommendation measure was accounted for by the Clinical and Risk Management scales of the HCR-20. Similar results were obtained in two subgroups whose index offence did or did not contain a sexual element. Parole decisions for those cases with a sexual element were unrelated to scores on the SVR-20. The results suggested that the risk assessments presented by professionals were largely dependent upon clinical impressions, rather than empirically demonstrated risk factors. These impressions have

a poor to zero relationship with the risk of subsequent reconviction, suggesting that the parole decision-making is not genuinely risk-based. Implications and suggestions for changes in the system are discussed.

Introduction

Life sentence prisoners ("lifers") pose special problems for the criminal justice system. First, because sentences are indeterminate, there is no automatic right of release and this can result in prisoners serving very long periods; participants in this study had served up to 40 years, and some were up to 26 years past the expiry of their minimum term. This may be regarded as inhumane and is certainly expensive. Second, most lifers can be paroled if their risk to the public is judged to be low enough, which requires a means of making that judgement. Third, at times of increasing punitiveness in society sentences tend to be lengthened, and according to a number of observers western society is currently going through such a time (Griffin & O'Donnell, 2012; Zinger, 2012). In England and Wales this has led to an increase in the prison population which is partly due to an increase in the length of imprisonment of indeterminate-sentence prisoners, and partly due to the increased imposition of such sentences, particularly for sexual and violent offenders (Ministry of Justice, 2013b).

The parole system in England and Wales

Since the abolition of the death penalty in the UK a life sentence has been mandatory in the event of a murder conviction. However, a so-called "discretionary life sentence" may be imposed for some other serious offences, notably rape and manslaughter. The concept of a minimum term, previously known as the "tariff", was introduced as a means of tailoring an indeterminate sentence to the seriousness of the offence. The

minimum term is imposed by the sentencing judge, and is a term which must be served before indeterminate-sentence prisoners can be considered for parole. Only in extreme circumstances can the Secretary of State for Justice parole a prisoner before the expiry of the minimum term. This is usually only granted for humanitarian reasons, such as a prisoner's terminal illness. After expiry of the minimum term all lifers can apply for parole, although it is not automatic. It is up to the Parole Board to determine a prisoner's level of risk, and it may only direct release if it judges the "risk to life and limb" to be "minimal" (Parole Board for England and Wales, 2012).

Parole applications can be considered on the basis of a consideration of documentary evidence alone or by way of a hearing. Each case is considered by a panel of three members of the Parole Board. In lifer cases the panel will normally include a psychologist or psychiatrist, and the panel is chaired by a judge or senior lawyer. The documentary evidence before them will include reports from probation officers, prison staff, and sometimes psychologists and psychiatrists, as well as an account of the offence and details of the prisoner's previous criminal history. If the case proceeds to a hearing, the panel will hear submissions from the prisoner in person, and oral evidence from those who have presented reports. The prisoner may be legally represented, and may call witnesses to support his case, including expert witnesses such as psychologists. The parole panel can direct release, in which case the Secretary of State for Justice must comply. Alternatively, if the prisoner is in closed prison the panel can recommend his transfer to an open prison preparatory to release, but this is not binding on the Secretary of State.

Minimum terms for life sentences were not introduced immediately after the abolition of the death penalty, and some older lifers never received one. Technically,

therefore, they were eligible for parole as soon after sentencing as a parole panel could be convened. In practice, this was extraordinarily unlikely. Since 2005 determinate-sentence prisoners in the UK have all been conditionally released partway through their sentence, so the Parole Board no longer considers their cases unless their release has been revoked and they are applying to be released again. Therefore, almost all of the Parole Board's decisions are now taken in respect of lifers and other indeterminate prisoners (Parole Board for England and Wales, 2012).

There is another group of indeterminate-sentences imposed for persistently repeated offences, normally sexual or violent offences. These are known as "indeterminate sentences for public protection" (IPPs) (Ministry of Justice, 2013b), or more colloquially as "two strikes and you're out". Typically, these feature a relatively short minimum term. This research did not include any prisoners sentenced under these arrangements.

Previous research on parole criteria

One might have expected that considerable attention would have been paid to the increasing burden imposed by life sentence prisoners on resources, and the possibility of reducing this by carefully targeted parole. Typically, this seems not to have been the case. A number of papers have examined the parole process, but these have mostly been in jurisdictions where parole is available to determinate-sentence prisoners. In a review of North American literature, Caplan (2007) did not find any studies where lifers were considered separately. Studies have tended to combine parole applicants irrespective of sentence, offence type, and often gender (Matejkowski et al., 2011; Morgan & Smith, 2005a, 2005b). Others have studied the effect of specific factors such as gender (Scott,

1974) or ethnicity (Huebner & Bynum, 2008; Welsh & Ogloff, 2000) but have failed to consider different categories of prisoner, and specifically life sentence prisoners.

Another important issue, given that risk to the public is intended to be the criterion for deciding whether to parole lifers or not, is the type of risk assessment used by parole boards. A number of studies have examined this. They have generally found that objective risk assessments were not favoured by parole boards, and subjective or "clinical" assessments were preferred. In the USA Pogrebin et al. (1986) developed a statistical predictor of parole success intended to help the Colorado Parole Board, but the Board refused to use it. In a Californian study Holland et al. (1978) found that, although an objective risk measure was available, the Parole Board was not influenced by it. They were unduly influenced by the seriousness of the most recent offence, and their decisions were unrelated to subsequent performance on parole. On the other hand, Bonham et al. (1986) found that the Kansas Parole Board were influenced partly by a statistical measure of recidivism risk.

Three studies have examined risk assessment and parole decisions for indeterminate-sentence prisoners in England and Wales. Hood and Shute (2000) found that the Parole Board generally overestimated risk, and that this was especially true for sex offenders. A Home Office statistical predictor of recidivism was available to the researchers, but not to the Parole Board, whose decisions were nonetheless correlated with it. This suggested that the Board was at least partly influenced by factors genuinely reflecting risk. The Hood and Shute study was not specifically concerned with indeterminate-sentence prisoners. However, Padfield and Liebling (2000), in an observational study of discretionary life sentence parole applicants, found that there was little systematic risk assessment, and that the Board's decision-making was essentially clinical in style and

based on risk-related beliefs of unknown accuracy. Bradford and Cowell (2012) reported that there was no correlation between statistical risk scales and parole decisions for IPP prisoners. They also found that the recommendation of the offender manager (external probation officer) was strongly predictive of the parole decision, despite the fact that this professional had less interaction with the prisoner than any other. It was not clear what influenced the offender manager's recommendation itself. A similar influence of reports from professionals has been reported in other studies (Heinz et al., 1976; Morgan & Smith, 2005b; Winfree et al., 1990) and in those studies it was likewise unclear how the influential professional recommendation itself had been arrived at.

As mentioned above, Hood and Shute (2000) found that the Parole Board overestimated risk, especially for sex offenders. Other studies have found a similar overestimation of risk with respect to sex offenders (Bradford & Cowell, 2012; Hood et al., 2002; Welsh & Ogloff, 2000), with correspondingly low rates of parole. Indeed, Bradford and Cowell (2012) found that nonsexual violent offenders were more than eight times more likely to obtain parole than sex offenders. This is despite the fact that reconviction rates for sexual offenders are generally low (Hanson & Bussière, 1998; Hanson et al., 2003).

From this reading of the literature it appears that parole boards in general are not actually assessing future risk. However, the literature on risk factors and their measurement is extensive, particularly with reference to violent offending (Coid et al., 2007; Grann, Danesh, & Fazel, 2008; Mokros et al., 2010; Singh & Fazel, 2010; Singh, Grann, & Fazel, 2011; Yang, Wong, & Coid, 2010) and sexual offending (Beech, Fisher, & Thornton, 2003; Grann et al., 2008; Hanson & Bussière, 1998; Hanson &

Morton-Bourgon, 2009; Hanson et al., 2003; Harkins et al., 2010; Singh & Fazel, 2010; Wakeling et al., 2011).

It should also be noted that the literature is clear that statistical ("actuarial") methods of prediction are the most accurate predictors of reconviction, that structured clinical judgements are less so, and unstructured clinical judgement ineffective (Coid et al., 2007; Hanson & Morton-Bourgon, 2009). However, even statistical scales achieve only moderate accuracy (Gray et al., 2004; Grubin, 2011; Hanson & Thornton, 2000) and attempts to enhance this by the addition of clinical information have not been very successful (Campbell & DeClue, 2010; DeClue, 2013; Gray et al., 2004; Hanson & Morton-Bourgon, 2005). Grubin (2011) has found evidence that the Risk Matrix 2000 (Thornton et al., 2003), a statistical predictor commonly used in the UK to assess the risk of both sexual and violent offending, *reverse*-predicted the seriousness of offending. That is, it identified individuals who would reconvict quickly rather than those who would commit grave offences. In Grubin's study those who committed the most serious offences subsequent to the Risk Matrix 2000 assessment, including murder, had been assessed as low in risk. This raises doubts as to whether risk is a unitary concept or a combination of at least two dimensions, namely speed and severity of reconviction. Current assessments merely predict repetition.

Statistical predictors have been criticised for having confidence limits so wide that they cannot usefully be applied to individuals (Hart, 2005; Hart, Michie, & Cooke, 2007; Morrissey, Mooney, Hogue, Lindsay, & Taylor, 2007), and for failing to give enough weight to age-related risk reduction (Barbaree et al., 2003; Wollert, 2006). Structured anchored clinical judgments (SACJs) are sometimes claimed to be of greater practical value when dealing with individuals. SACJs require an assessment to be made

of whether a characteristic, thought to be relevant to risk, is present in the individual being assessed. Typically, this is done on a three-point scale, according to whether it is absent, partially or possibly present, or fully present.

Three assessments have been developed by Webster, Hart and colleagues: the HCR-20 for nonsexual violence (Webster, Douglas, Eaves, & Hart, 1997), the SVR-20 for sexual violence (Boer, Hart, Kropp, & Webster, 1997), and the SARA for domestic violence (Kropp, Hart, Webster, & Eaves, 1995). The authors of these instruments have advised against creating a numerical score out of the item ratings for clinical use, preferring that assessors should use their professional judgement to weigh these appropriately and combine them into an overall rating of low, medium or high risk.

However, it is not clear that humans can perform this task, even if they are trained clinicians, and there is no agreed operational definition of low, medium or high risk. In practice most research has used a numerical scale, which the authors themselves advise for research purposes (Webster et al., 1997). Several studies have also shown that, although a risk assessment instrument as a whole may be predictive, some of its items or subscales may not be. For example Coid et al. (2007) found the Historical scale of the HCR-20 to be the strongest predictor of future violence, as did Dietrich (1994), with the Clinical and Risk Management scales performing less well. Coid et al. (2011) found that only seven out of 40 items of the HCR-20 and PCL-R predicted violence. Dernevik (2004) found that over 40% of the variance in HCR-20 ratings was attributable to the emotional reactions which assessors had to the offenders being assessed.

Given these findings, it is surprising that instruments like the HCR-20 and PCL-R have been as well supported as they have been in the literature. However, it may simply be that they reflect general criminality, and that the more items present in an individual

case, the greater the risk of subsequent recidivism. This was the view of Kroner, Mills, and Reddon (2005), who found that the items of individual scales were more or less interchangeable, and that a mixed scale constructed out of items from four well-known risk assessment scales (the PCL-R, the LSI-R, the VRAG, and the GSIR) predicted reconviction as well as any of the "genuine" scales. This was supported by Coid et al. (2011), who found that most items on three scales (the PCL-R, HCR-20, and VRAG) were not independently predictive. They constructed a mixed scale from those that were, but reported a negligible increase in predictive accuracy. They suggested that there may be an inherent limit to the predictability of criminal behaviour.

Although the PCL-R is intended to be a measure of psychopathic personality, and not risk *per se* (Boccaccini, Turner, & Murrie, 2008; Hare, 2003b) it is in practice widely used as a risk assessment measure. The PCL-R produces four "Facet" scores which combine into two Factor scores. Facet 1 contains "interpersonal" items, such as those relating to superficial charm and manipulativeness, while Facet 2 contains "affective" items, reflecting such characteristics as lack of emotional response and callousness. Facet 3 consists of "lifestyle" items such as impulsivity and parasitism, while Facet 4 features "antisocial" items relating to criminal history and behaviour problems. Facets 1 and 2 are combined to form Factor 1, which is intended to reflect core psychopathic personality characteristics. Facets 3 and 4 are similarly combined to form Factor 2, which reflects criminality and a socially deviant lifestyle.

Several studies have reported zero reliability for Factor 1, when used by clinicians "in the field", as opposed to groups of academic researchers; they have also shown the scoring of this factor to be biased according to whether assessors were hired by the defence or the prosecution (Boccaccini, Turner, et al., 2008; Murrie, Boccaccini,

Caperton, & Rufino, 2012; Murrie, Boccaccini, Johnson, & Janke, 2009), and even by the assessor's own personality characteristics (Edens et al., 2010; Miller et al., 2011; Murrie et al., 2012). However, Mokros et al. (2010) found that Factor 2, in combination with an individual's age, made a good predictor of future violence. Factor 2 is seen as more objectively scored (Rufino, Boccaccini, & Guy, 2011), as it is largely based on an offender's criminal record. Factor 1 relies on clinical judgement, and may therefore be subject to the same shortcomings that Dernevik (2004) reported for the HCR-20. This suggests that decision makers should not rely too heavily on individual item or scale scores, and should be aware that some are more predictive than others.

Clinical risk assessment

Given that parole boards in many jurisdictions have not always welcomed statistical risk assessments, preferring to retain the discretion to decide as they wish, the question of how closely their decisions parallel more objective assessments of risk could be regarded as critical. Risk to the public is generally the criterion which parole boards are supposed to apply in all the jurisdictions studied above, and specifically in the UK. Previous studies have either excluded life sentence prisoners or mixed them with determinate-sentence prisoners, so that no specific conclusions can be drawn about them. This study therefore proposed to examine the factors relating to parole decisions in a group consisting only of life sentence prisoners in England and Wales, with specific reference to the influence of risk assessments on the parole decision.

Both Rufino, Boccaccini, Hawes, and Murrie (2012) and Murrie et al. (2009) found partisan allegiance in PCL-R ratings, but there is no reason to think that other risk assessment instruments would be immune to such influences. Indeed, the possible

consequences of an inappropriate release of a violent prisoner would make it highly likely. Research by Dernevik (2004) suggests that the HCR-20 may also be affected by emotional responses on the part of the rater. Rufino et al. (2011) investigated a related problem, namely the reliability of the more subjective risk assessment scale items. In a study of the PCL-R and HCR-20 they found that professionals agreed strongly as to which items were the most subjective. Ratings of subjectivity correlated negatively with published reliabilities for the items and subscales of both instruments. They also found that the subjectivity of items was inversely related to their predictive utility. In other words, the more clinical judgement was involved, the less reliable and accurate the risk assessment. As Kahneman (2011) has pointed out, people tend to avoid potential failure more keenly than they are attracted to potential success. Essentially, professionals who prevent a "safe" applicant from gaining parole suffer no consequences, but may suffer public criticism or worse if they release one who is "unsafe". Under the circumstances, we might expect the weaknesses of a clinical assessment to be increased, and also the clinical judgement elements of a structured assessment. This is consistent with previous research showing that parole boards were conservative in deciding to release prisoners, and that their methods of risk assessment were predominantly subjective.

The present study

The literature suggests that parole boards and the professionals who advise them do not estimate risk well, or make good use of statistical indicators of risk, being more inclined to fall back on relatively subjective or clinical measures, whose weaknesses have been outlined above. The practical implications of this for life-sentence prisoners are considerable. First, it implies that many low-risk individuals who could safely be released are likely to serve longer than necessary in excess of their minimum terms.

Second, it implies that parole decisions are not likely to be strongly related to relatively objective measures of risk.

Since a large part of the candidate's work consisted of reporting on parole applicants, and performing risk assessments, he had access to a large number of life-sentence prisoners and their prison files. Approximately 50 such assessments were carried out per year. Therefore, in the two years of the degree programme it was expected that a sample of 100 life-sentence prisoners could be obtained. Not all of these would be parole applicants, as the candidate was also asked to perform risk assessments for reasons other than parole. However, the majority would be parole applicants, and data from those who were not could still be used for analyses of relationships between risk assessment instruments. It is usual for professional witnesses to be informed of the outcome of parole hearings which they have attended, so that information would also be available.

Hypotheses

It was generally hypothesised that parole decisions, and the length of time served by lifers in excess of their minimum terms, would correlate with those scales which have been shown to be least predictive of recidivism, but most related to the parole applicants' current clinical presentation. These are the characteristics of which they will be most aware, either through their own interviewing of the prisoner, or through reports made by professionals. Following the literature, it was hypothesised that decisions for those with sexual offences would be particularly influenced by the characteristics of the offending itself.

On the rationale that some scales on risk assessment instruments are more clinical in nature and others more objective, it was further hypothesised that the scales would

intercorrelate accordingly. That is, clinical scales on one would correlate with clinical scales on another, and more objective scales would also intercorrelate. At first sight, this might appear counterintuitive, as the poor reliability reported for some of the more clinical scales would preclude correlations of any significance. However, the studies which have been cited refer to inter-rater reliability. As there was only one rater for all the scales used in this study, inter-rater reliability was not a factor. From the point of view that personal presentation is the basis for all of these more clinical scales, correlations between them might reasonably be expected within the work of a single rater. Poor inter-rater reliability would still be expected, because raters would each have their own methods of interpreting personal presentation, and because the presentation of the offender might vary between raters. For example, an offender might well present as more guarded or hostile towards a prison psychologist than towards an independent psychologist, because the latter has been retained on his behalf and may be seen as more "defence oriented" than a member of the prison staff.

Specific hypotheses are listed below.

Hypothesis 1

That significant positive correlations would be found between the more subjective scales of risk assessment instruments (Facets 1 and 2 of the PCL-R, the Clinical and Risk Management scales of the HCR-20, and the Future Plans scale of the SVR-20).

Hypothesis 2

That significant positive correlations would be found between the less subjective scales of risk assessment instruments (Facets 3 and 4 of the PCL-R, the Historical scale of the HCR-20, and the Psychosocial Adjustment scale of the SVR-20).

Hypothesis 3

That significant positive correlations would be found between the more subjective scales of risk assessment instruments and the amount of time served in excess of the minimum term ("Years over").

Hypothesis 4

That significant negative correlations would be found between the more subjective scales of risk assessment instruments and professional recommendations in favour of progress.

Hypothesis 5

That significant positive correlations would be found between professional recommendations and the Parole Board's final decision.

Hypothesis 6

That significant negative correlations would be found between the more subjective scales of risk assessment instruments and Parole Board decisions in favour of progress.

Method

Ethical approval

The research was conducted in accordance with standards drawn up by the British Psychological Society and the Health and Care Professions Council. Ethical approval for this project was given by the Science, Technology, Engineering and Mathematics Ethical Review Committee of the University of Birmingham. The Committee approved

the final choice of wording for the consent form, a copy of which is in Appendix 3. Informed consent was given separately, and recorded on a different form, for the production of a report based on the interviews. Consent was obtained to use all data from the assessment interviews and psychometric testing for the research project.

Data were collected during interviews with participants and included psychometric test data (used for the psychological report), ratings on risk assessment schedules, and criminological information derived from participants' parole dossiers, or a similar dossier if he was not a parole applicant. All dossiers were supplied by the participants' legal advisers on their instructions. As well as obtaining a signature on the consent form, the candidate talked through the research project with each participant, explaining that the purpose was to investigate the factors involved in parole decisions about lifers, and that the decision whether or not to participate was entirely free and would not affect the psychological report. Participants were also told what information would be collected for the research and how the data would be stored and eventually disposed of. They were also told how the research results would be publicised after completion, and how they could withdraw from the project if they changed their minds about participating. None refused consent, and in the event none chose to withdraw from the project later. This 100% cooperation may reflect lifers' attitudes noticed by the candidate over a number of years, namely, that they often feel their interests are forgotten and that parole decisions are arbitrary and the process ought to be investigated. Indeed, many participants expressed such opinions during the consent process.

Procedure

Every participant was interviewed for a total of about four hours, including approximately one hour of psychometric testing for the psychological report. All assessments were carried out by the candidate, who had completed appropriate training in in their use. PCL-R information was collected using the interview guide and booklet published for the purpose (Hare, 2003a). Collateral information was obtained from case files of the individual in question. This took approximately two hours per case. The same sources were used to score the HCR-20 and SVR-20. Scoring of the three instruments took approximately a further two hours. The files typically contained an account of the index offence, judge's comments at sentencing, an account of the participant's progress during his current sentence, a police list of previous convictions, and recent reports on his behaviour from prison staff. They also contained a printout of an OASys (Offender Assessment System) report; this is a standard computerised prison report on offending and criminogenic needs (Howard & Dixon, 2012). Files also contained reports from the man's offender manager (home probation officer), offender supervisor (probation officer seconded to the prison), prison chaplain, medical officer and a set of reports from uniformed prison staff. Most also contained a prison psychological report to the Parole Board, and reports of any offending behaviour programmes undertaken. Many also contained psychological and psychiatric reports dating back a number of years, numbers and types varying according to the different histories of those concerned. Files were typically between 150 and 200 pages in length, though it was not always necessary to read every section of a file in order to obtain the information required.

Parole decisions are given within two weeks of the parole hearing, and it is usual that solicitors convey the outcome to independent psychologists and others who have given evidence. In rare cases they failed to do this and were followed up by telephone.

Measures

The following data were recorded:

1. Age in completed years at time of assessment ("Age now").

2. Length of minimum term or tariff (years).

3. Time actually served so far (years).

4. Time (complete years) served in excess of the minimum term ("Years over").

5. Number of offences, including the index offence, in each of seven categories. The rationale for these categories is that they represent different types of behaviour. Often in criminological research offence types are grouped in legal categories representing very heterogeneous kinds of behaviour, and this was an attempt to distinguish offence types which are behaviourally different, regardless of which particular statutes they happen to break. It was also an attempt to concentrate on significant offences, rather than the relatively minor miscellaneous offences which are common in the records of repeat offenders. The system of categorising offences here has been adapted from a system suggested by Hugh Marriage, OBE, former Crime Reduction Commissioner for the South-East Region (personal communication). The categories are:

    a. Sex offences (those involving sexual *contact* only). Does not include non-contact offences with sexual connotations, like indecent exposure, stealing women's clothing, or prostitution, etc.

b. Violent offences (those involving assault upon the victim). Includes Rape, but excludes robbery unless accompanied by assault.

c. Robbery (theft using interpersonal force/threats). Includes bank holdups, mugging, handbag snatching.

d. Burglary (breaking into enclosed premises for theft). Includes shops and offices as well as homes.

e. Deception (using deception for gain). Includes fraud, forgery, distraction burglary.

f. Theft (without force, burglary, or deception). Includes shoplifting, stealing by finding, but not handling stolen goods.

g. Vehicle-related offences (not for gain). Includes taking a vehicle without the owner's consent, driving without insurance, dangerous driving, drunk driving, etc).

6. Other criminal history measures:

a. Age at first conviction (whether adult or juvenile).

b. Mean prior sentence (Mean sentence length in months for all *previous* custodial sentences).

c. Latency (time in months between offences, not counting time spent in custody, i.e., when not free to commit most offences).

d. Total prior convictions (number of *previous* appearances at court when at least one conviction was imposed).

7. Index offence, recorded as murder (including two cases of attempted murder), manslaughter, or rape. These were the only three kinds of index offence in the sample.

8. Scores on the PCL-R (individual items, Facets, Factors, and Total score).

9. Scores on the HCR-20 (individual items, Historical, Clinical and Risk Management scores separately, and Total score; also, subscales and Total minus the psychopathy item).

10. Scores on the SVR-20 (individual items, Psychosexual Adjustment, Sexual Offences and Future Plans, and Total score; also, subscales and Total minus the psychopathy item).

11. For parole applicants, the recommendations by external and seconded probation officers, and that of the prison psychologist (if any) were also recorded. All of these report writers made recommendations in most parole cases. It is very unusual for lifers to be released directly into the community; they normally spend a period in open prison first. Those already in open prison might progress to the community or might stay where they were. This meant two types of progress were possible, but for the purposes of this project they were combined. Recommendations were coded in the following way: recommendation against progress (0), conditional recommendation for progress (1), or unconditional recommendation for progress (2). This was not a simple binary variable, because a conditional recommendation was occasionally made whereby progress was recommended on the condition that the prisoner undertook something first, such as an offending behaviour booster course. This suggested a three-point scale for professional recommendations, but in practice this produced cross-tabulations where several individual table cells had values of less than 5, which is not recommended for the chi-square test (Field, 2009). The few tentative recommendations were therefore recoded as recommendations against (0) and

recommendations for a progressive move were coded as 1. However, the three-point scale was not abandoned and was used in analyses reported below.

12. Also for parole applicants, the eventual Parole Board decision was recorded. This was coded as a decision either to grant or recommend a progressive move (1), or not (0). It was recognised that this was not ideal, because a progressive move could represent two different things. For a prisoner in closed conditions it could be a recommendation for a move to an open prison, or (very rarely) a direction to release. For a prisoner already in open conditions it could only be a direction to release. However, distinguishing between the two decisions would require a large sample of prisoners in each condition and this was not available within the timescale of the research. Only four of the 63 applicants were seen in open conditions, and of the other 59 only two were released directly from closed conditions, so "progress" almost always meant being recommended for open prison conditions.

Sample

Table 9 shows descriptive statistics for the entire sample. This consisted of 100 consecutive cases referred to a private psychological practice for psychological assessment. All participants were serving either a mandatory life sentence (i.e., murderers) or discretionary life sentence in England or Wales. That is, none of them were serving IPP sentences (see Introduction). All were interviewed by the candidate. Seventy-six had been given their life sentences for murder, two for attempted murder, ten for manslaughter, and twelve for rape. The two cases of attempted murder were combined with the murder cases for the purposes of statistical analysis. Eighty-four of

the men were eligible for parole, either because the minimum term of their life sentence had now been served or because they had never been given a minimum. These cases had been referred because their legal representatives felt that a psychological report, including a risk assessment, might help their parole application. Other cases were referred in connection with some other decision, such as an application to have their security category downgraded.

Most participants (N=84) had been given a minimum term which had expired, or had never been given one, and were therefore eligible for parole. Descriptive statistics for this group are summarised in Table 10. The mean age for this group was over 49 years, but 14 were over 60 and one was 79 years old. The range of criminal histories was also

Table 9: Descriptive statistics for the whole sample (*N*=100)

	Mean	S.D.	Range
Age now	48.08	11.61	22-79
Minimum term (years)*	12.95	4.44	<1-25
Years served	19.64	8.20	2-40
Years over*[a]	7.00	6.54	<1-26
Age at first conviction (years)	17.26	6.87	9-48
Mean prior sentence	10.36	13.35	0-60
Mean latency (months)	16.48	21.85	<1-143
Total prior convictions	8.38	6.14	0-23

* excludes those with no minimum term (*N*=87)

[a] additionally excludes those who had not yet completed their minimum term (*N*=71)

very varied, some participants having none and others having more than 20 previous convictions for which they had served a number of prison sentences averaging anything up to five years. The statistical analyses were done before three of the individuals had had their hearings, and four others dismissed their solicitors before their hearings so that the eventual outcomes of their hearings were not known, making the effective size of this group 63 for several analyses.

Analysis

Because of the way that the data were collected, that is, the fact that it was a convenience sample, not all of the data were available for every participant. This mainly reflected the different purposes for which assessments had been carried out. The PCL-R

Table 10: Descriptive statistics for parole-eligible participants (*N*=84)

	Mean	S.D.	Range
Age now	49.57	10.18	31-79
Minimum term (years)*	12.68	4.42	<1-20
Years served	21.05	7.80	2-40
Years over*	7.00	6.54	0-26
Age at first conviction (years)	16.89	6.32	9-48
Mean prior sentence length (months)	10.64	13.30	0-60
Mean latency (months)	13.87	13.55	<1-63
Total prior convictions	8.83	6.05	0-23

*excludes those with no minimum term (*N* = 71)

was completed for 94 of the 100 participants, the HCR-20 for 95, and the SVR-20 for 36. The SVR-20 was, of course, not appropriate for those without sexual offences. The HCR-20 was, however, carried out on most of those who were regarded as primarily sexual offenders. This was because the index offence was usually rape, which is violent as well as sexual, and implies a risk of future violence (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005), and because they had more previous violence offences than those who had received their life sentences for nonsexual violence. Similarly, the SVR-20 was carried out on men whose index offence was murder, but contained a sexual element, even if they had not been convicted of a sexual offence.

Age and most criminal history information were available for all participants, but in two cases the list of previous convictions was illegible because of poor photocopying. The individuals concerned were able to supply some of the missing information, but were often uncertain about dates and circumstances of previous offences, so this information was omitted as unreliable.

All data were entered into a Microsoft Excel spreadsheet which was later imported into IBM SPSS version 19. All statistical analyses were carried out using this program. Where missing information affected the numbers to be used in any given analysis, incomplete cases were excluded listwise rather than pairwise. This has the disadvantage of losing some data. However, excluding cases pairwise essentially means that analyses are often carried out on slightly different samples. Experimentation suggested that any differences would actually be small.

Results

Before carrying out any analyses the variables were examined for skewness and kurtosis, as parametric statistical tests require data to be normally distributed. The

following variables displayed significant skewness: PCL-R Facet 1, PCL-R Factor 1, PCL-R total score, HCR-20 C scale, HCR-20 R scale, SVR-20 FP scale, Years Over, Age at First Conviction, Mean Prior Sentence, Latency, and Total Prior Convictions. The same scales exhibited significant kurtosis, except for PCL-R total score, Years Over, HCR-20 R scale, and SVR-20 FP scale. Although parametric tests are generally regarded as quite robust to violations of their assumptions, they are less so when sample sizes are small (Field, 2009). Accordingly, nonparametric tests were used for many of the analyses, and this has been indicated in the tables of results.

Since the parole process was the main focus of the research, the initial analyses were conducted only on parole-eligible participants. Table 11 shows the criminological measures for those participants, broken down by index offence. One-way analysis of variance showed no significant difference in age between the offence categories. There was a significant difference in the minimum term awarded (F=29.26, p<.001), but no difference was found in the length of time actually served. Post hoc comparisons with Scheffé's test showed that rape attracted a longer minimum term than manslaughter (p<.01), and murder attracted a longer minimum term than either (p<.001). Logically, if minimum terms differ between groups but actual time served does not, one would expect that the groups would also differ in how far beyond their minimum ("Years Over") they had served. The variable Years Over was significantly skewed, but multiple Mann-Whitney U tests were used to compare offence categories, necessarily excluding the 14 cases who had never been given a minimum term. This showed that those convicted of manslaughter were significantly further past their minimum term than murderers (p<.02) but no other comparisons were significant, despite the large mean difference in Years Over between the rape and manslaughter groups. This was

Table 11: Criminological measures for parole-eligible participants broken down by index offence (*N*=84)

Index Offence	Murder/attempt (*N*= 62)			Rape (*N*=12)			Manslaughter (*N*=10)		
	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range
Age Now	49.45	9.30	31-69	45.92	9.26	34-58	54.70	14.74	38-79
Minimum term (years)*	14.33	3.28	7-20	9.67	3.14	5-14	4.67	3.78	0-10
Years served	21.45	7.26	10-40	17.67	8.57	5-32	22.60	9.73	2-36
Years over*	5.87	5.50	0-21	8.00	7.03	0-19	15.00	9.14	0-26
Mean prior sentence (months)	8.76	10.84	0-52	21.98	20.31	0-60	8.27	10.83	0-36
Prior sex offences	.26	.75	0-4	2.00	4.18	0-15	.40	.97	0-3
Prior violent offences	1.48	1.83	0-7	3.58	4.25	0-15	.50	.71	0-2
Age at first conviction	17.24	7.10	9-48	15.00	3.16	9-20	17.00	2.91	12-23
Mean latency (months)	14.03	14.20	0-62	9.88	6.51	2-24	15.07	18.35	1-63
Total prior convictions	8.50	6.21	0-23	10.42	6.02	1-20	9.00	5.23	2-20

*excludes those with no minimum term (*N*=71)

surprising, but when a box plot was generated of the three distributions, it was apparent that the distributions of Years Over in the rape and manslaughter groups overlapped considerably. This is shown in Appendix 5.

Other measures displayed significant skewness and/or kurtosis, and the Kruskal-Wallis test was used to make comparisons. These showed that the three offence groups differed in their history of sexual ($p<.001$) and violent ($p<.05$) offending prior to the index offence. Mann-Whitney U tests were used to test for differences between pairs of index offence types. Rapists specifically had significantly more *previous* sexual offences ($p<.001$) than either of the other groups but also significantly more previous violent offences ($p<.05$). The murder and manslaughter groups did not differ significantly in the number of previous sexual or violent offences. The difference in mean prior sentence length suggests that those who were given their life sentence for rape had (in the opinion of the courts) more serious previous convictions than those who were given it for violence. This raised the possibility that they might be more antisocial in general, and perhaps more psychopathic. Rapists exhibited significantly higher PCL-R Total scores than other lifers ($p<.007$). Comparisons were therefore made between rapists and others on the four facets of the PCL-R. Rapists scored significantly higher only on Facet 3 ($p<.008$). Facet 3 is characterised as a "lifestyle" factor by Hare (2003b), and comprises five items relating to need for stimulation, parasitism, lacking goals, impulsivity and irresponsibility. One might question whether stimulus hunger and impulsivity are lifestyle factors or aspects of personality, but either way the result is not easy to interpret.

A number of the murder convictions related to crimes in which a rape had taken place during the course of the murder, but the prisoner had not been charged with sexual

offences separately. In other words, some of the murderers were also rapists but were not charged with that. When the offences were coded according to whether they contained such a sexual element or not, rather than according to the legal definition of the offence, no differences in Facet scores were found. The apparent difference in Facet 3 could therefore be due to sampling differences, there being only 12 cases convicted of rape, but 34 convicted of an offence with a sexual element.

It was also possible to consider the context of homicides, as a full account of offences was given in the files. Homicides were therefore divided into "domestic" and others. "Domestic" was defined as the killing of someone who had been a member of the offender's household for at least six months. Other homicides included sexually motivated murders of strangers and acquaintances other than intimate partners, killings in the course of other offences such as robbery, and killings resulting from fights. There were two cases of witnesses to other offences having been murdered to silence them, and two in which drug traffickers had killed competitors. A considerable number of killings were committed under the influence of alcohol, and sometimes drugs as well. No differences were found between domestic and other cases in respect of PCL-R facets or the minimum term imposed. When cases with child victims were excluded, no differences were found in PCL-R facets, minimum terms, or other criminological measures. This does not suggest any great difference between domestic and other killers.

Indeterminate sentences are given in order to permit a prisoner's release when it is judged that he poses a minimal risk to the public, so in principle those serving the longest periods should be those displaying the greatest evidence of risk. The following analyses examined the relationship between three widely used risk measures.

Relationships between risk assessment instruments

Studies of the correlations between risk assessment instruments did not need to be
restricted to parole applicants, so these were based on the total sample. Table 12 shows
correlations between the various subscales of the HCR-20, SVR-20 and PCL-R. The
PCL-R score itself is one of the items on the Historical subscale of the HCR-20 (item
H7) and the Psychosocial Adjustment subscale of the SVR-20 (item 3); this item was

Table 12: Correlations (Kendall's τ_b) between PCL-R facets and subscales of the HCR-20 and SVR-20

	HCR-20 (excluding psychopathy item)			SVR-20 (excluding psychopathy item)		
PCL-R	Historical	Clinical	Risk management	Psycho-social adjustment	Sex offences	Future plans
Facet 1	.02	.20	.20	.14	.29	.42
Facet 2	-.01	.32	.26	.03	.09	.43
Facet 3	.40	-.02	.11	.44	.17	.06
Facet 4	.62	-.14	.10	.40	.09	.05
Light shading: *p*<.05 Dark shading: *p*<.01 (all one-tailed)	HCR-20 (excluding psychopathy item)	Historical		.46	.04	.06
		Clinical		-.15	.07	.37
		Risk mgmt.		.18	.02	.50

86

therefore removed from the scores of both to avoid artificial correlations between them. Kendall's τ_b was used, as the nature of the data would create a large number of ties, and this statistic is less affected by these than other correlation coefficients (Field, 2009).

Hypothesis 1 predicted that significant positive correlations would be found between those subscales considered to reflect more subjective judgement, namely, Facets 1 and 2 of the PCL-R, the Clinical and Risk Management subscales of the HCR-20, and the Future Plans subscale of the SVR-20. Table 12 shows that significant positive correlations between all of these subscales were found and the hypothesis is therefore supported.

In addition, one unexpected correlation was found between Facet 1 of the PCL-R and the Sexual Offences subscale of the SVR-20 (τ_b=.29, p<.05). Considering the individual PCL-R items, the Sexual Offences subscale was found to correlate with item 1, Superficial Charm (τ_b=.39, p<.005), item 2, Grandiose Sense of Self Worth (τ_b=.32, p<.01) and item 5, Conning/Manipulative (τ_b=.33, p<.01), but not with item 4, Pathological Lying. Apart from this, no relationships were found between the Sexual Offences subscale of the SVR-20 and any subscales of either of the other instruments. The relatively small sample size (N=36) could have been one reason for this, but the Sexual Offences subscale reflects the character of previous offending rather than the offender's current presentation, so correlations with the clinical measures were not expected. However, no correlations were found between this subscale and any criminological measure.

Hypothesis 2 predicted that significant positive correlations would be found between those items thought to reflect less subjective judgement, namely, Facets 3 and 4 of the PCL-R, the Historical subscale of the HCR-20, and the Psychosocial Adjustment

subscale of the SVR-20. This was because they were thought to be more firmly anchored in objective information about the offender's previous criminal record and thus less susceptible to variations in individual judgement. Table 12 shows that significant correlations were found between all of these subscales, and Hypothesis 2 is therefore supported. No correlations were found other than those predicted.

Taken together, these findings support the idea that Factor 2 of the PCL-R, the Historical scale of the HCR-20, and the PA scale of the SVR-20 all reflect general antisociality, which is measurable with some reliability from a person's previous criminal history. On the other hand Factor 1 of the PCL-R, the C and R scales of the HCR-20, and the FP scale of the SVR-20 may reflect the current clinical presentation of the person being assessed.

Risk measures and Years Over

Hypothesis 3 predicted positive correlations between the length of time served in excess of the minimum term and those subscales considered more subjective. In other words, positive correlations were predicted between Years Over and Facets 1 and 2 of the PCL-R, the Clinical and Risk Management subscales of the HCR-20, and the Future Plans subscale of the SVR-20. However, it was likely that those who had served longer terms would be older, so a correlation between Years Over and Age Now was calculated. Kendall's τ was .35 ($p<.001$, one-tailed), as predicted. Therefore, partial correlations were used, controlling for age. This analysis necessarily excluded those participants who had never been given a minimum term. It also excluded any participants who had not yet completed their minimum term, as they were not yet eligible for parole and had

been reported on by the candidate for other reasons. This reduced the sample to 71 participants.

PCL-R scores were considered first. Contrary to the hypothesis, no significant correlations were found between Years Over and the PCL-R Total, Factor, or Facet scores. The same result was obtained when age was not controlled for. A similar result was obtained for the subscales of the HCR-20, although there was a weak correlation between Years Over and the HCR-20 total score (r=.25, p<.05, one-tailed). Given that this was the only marginally significant result among several hypotheses, it may be concluded that little weight should be attached to it.

The result for the SVR-20 was rather different. Neither the Psychosocial Adjustment nor the Future Plans subscale was significantly correlated with Years Over. However, there were substantial correlations between Years Over and both the Sexual Offences subscale (r=.47, p<.05, two-tailed) and the SVR-20 Total score (r=.39, p<.05, two-tail). The Sexual Offences subscale concerns the character of a prisoner's entire sexual offence history, not just the index offence, and includes items relating to the intrusiveness of sexual offending and the use of weapons or threats. It therefore provides information about how unpleasant or frightening the offences were for the victims, as well as the attitudes of the offender himself. It is not clear that this information has any genuine relationship to risk (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005).

Risk measures and professional recommendations

The remaining analyses were carried out using only the 84 participants who were eligible for parole, whether or not they had been given a minimum term. On average,

professionals recommended 38% of applicants for a progressive move. However, the rates differed between professionals. The offender manager (external probation officer) recommended 48% of cases for progression, the offender supervisor (probation officer seconded to the prison) 43%, and the prison psychologist 22%. Taking 38% as the expected rate, a chi-square test was performed for each of the three professional groups. Only the prison psychologists' number of recommendations differed significantly from that expected at a rate of 38% (chi-square = 4.57, $p<.05$). Contingency tables were created (see Appendix 4) and Cohen's Kappa applied. This showed that the probation officers agreed significantly with each other (Kappa=.35, $p<.05$, 68% of cases). The prison psychologist agreed significantly with both the offender manager (Kappa=.30, $p<.05$, 68% of cases) and the offender supervisor (Kappa= .35, $p<.05$, 72% of cases).

It is not immediately apparent why prison psychologists should be approximately half as likely as probation staff to recommend parole. It is possible that the answer lies partly in professional status and experience. Although the precise number was not noted, many of the prison psychologists had trainee status, whereas none of the probation officers did. It is also possible that the psychologists, who were heavily involved in running offending behaviour programmes, may have been more aware of continuing treatment needs which they felt were exhibited by the prisoners. This would be consistent with the finding by Welsh and Ogloff (2000) that sex offenders were less likely to be paroled if they had completed sex offender treatment. Again, it is possible that they made different use of risk assessments. No comparison was made between assessments made by the candidate and those of his prison colleagues, but there may have been systematic differences. This would be consistent with the finding that

prosecution and defence experts are influenced by "allegiance" (Murrie et al., 2009). It has not been possible to answer this question on the basis of the current research.

Hypothesis 4 was that recommendations made by professionals would correlate negatively with the more subjective subscales of risk assessment instruments, namely, Facets 1 and 2 of the PCL-R, the Clinical and Risk Management subscales of the HCR-20, and the Future Plans subscale of the SVR-20. For each category of professional

Table 13: Point-biserial correlations between expert recommendations and risk assessment instrument subscales

Risk assessment instrument	Offender manager	Offender supervisor	Prison psychologist
PCL-R Facet 1	-.27	-.07	-.11
PCL-R Facet 2	-.55	-.32	-.06
PCL-R Facet 3	.17	-.04	-.20
PCL-R Facet 4	.12	.01	-.01
HCR-20 Historical	.03	.02	-.01
HCR-20 Clinical	-.43	-.37	-.33
HCR-20 Risk Management	-.49	-.23	-.26
SVR-20 Psychosocial Adjustment	-.05	-.04	.03
SVR-20 Sexual Offences	-.09	.17	.33
SVR-20 Future Plans	-.29	-.18	-.05
Light shading: $p<.05$; dark shading: $p<.01$ (all one-tailed)			

assessor point-biserial correlations were calculated between their recommendation and the subscales of the risk assessment instruments. The results are shown in Table 13.

It is clear that Hypothesis 4 has been upheld largely in respect of offender managers, partially in respect of offender supervisors, and barely at all in respect of prison psychologists. It is also clear that most of the correlations are rather weak, suggesting that professional recommendations are not strongly related to most of the subscales of risk assessment instruments. Where they are, however, they are related to the more subjective subscales, as predicted. Nonetheless, no significant correlations were found between professional recommendations and any of the SVR-20 subscales. This suggests that, consistent with other research, sex offenders are treated differently from other prisoners when it comes to assessing risk.

To pursue this further, professional recommendations were crosstabulated with index offence type in 2 x 2 tables, coded not according to the legal conviction but according to whether or not the offence contained a sexual element. Cohen's Kappa was applied, and no significant relationship was found between the presence of a sexual element and the recommendations of offender managers or prison psychologists. However, in the case of offender supervisors there was a clear bias against recommending progress if the index offence contained a sexual element (Kappa= -.39, $p<.005$, one-tailed). It is not immediately obvious why one professional exhibited a bias that the others did not, but it is possible that the answer lies in the amount of exposure experienced by that professional. In the UK, offender supervisors have more personal contact with their cases than prison psychologists, who normally only see people by appointment, or offender managers who may be based many miles away. The latter may only have occasional telephone contact with the offenders with whom they deal.

It is worth remembering that all of the risk assessments used for this research were conducted by the candidate. The candidate's risk assessments were not normally seen by the other report writers before they made their own recommendations. Therefore, although the latter may *correlate* with the candidate's risk assessments, they cannot have been *influenced* by them in most cases. This may not apply to the Parole Board, who would have seen the report before making their decision.

Professional recommendations and the Parole Board decision

Hypothesis 5 predicted that there would be significant correlations between the recommendations of professionals and the decision of the Parole Board. As these were all binary measures 2 x 2 contingency tables were created (see Appendix 4) and Cohen's Kappa applied. These showed that the Parole Board agreed significantly with the offender manager (Kappa=.44, $p<.005$, 68% of cases), the offender supervisor (Kappa=.33, $p<.05$, 67% of cases) but not the prison psychologist (Kappa=.25, $p<.08$, 66% of cases). Hypothesis 5 was therefore mainly upheld. The result for the prison psychologist approached significance, and although it represents weaker agreement it might well have become significant had the sample size been larger. The much greater level of agreement with the external probation officer matches the finding in other research that such officers' recommendations were influential.

There were cases in which professionals did not agree with each other, or agreed partially. There were also cases in which their support for a progressive move was conditional upon the offender undertaking some offence-related work, and was therefore regarded as tentative. As mentioned above (see Measures, paragraph 11) professional recommendations were coded on a three-point scale reflecting this range of

opinion. The scores were then added to create a combined Professional Recommendation score ranging from 0 to 6. This is certainly a crude measure, but it does combine the recommendations of professionals in a way which reflects the varying strengths of each, rather than assigning them all the same weight. This measure correlated very significantly with the Parole Board's decision (Kendall's τ_b= .59, p<.001, one-tailed). It is therefore clear that the Parole Board's decision was predictable to a significant extent from the professional recommendations.

Risk measures and the Parole Board decision

Hypothesis 6 predicted that negative correlations would be found between the Parole Board's decision and the more subjective subscales of risk assessment instruments. These were defined as Facets 1 and 2 of the PCL-R, the Clinical and Risk Management subscales of the HCR-20, and the Future Plans subscale of the SVR-20. Table 14 shows point-biserial correlations between these subscales and the Parole Board's decision. As with the correlations between risk assessment instruments and professional recommendations, Hypothesis 6 was upheld with respect to the PCL-R and HCR-20, but not the SVR-20. Once again, this suggests that sex offenders are seen in a different way by those who have to deal with them professionally.

From the literature it had appeared that sex offenders (rapists in this sample) might be less likely than others (nonsexual violent offenders) to obtain a positive decision from the Parole Board. In fact, 40% of offenders in each group obtained a positive decision, indicating no relationship between the Parole Board's decision and whether the index offence was rape or homicide. The same result was obtained when the index

offence was coded according to the presence or absence of a sexual element, rather than the legal definition of the conviction.

Table 14: Point-biserial correlations between subscales of risk assessment instruments and the Parole Board's decision	
Risk assessment instrument subscale	r_{pb}
PCL-R Facet 1	-.41
PCL-R Facet 2	-.48
PCL-R Facet 3	-.20
PCL-R Facet 4	-.06
HCR-20 Historical	-.09
HCR-20 Clinical	-.29
HCR-20 Risk Management	-.38
SVR-20 Psychosocial Adjustment	-.18
SVR-20 Sexual Offences	-.02
SVR-20 Future Plans	-.28
Light shading: $p<.05$; dark shading: $p<.01$ (all one-tailed)	

Exploratory regression analyses

Many previous studies in this field have used regression analyses to determine the variables which predict the outcome of parole proceedings. Unfortunately the size of sample in this study is small for this purpose, particularly considering that participants who were not applying for parole are excluded, and the large number of variables which might be entered into any regression calculation. However, the general hypothesis in this project has been that parole decisions can be predicted from a small number of variables, and will reflect recommendations from professional reports, which in turn will be influenced largely by the current clinical presentation of parole applicants. Therefore, whilst a larger sample would certainly be preferable, it was considered worthwhile performing regression analyses and entering only a small number of potential predictors. The first analysis concerned the prediction of the parole decision, and the second the prediction of the professional opinions which appeared to influence it. There are several methods of conducting regression analysis, but Field (2009) and Howell (1997) both recommend against using stepwise methods for hypothesis testing as opposed to data exploration. Accordingly, the forced entry method was used. The combined Professional Recommendation score was entered into a binary logistic regression analysis with the Parole Board decision as the dependent variable. The result was that Professional Recommendation correctly predicted 84% of the Parole Board decisions (84% of progress decisions, and 85% of non-progress decisions). This was an improvement on the 68% predicted by the offender manager's recommendation alone (the best predictor of the three professionals). Table 15 gives more details.

Table 15: Results of binary logistic regression analysis with Professional Recommendation as the independent variable and Parole Board decision as the dependent variable (all parole-eligible men, *N* = 84)

	B	SE	Wald	df	Sig.	Exp(B)
Professional Recommendation	1.08	.36	8.94	1	.003	2.95
Constant	-2.06	.72	8.08	1	.004	.13

Although regression is essentially a correlational technique, and not proof of causality, this analysis does suggest that the Parole Board decision is strongly influenced by the advice of professionals. However, as noted earlier, this simply moves the question one stage further back in the process. It is still not known what determines the expert decisions in their turn, although, following earlier literature, it was predicted that predominantly clinical variables would be the major influences. A multiple linear regression analysis was therefore carried out with Professional Recommendation as the dependent variable. Independent variables considered to be predominantly clinical were PCL-R Facets 1 and 2, the HCR-20 Clinical scale and the HCR-20 Risk Management scale. However, regression analysis is based on the assumption that independent factors are not strongly correlated. PCL-R Facets 1 and 2 have widely been found to correlate (Hare, 2003b), and they did in this sample (τ =.355, p<.001). The HCR-20 C and R scales also correlated to exactly the same degree, though correlations between the PCL-R and HCR-20 scales were lower. The PCL-R Facets were therefore combined into Factor 1, and the C and R scales into a combined scale termed "C+R". The two independent variables were entered into the regression by the forced entry method but

only the C+R scale produced a significant result. A simple linear regression analysis

was then run with this scale as the sole independent variable. The result is shown in

Table 16.

Table 16: Results of linear regression with C+ R measure as the independent variable and Professional Recommendation as the dependent variable (all parole-eligible men, $N = 84$)

	Beta	t	Significance
Constant		7.22	.001
C+ R measure	-.65	-4.41	.001

F=19.44, p<.001; Adjusted R^2 = .40

It is recognised that there are limitations of this analysis. In particular, both of the

regression analyses presented were carried out on samples which are small for this

purpose. The fact that only one independent variable was used does mitigate this

objection, but only to some extent. However, the main purpose of these analyses was to

illustrate how the pattern of data in this project is consistent with two findings of earlier

studies with other types of offender: the influence of professional reports on parole

decisions, and the predominance of clinical or subjective assessments of risk. In this

case, 40% of the variance in recommendations can be accounted for by a model using

only one clinical measure. However, it is again important to note that probation officers'

and prison psychologists' recommendations were made before those professionals had

read the candidate's report and been made aware of the HCR-20 scores. In other words,

the regression analysis does not imply that professionals were directly influenced by the

psychological report on the offender. Rather, it implies that professional

recommendations and HCR-20 scores both reflect the clinical presentation of the

offender.

It was thought possible that the C+ R measure might not be so predictive with sexual

offenders, as it is derived from an instrument intended to assess risk in nonsexually

violent offenders. It was possible that cases with a sexual element might be assessed

differently by professionals. As there were only 12 rapists in the sample a separate

regression analysis was not practicable for this group. However, index offences had also

been coded according to whether they contained a sexual element ($N = 33$) or not ($N =$

50), as many of the homicide cases did. Further exploratory regression analyses could

therefore be carried out on these two groups. Table 17 shows that the result for cases

Table 17: Results of linear regression analysis with C+ R as the independent

variable and Professional Recommendation as the dependent variable

(parole-eligible cases without a sexual element, $N = 50$)

	Beta	t	Significance
Constant		5.10	.001
C+ R	-.68	-2.95	.01
F=8.73, *p*<.01; Adjusted R^2 = .41			

without a sexual element was very similar to that for the whole group of parole-eligible

cases. However, this might be expected because the majority of parole-eligible men

would have fallen into this group in any case.

Table 18: results of linear regression analysis with C+ R as the independent

variable and Professional Recommendation as the dependent variable

(parole-eligible cases including a sexual element, $N = 33$)

	Beta	t	Significance
Constant		4.80	.001
C+ R	-.61	-3.02	.01

F=9.09, *p*<.01; Adjusted R^2 = .34

Of greater interest is the result for those cases in which there was a sexual element. Table 18 shows the results of the same analysis conducted with this group. As can be seen, the result is again very similar, although the proportion of variance in professional recommendation accounted for by the C+ R measure is not so high. Nonetheless, these analyses, although exploratory and conducted with small samples, do not suggest any substantial difference in the way that risk is assessed in cases with and without a sexual element. This is despite the availability of a risk assessment instrument, the SVR-20, specifically intended for use with sexual offenders.

An anomalous group

Finally, some results will be presented which might be expected in the light of the general stance taken in this thesis, but which in practice are difficult to interpret. These results relate to a small group of men who had never received a minimum term. When the death penalty for murder was abolished the arrangements for deciding minimum terms had not been put in place. For some years, therefore, offenders received life sentences without any minimum term being stated. Of the 84 parole-eligible cases in

this study, 13 fell into this group. Comparisons between their mean scores on a number

of variables were carried out. As no specific prediction had been made, and this analysis

was essentially exploratory, Bonferroni's correction was applied and a minimum

acceptable significance level set at $p<.003$. On this criterion the two groups did not

differ with respect to any of the criminological variables, except that the "no minimum"

group had fewer robbery offences (in practice, none) and fewer deception offences. As

expected, they were also older by almost 11 years, and had been involved with the

justice system for almost nine years longer in connection with the index offence.

However, this does not mean that they had served nine years longer in prison. An

important difference between the two groups was that the "no minimum" group were

significantly more likely to have been released from their life sentence at some point

and then recalled to prison because they had failed supervision ($p<.001$). In no case was

this due to a repetition of the index offence, and in most cases it was not due to an

offence at all. Nonetheless, of the 71 men who had been given a minimum term, three

(6%) had at some point been released and recalled after a short time. The corresponding

figure for the "no minimum" group was 11 out of 13 (85%). Some of these had been in

the community for a substantial length of time before recall. The longest had been at

liberty for 12 years. It was therefore clear that the group without a minimum term were

very different from the other men eligible for parole. This makes comparisons with

other groups problematic and difficult to interpret. Even so, these men had been

involved with the criminal justice system for a great deal longer than others with similar

offences. Furthermore, many had been recalled to prison from supervision in the

community without having committed any new offences. Given the general position

taken here it is logical to suggest that this may not always have been because of genuine

indicators of greater criminal risk, but because clinical assessments caused them to be *perceived* as higher in risk. Professional assessments are of crucial importance in such cases, because when lifers are recalled the decision is taken by the probation service, often at very short notice and without consulting other agencies. It was therefore predicted that the "no minimum" group would have higher scores on those subscales of risk assessment instruments which have been identified as more subjective in nature. Multiple t-tests were performed to test for differences between these two groups of prisoners on each subscale.

As predicted, the "no minimum" group obtained higher scores on PCL-R Facet 1 ($p<.01$), PCL-R Facet 2 ($p<.01$), the HCR-20 Clinical subscale ($p<.001$), the HCR-20 Risk Management subscale ($p<.01$) and the SVR-20 Future Plans subscale ($p<.05$). No other comparisons were statistically significant. This suggests that the criminal risk of those who had had prolonged contact with the criminal justice system was being assessed in a similar way to that of parole applicants. It also suggests that the decision to recall men to prison may be taken on similar criteria to the decision to parole them in the first place. Once again, the decisions cannot have been influenced directly by the scores, because they were not known to the professionals involved in the recall decisions. It is acknowledged that these group sizes are small, and that other interpretations are possible, but these results are consistent with the general hypothesis advanced in this thesis.

Discussion

The hypotheses which formed the basis of this research were derived from previous literature, mostly conducted on determinate-sentence prisoners, which broadly suggested two main effects. First, it suggested that Parole Board decisions would be

strongly influenced by the recommendations of professionals. Clearly, there is nothing inherently wrong with that: indeed, one might argue that parole decisions ought to be influenced by the recommendations of professionals whose expertise enables them to assess risk accurately. The second suggestion, however, was that professionals do not make recommendations on the basis of any special expertise, but on the basis of essentially subjective or "clinical" factors which are related to risk weakly, if at all. This research supports both of these findings, and extends them to life sentence prisoners.

The first set of findings relates to relationships between widely-used risk assessment instruments. Predicted relationships between those scales were found. These were in line with the findings of Rufino et al. (2011), who found that professionals agreed significantly about the objectivity of scoring guidelines for different instruments. They also found that interrater reliabilities were higher, the more objective those guidelines were perceived to be. Rufino et al. studied only the PCL-R and the HCR-20, but Facet 4 of the former and the Historical scale of the latter, which they identified as the most objectively-scored, were significantly correlated in this study, and both correlated significantly with the Psychosocial Adjustment scale of the SVR-20. These scales were also the most strongly related to reconviction in previous studies (Coid et al., 2007; Coid et al., 2011; Dietrich, 1994; Edens et al., 2010; Gray, Taylor, & Snowden, 2008; Mokros et al., 2010; Yang et al., 2010), and were not related in this study either to professional recommendations or to the Parole Board decision itself. This does not suggest that the Parole Board decision is strongly influenced by those indicators of risk which are supported by research.

It was predicted that correlations would be found between the more subjective or clinical scales and the variable Years Over (years served beyond the minimum term).

That is, higher ratings on these scales should be associated with parole-eligible men serving more time beyond their minimum. These expected correlations were not found. On the basis that professional recommendations may be influential in parole decisions, negative correlations were predicted between the various clinical scales and professional recommendations. This was generally supported, although no significant correlations were found between the SVR-20 scales and professional recommendations. There were, however, no unexpected correlations between any of these recommendations and the more objective scales of these instruments. That is, the more objective risk-related measures did not correlate with professional recommendations. There was therefore no support for the idea that professional recommendations were based on objective risk assessment, and considerable support for the idea that they were based largely on clinical or subjective considerations. It was expected that professional recommendations would show a strong correlation with the actual decision taken by the Parole Board. This was confirmed, although less strongly for the prison psychologists than the other professionals, perhaps because they recommended progressive moves at a much lower rate. It was also predicted that sex offenders would be less likely than violent offenders to obtain a favourable decision, but this was not supported. The fact that the C and R scales of the HCR-20 and Facets 1 and 2 of the PCL-R all correlated with the parole decision again supports the idea that it is clinical assessment and not objective assessment which influences the decision. A combined score representing the experts' recommendations predicted the Parole Board decision correctly in 84% of cases, similar to previous findings. A combined clinical scale accounted for almost 40% of the variance in the expert recommendation. Previous research suggests that such scales do not relate strongly to the risk of reconviction, if at all (Coid et al., 2007; Coid et al.,

2011; Mokros et al., 2010) and, where they do, measure only general criminal risk and not the more specific risk (such as violence risk) which was intended (Coid et al., 2007; Kroner et al., 2005).

The sample contained a group of parole-eligible men who had never been given a minimum term, and whose involvement with the criminal justice system in connection with the index offence had therefore been much longer. Apart from being older and having slightly different offence histories (especially an absence of robbery), they did not display any differences from other parole-eligible men on the variables measured. That is, their criminal histories were not found to be substantially different. These men had higher mean ratings on Facets 1 and 2 of the PCL-R, the C and R scales of the HCR-20, and the FP scale of the SVR-20, as predicted. These findings are not consistent with the lack of support for Hypothesis 3 (that these subscales would correlate with Years Over), but it may be that these risk-scale ratings are related to recall decisions in the same way as they are related to parole decisions. In that case high ratings would be expected in a subgroup of individuals whose involvement with the criminal justice system is prolonged. The fact that this group had high rates of release and recall to prison after failing supervision is consistent with this expectation. It must, however, be acknowledged that this part of the study is especially small in scale, and any interpretation should be regarded as tentative.

These results generally confirm the findings of earlier research with respect to the preference for clinical or subjective assessments in the expert recommendations (Bradford & Cowell, 2012; Garber & Maslach, 1977; Hood & Shute, 2000; Padfield & Liebling, 2000; Pogrebin et al., 1986; Porter et al., 2009), and the role of others' recommendations in the parole decision (Bradford & Cowell, 2012; Heinz et al., 1976;

Hood & Shute, 2000; Morgan & Smith, 2005a, 2005b; Zinger, 2012). The results also extend the findings of other researchers to include life sentence prisoners, who had previously not been extensively studied. Although the sample size in this research was much smaller than that in Hood and Shute (2000), a similar proportion of parole decisions was predicted by a measure of the expert opinion. As in Hood and Shute (2000) and Bradford and Cowell (2012), a larger proportion of parole decisions agreed with the recommendation of the offender manager (external probation officer) than that of any other professional report writer, despite this being the professional who had least contact with the offender. It could be argued that the offender manager, being based in the community, is more aware of the conditions to which the prisoner will be returning. However, even if this were shown to be significantly risk-related (and it has not been), it is clearly offset by the lack of knowledge of the offender concerned. Since this is the basis of the risk assessment, knowledge of the individual offender ought to be crucial.

It could, however, also be argued that those statistical risk indicators which have been shown to be relevant to risk in determinate-sentence men may not apply to risk in lifers, who are generally older and have spent much longer in prison. On this view, statistical predictors which relate to the man's behaviour a couple of decades earlier will not have much relevance to his risk today. It is fair to say that prediction generally becomes less reliable the further ahead one is trying to predict (Crighton & Towl, 2008), and by the same token predictive measures may be less useful the longer ago the measurements were taken. Although there may be truth in this argument, it does not render subjective methods of risk assessment any better. Rather, it suggests that more objective measures may be less useful for lifers than they are for those serving shorter

prison terms, especially as none exist which have been developed on lifer populations. This makes prediction more problematic rather than less.

Although risk is generally stated to be the paramount consideration of the Parole Board (Ministry of Justice, 2013a), there may be other considerations. In particular, progress to open conditions permits an offender a graduated return to normal society, rather than sudden release into a world which may have changed considerably since the start of his sentence. These considerations are not, of course, independent of risk, particularly when it is assessed by the use of a structured clinical instrument. Indeed, both the HCR-20 and the SVR-20 contain sections intended to guide future risk management. In the end, even rehabilitative efforts are intended to manage and reduce risk. Indeed, the evidence reported both here and in previous research (Bradford & Cowell, 2012; Heinz et al., 1976) that parole boards tend to follow the recommendation of external probation officers may reflect both considerations. External probation officers are likely to be aware of the local conditions to which the offender will return, and this may have implications both for his integration into the community and for risk.

The uncertainty inherent in predictions of this kind, and the potentially high cost of making a mistake, are likely to foster certain kinds of error. First, the likelihood is that a large amount of information will be sought, even though most of it is actually irrelevant to the question of recidivism risk. This gives the illusion that one understands the process involved, and can better predict what will follow. This is the error described as the "illusion of validity" by Kahneman (2011, p.209), who emphasises that humans are very good at creating a continuous narrative out of inadequate information, and then rationalising whatever decisions which they have actually taken. Tversky and Kahneman (1974) have drawn attention to the "anchoring" bias, which is the tendency

for decision-makers to be influenced by a prior suggestion regarding the outcome. It is probably shown here in the tendency of the Parole Board to follow expert recommendations, especially that of the offender manager. This is similar to the findings of other researchers (Bradford & Cowell, 2012; Heinz et al., 1976; Hood & Shute, 2000). The "availability" bias is likely to influence decisions in the direction of the information which is available, rather than that which is relevant. These biases will inevitably affect the quality of release decisions as they do the quality of economic decisions (Kahneman, 2011; Tversky & Kahneman, 1974).

The implications of the research are significant. Clearly, if the Parole Board serves mainly to confirm the decisions of other experts, that raises a question about its utility (Zinger, 2012). However, if the experts themselves are relying mainly on clinical or subjective assessments which are not significantly related to risk, that raises a question about the entire process. The latter question is by no means new. It was raised sixty years ago by Meehl (1954), who even then was able to show that statistical predictors of risk performed better than clinical predictors. He also suggested that clinicians who are invited to make a contribution will usually do so (Meehl, 1973), despite the fact that it may be irrelevant to the point at issue. That in itself may exemplify the illusion of validity and the availability bias.

Following publication of the work by Hood and Shute (2000) parole for determinate-sentence prisoners was abolished in the UK, and replaced by supervised "conditional release". This is essentially parole for all, although it can be revoked in the event of misbehaviour. However, the parole system remained for indeterminate-sentence prisoners because they do not have a fixed release date. Given that the Parole Board is formally required to make a judgement about risk it is concerning that there does not

seem to be an adequate basis for making this judgement. Furthermore, the Parole Board's decisions are not related to those risk indicators which are already known. For example, they correlate with Factor 1 of the PCL-R rather than Factor 2, and they are not related to age, which is a known and highly significant risk factor for sexual offenders (Barbaree et al., 2003; Wakeling et al., 2011), violent offenders (Mokros et al., 2010; Quinsey et al., 2006), offenders in general (Sampson & Laub, 2003) and parolees (Porter et al., 2001). With poor risk assessment, and the need to maintain public confidence in the parole system, the only strategy for limiting the release of dangerous individuals who reoffend seriously ("false negatives") is to hold down the number of *all* releases. A necessary corollary of this is that many offenders who could safely be released will be retained in custody ("false positives"). Sir David Latham, former Chair of the Parole Board, has complained that this already happens (Hill, 2010). This is consistent with earlier UK research (Bradford & Cowell, 2012; Hood & Shute, 2000; Hood et al., 2002). Latham also stated that society should decide what level of risk it is prepared to tolerate, but there is no indication that any politician is prepared to state publicly that *any* level of risk is tolerable.

There is little doubt that a statistical predictor could predict parole success or failure better than any other method (Hoffman, 1994; Hood & Shute, 2000; Meehl, 1954; Pogrebin et al., 1986; Quinsey et al., 2006), especially if it were developed specifically for a UK life sentence population. However, the rational use of a statistical method does have one drawback: it would require the specification of a tolerable level of recidivism and/or parole failure. As was suggested in the previous paragraph, this is politically unlikely. One alternative might be to create a presumption in favour of release at the conclusion of the minimum term. If this could only be extended in the light of objective

evidence of continuing risk, the Parole Board would still be able to extend custody if that evidence was available. However, they would have to make a case for doing so, and that case could be legally tested if it was disputed. A statistical risk assessment instrument would be of value in this process if it were shown reliably to identify dangerous individuals. One current problem is that statistical instruments are often better at identifying low risk than high risk individuals (Campbell, 2011; Campbell & DeClue, 2010). Another problem is that the applicability of instruments often varies depending on the type of offence and the type of individual (e.g., with respect to ethnic group or gender) to whom it is applied (Bartosh, Garby, Lewis, & Gray, 2003; Coid et al., 2009; Craig & Beech, 2009; Harris et al., 2003; Helmus, Hanson, Thornton, Babchishin, & Harris, 2012; Urbaniok et al., 2007). This may limit the utility of instruments developed in other countries, or standardised on other kinds of offender.

If inadequate risk assessment underpins Parole Board decisions, this logically implies not only that low risk individuals are being retained unnecessarily in prison, but also that some high risk individuals may be inappropriately released. This study did not directly examine this possibility, but the study by Porter and ten Brinke (2009), which suggested that psychopaths were adept at talking their way into parole, suggests that it is well worth investigating.

In the previous section ("An anomalous group") it was suggested that the risk assessment behind recalls from parole may be little different from the risk assessment used in paroling prisoners in the first place. If that is the case, it has implications for the effectiveness of recall decisions, since it is likely that most parolees who are recalled do not in fact pose a significant risk to the public, and have not committed further offences (Vito et al., 2012). However, when a prisoner is recalled he is delivered by the police

directly to a closed prison. Recalled prisoners are supposed to be granted a hearing in front of the Parole Board within six months, but one participant in this research had been waiting for two years. In any case, the prisoner will then have to start the whole process from the beginning, regardless of whether the Parole Board decides the increase in his risk was serious, or occurred at all. Given the establishment of an accurate risk assessment process one possible improvement in the system might be to provide a graded response, or degrees of recall. Thus, someone who was judged to have increased in risk only slightly might be recalled to a hostel where he would be subject to increased supervision. Another offender whose increase in risk was more serious might be returned to an open prison, and only the most serious cases would have to return to closed conditions. One advantage of this system would be that parolees might be better able to confide in the probation officers if they were having difficulties. At the moment, the prospect of causing anxiety in his supervising officer and precipitating a return to closed conditions is likely to deter a parolee from doing this. This is particularly likely where the offender himself recognises that he is tempted to reoffend and is seeking support.

Strengths and limitations

This study clearly has an important limitation in that it is based on a convenience sample. It cannot be shown that the sample of life sentence prisoners used is representative of the entire population of male life sentence prisoners, and females were not included at all. The participants had all been referred by their legal representatives because they felt that psychological evidence might be an issue in their cases, and this in itself may be a selection effect, perhaps selecting only the more complicated cases.

On the other hand, it was a relatively pure sample in that it consisted only of lifers and not a mixture of sentence types. The sample size was probably adequate for most of the analyses conducted, and probably inadequate for conducting regression analysis. However, its findings do not exist in isolation; they confirm those of previous studies conducted on other kinds of prisoner, which may increase confidence in the results.

One further feature of this research is that all of the psychometric assessment was carried out by the candidate. Although this rules out unreliability due to differences between assessors, it cannot exclude the possibility that the candidate himself had a bias. He had completed Hare's Darkstone training for the PCL-R and achieved "excellent" interrater reliability (intraclass correlation = 0.97), but training for the other instruments does not involve specific interrater reliability training. The risk assessment instruments were scored separately, rather than in one combined session, but this cannot exclude the possibility of some kind of unconscious bias. However, although this could have affected the correlations between scores on risk assessment instruments, it could not have produced the correlations between those scores and the recommendations of other professionals, which were made independently.

FOREWORD TO CHAPTER 4

The study reported in Chapter 3 confirms the finding reported in earlier research that parole boards were not much influenced by those sections of risk assessment instruments which are the most predictive of reconviction. It also confirms the widespread finding that professionals making recommendations about parole decisions are not strongly influenced by them either. Both the professionals and the Parole Board appeared to prefer a more subjective, clinically-based assessment.

This raises the question of what these instruments are actually measuring. The "coffee can" study (Kroner et al., 2005) suggested that they were all measuring much the same thing, namely, criminal risk. However, most of these risk assessment tools do not claim to be psychometric instruments. That is, they do not claim to be measuring an enduring characteristic of the person being assessed. Indeed, the authors of these instruments would accept that risk can vary, and would suggest that scores obtained using their instruments would vary accordingly (Boer et al., 1997; Webster et al., 1997).

There is, however, one exception to this. This is the Hare Psychopathy Checklist-Revised, which is widely used as a risk assessment instrument, but was actually intended as a measure of personality. This was the origin of the instrument, as previous definitions of psychopathy had not been standardised. As Hare (2003b) states "Properly used, the PCL-R provides a reliable and valid assessment of an important clinical construct — psychopathy. **Strictly speaking, that is all that it does**." (page 15, emphasis in original). The fact remains that it has been extensively marketed and used as a risk assessment instrument, because its scores were found to correlate with reconviction. It regularly happens that panels of the Parole Board for England and

Wales will be presented with PCL-R scores as part of a risk assessment by professionals, and will even ask for such an assessment to be carried out if it has not been.

However, recent research has raised doubts about the reliability of the PCL-R and its validity as a personality measure. Consequently, similar questions may be asked about its reliability and validity as a measure of criminal risk. In particular, if its interrater reliability is doubtful, it can be of little value as a risk assessment instrument. The following chapter therefore attempts a critique of the PCL-R. To review all of the research conducted on this instrument would require a book rather than a chapter, which must inevitably be something of an overview. The intention, however, is to address the shortcomings of the PCL-R, rather than to provide a complete evaluation of its use in all circumstances.

CHAPTER 4: CRITIQUE OF A PSYCHOMETRIC INSTRUMENT: THE PSYCHOPATHY CHECKLIST-REVISED (PCL-R)

Introduction

The Psychopathy Checklist-Revised (PCL-R) was developed by Robert Hare (Hare, 2003b) in a programme of research which began in the 1970s and continues to the present day. His purpose was to standardise the measurement of psychopathy, because at that time there were many different competing definitions of psychopathy and no standardised method of assessing it. Hare (Hare, 2003b; Hare, Hart, & Harpur, 1991) began by reviewing the literature which was already available, basing his work on the classic accounts of Buss (1966), Cleckley (1976), Craft (1965), and McCord and McCord (1964). Although these writers have concentrated on different aspects of psychopathy, there has been general agreement that psychopaths exhibit recognisable traits, such as impulsivity, irresponsibility, lack of remorse and empathy, grandiosity, and a variety of antisocial behaviours.

The PCL-R was developed from an earlier seven-item scale, and a 16-item instrument used by Cleckley (1976), although Hare (2003b) found the latter difficult to score. He therefore decided to construct his own scale based on statistical analyses of ratings on various items, to see which discriminated best between high and low scoring prison inmates. Even at this stage, the development programme was using criminal behaviour as a validity criterion measure. The original Psychopathy Checklist (PCL) (Hare, 1980) contained 22 items. Factor analysis suggested two underlying factors,

Factor 1 being associated with the core personality characteristics of the psychopath (i.e., selfishness, callousness and the remorseless use of others), while Factor 2 reflected an unstable and antisocial lifestyle. Cooke and Michie (2001) proposed an alternative analysis resulting in a three-factor model. This split the original Factor 1 into two and omitted four items from the original Factor 2. Hare (2003b) did not accept this reanalysis, pointing out that reducing the number of items reduces the number of potential factors, and querying the rationale for the choice of items by Cooke and Michie. Although he continued to argue for a two-factor solution, and the PCL-R has been modified to accommodate two "facets" within each factor, the distinction between the two underlying factors remaining.

Hare soon came under pressure to publish the instrument for use in forensic settings. The PCL-R was published in 1991 (Hare et al., 1991) and has since become widely used as a risk assessment instrument in its own right, as well as an item in other risk assessment scales, such as the Violence Risk Assessment Guide and Sex Offender Risk Assessment Guide (Quinsey et al., 2006), the HCR-20 (Webster et al., 1997) and the SVR-20 (Boer et al., 1997). Hare (2003b) has stressed that the PCL-R is intended to assess personality rather than criminal risk.

Description of the PCL-R

The PCL-R consists of 20 items, which are shown in Table 1. Each item is rated by the examiner on a three-point scale according to whether the characteristic in question is thought to be absent (0), present to some extent or maybe present (1), or definitely present (2). Thus the score can range from 0 to 40. Although a psychometric *instrument*,

Table 19: PCL-R items, Facets and Factors (numbers as on the PCL-R scoring sheet)

Factor 1: interpersonal/affective	Facet 1: interpersonal	1	Glibness/superficial charm
		2	Grandiose sense of self-worth
		4	Pathological lying
		5	Conning/manipulative
	Facet 2: affective	6	Lack of remorse or guilt
		7	Shallow affect
		8	Callous/lack of empathy
		16	Failure to accept responsibility for own actions
Factor 2: socially deviant/antisocial	Facet 3: lifestyle	3	Need for stimulation/proneness to boredom
		9	Parasitic lifestyle
		13	Lack of realistic, long-term goals
		14	Impulsivity
		15	Irresponsibility
	Facet 4: antisocial	10	Poor behavioural controls
		12	Early behavioural problems
		18	Juvenile delinquency
		19	Revocation of conditional release
		20	Criminal versatility
"Orphan" items		11	Promiscuous sexual behaviour
		17	Many short-term marital relationships

the PCL-R is not a psychometric *test* in the traditional sense (i.e., a questionnaire),

despite popularly being referred to as such (Ronson, 2011). It appears that the question

117

of treating the ratings "present to some extent" and "maybe present" as equivalent has never been raised. However, one reflects uncertainty that a characteristic is present at all while the other accepts that it is, but not in the maximum quantity. The issue of whether a three-point scale is sufficient also appears not to have been examined.

The professional manual contains definitions of the 20 items, with guidelines for scoring them. Hare suggests that a score of 30 or more be taken as indicating the presence of psychopathy. Initially it was thought that this cut-off should vary from one culture to another, as some analyses suggested that the same score represented different levels of psychopathy in different populations. For example, a score of 25 was widely used in the UK at one time, following Cooke and Michie (1999), whose work, based on item response theory, suggested that this represented the same degree of the underlying trait as a score of 30 in America. However, as the body of research has increased there has been a tendency to accept the cut-off of 30 internationally (Hare, 2003b).

Level of measurement

A ratio scale is the ideal form of measurement for scientific purposes, but traditional psychometric measures do not reach this ideal (Kline, 2000). As they lack a true zero point they are assumed to be interval scales, although it is not always clear that they are: it is not clear that the difference between two adjacent points on the scale is always the same no matter which two we choose.

In the case of the PCL-R, the presence of a zero point might appear to indicate a ratio scale. However, this is artificially imposed by the method of scoring, and it is at best an interval scale: one cannot claim, for example, that someone with a score of 20 is twice as psychopathic as someone with a score of 10. Kline (2000) suggests that the

assumption of an interval scale is tenable provided that an instrument is of practical utility, though the establishment of ratio scales ought to be a long-term goal for psychology.

Reliability

Interrater Reliability. The PCL-R depends critically upon the assessor's ability to rate people's characteristics on a simple scale. Mindful of the known issues around the reliability of clinical judgement (Meehl, 1954, 1973; Quinsey et al., 2006; Quinsey & Maguire, 1986), Hare formed a company providing training intended to ensure interrater reliability. However, following challenges from practitioners his website (http://www.hare.org/index.html) now acknowledges that Hare's training is neither the only way to achieve proficiency in administering the PCL-R, nor proof of accuracy.

Studies of interrater reliability for the PCL-R show generally favourable results, at least among academic research groups. Using several samples totalling 925 prisoners, Hare, Harpur, Hakstian, Forth, Hart and Newman (1990) reported interrater reliabilities (intraclass correlation coefficients) from .82 to .93 for assessments based on single assessors' ratings, and .87 to .97 for assessments where two ratings were averaged. Schroeder, Schroeder, and Hare (1983) reported interrater reliabilities ranging from .84 to .93 in imprisoned offenders. Most researchers have studied male offenders, but Vitale, Smith, Brinkley, and Newman (2002) obtained reliabilities in excess of .95 for female offenders.

Hare (2003b) advises that an interview should be part of the PCL-R assessment, but states that it is possible to use file information alone, provided it is comprehensive. Grann, Långström, Tenström, and Stålenheim (1998) tested this in Sweden, obtaining a

reliability of .88 for PCL-R total scores. The corresponding figure for Factor 1 scores was .69 and for Factor 2 scores .89. The lower figure for Factor 1 may support Hare's belief that an interview is preferable for rating the interpersonal and affective facets of the PCL-R.

International support for the reliability of the PCL-R has come from studies in the UK (Hobson & Shine, 1998), the Netherlands (Hildebrandt, de Ruiter, de Vogel, & van der Wolf, 2002), while a paper by Hare, Clark, Grann, and Thornton (2000) includes a review of international studies. This work supports the interrater reliability of the PCL-R across different populations. However, recent work (Boccaccini, Turner, et al., 2008; Edens et al., 2010; Mokros et al., 2010; Murrie et al., 2012; Murrie et al., 2009) has found Factor 1 reliabilities as low as zero among individual practitioners "in the field", which seriously undermines the PCL-R's use in practical risk assessment as opposed to academic research.

One possible source of unreliability is subjectivity in clinical judgment. Hare (2003b) acknowledges that clinical judgment is required to score the PCL-R, but seems to assume that clinicians will make objective judgments, which is by no means certain (Neal & Grisso, 2014).

Test-retest reliability. Since Hare (2003b) regards psychopathy as a lifetime condition the test-retest reliability of the PCL-R should be high, and it is surprising that few studies have tested this. Hare cites only Schroeder et al. (1983), who obtained reliability figures ranging from .84 to .93 in five prison samples — comfortably above the generally accepted minimum of .70. However, they used a test-retest period of less than a year. Rutherford, Cacciola, Alterman, Mackay, and Cook (1999) obtained figures of only .60 to .65 for 225 male and female methadone patients over a two-year period.

For men (but not women), reliabilities were significantly lower for Factor 1 than Factor 2. Similarly, a study of 88 substance abuse patients over a one-month period yielded figures ranging from .60 to .74 (Alterman, Cacciola, & Rutherford, 1993). Importantly, the Schroeder et al study was conducted on the original Psychopathy Checklist. The two studies which yielded poorer figures were both conducted on the PCL-R itself. This highlights an important research gap, given that the PCL-R is meant to be a stable lifetime measure.

Internal consistency. Because the PCL-R is not a psychometric test it does not *prima facie* lend itself to some traditional analyses. For example, each item on the PCL-R is intended to tap a different aspect of psychopathy, which makes traditional split-half reliability testing problematic, although it should be possible with subscales. Hare (2003b) has reported several analyses of item-total correlations. However, a number of these relate to the original PCL rather than the PCL-R, which contains some different items. Hare reports analyses with the PCL-R itself, based on Swedish and UK offender samples, which show similar results. Item-total correlations range between about .40 and .60.

Hare (2003b) also reports Cronbach's alpha coefficients for these samples. These are .84 for the UK sample total score, and .87 for the Swedish sample. The factors and facets achieve lower coefficients, which cluster around .70, which is acceptable. Again, given the popularity of the PCL-R, and the fact that it has been published for 21 years, it is surprising that more studies of this type have not been carried out. Nonetheless, published studies suggest that it is a reliable instrument.

Hare (2003b) cites a number of (then unpublished) studies which have examined the PCL-R from the point of view of item response theory (IRT). His assessment of the

evidence from these studies is that the PCL-R "is a homogeneous scale in which all items contribute to the measurement of a single core construct or superordinate factor" (p. 69). He later states "one implication of the preceding analyses is that interpersonal and affective items (Factor 1) are more important in measuring and generalising the construct of psychopathy than most (not all) of the socially deviant, antisocial items (Factor 2)" (p. 75). However, if two factors exist and one yields more information than the other it is not clear that the PCL-R is in fact homogeneous. Indeed, Hare suggests that Factor 1 yields more information at higher levels of psychopathy and discusses the advisability of dropping some of those items which contribute less than others. It is hard to reconcile this with the poor reliabilities reported for Factor 1 "in the field". Bishopp and Hare (2008) used a multidimensional scaling approach to examine the structure of the PCL-R, concluding "While it is useful to describe the PCL-R in terms of two, three or four factors we should not preclude the possibility that it may be made up of many more" (p. 128). Again, this is hardly support for the PCL-R as a homogeneous scale.

Bolt, Hare, Vitale, and Newman (2004) found significant differences in the way that the PCL-R items yield information in female offenders, male forensic psychiatric patients, and male offenders assessed by file information only, when compared with a reference group of male offenders assessed by file information and interview (i.e., the standard method). In the cases of female offenders and male patients there were significant differences in more than half the items. Taken together, these studies suggest that the PCL-R's internal consistency, although widely assumed, is not as well established across different groups as one might wish, and most practitioners appear to believe.

Validity

The validity of a psychometric instrument is normally established by correlating it with various other measures. In the case of the PCL-R the choice of these can be problematic, and have led to serious criticisms of the instrument (see below). However, efforts have been made to establish validity by relating PCL-R scores to other instruments and to outcome measures, and norms for suitable groups have been provided. Much of this work has been reported in the PCL-R manual (Hare, 2003), and more has since been published.

 Concurrent validity. Hare (2003b) cites a number of studies which have correlated PCL-R scores with clinical ratings of psychopathy. Some of these were actually obtained using the original PCL; correlations between PCL/PCL-R total scores were in the range .80 to .90, which is acceptable. Several other studies are cited relating total scores to a diagnosis of Antisocial Personality Disorder (ASPD). These generally produce correlations of moderate size between total and Factor 1 scores, and higher correlations between ASPD and Factor 2. This is consistent with Hare's belief that Factor 2 is a measure of the antisocial behaviour associated with ASPD, whereas Factor 1 represents the "core" personality characteristics of the psychopath.

 Hare (2003b) also cites research which has correlated PCL/PCL-R total and factor scores with self-report measures of psychopathy. There are considerable problems in doing so, because many of these instruments did not necessarily use the same definition of psychopathy as Hare. In general the correlations are weak or modest (from .04 to .50) but in the expected direction. As Hare acknowledges (Hare, 2003b, p. 92) the correlations between Factor 2 and the self-report measures are generally higher, suggesting that the latter are measures of ASPD rather than genuine psychopathy.

123

The Personality Assessment Inventory (PAI) (Morey, 1991) has been extensively used with forensic populations, and contains three measures of antisociality (antisocial behaviour, egocentricity, and stimulus-seeking). Hare (2003b, p.94) reports three studies which have correlated these with the PCL-R. These showed modest correlations between the PCL-R total score and the antisocial scales of the PAI (.29-.53), and higher correlations with Factor 2 (.17-.61). Correlations with Factor 1 were only found in female offenders. Morey (2007) has added more recent studies, with correlations in the same range for male offenders, and correlations with Factor 1 again found only in female offenders. The explanation for the gender difference is not clear.

Shine and Hobson (1997) reported correlations between the Personality Diagnostic Questionnaire Revised (PDQ-R) and the PCL-R in English male prisoners. The PDQ-R provides scores corresponding to recognised personality disorders, and none of these correlated significantly with Factor 1 of the PCL-R, although many of them include emotional shallowness, lack of empathy, and so forth, which supposedly define that factor. Some PDQ-R scores correlated significantly with Factor 2, in particular those for antisocial, borderline, histrionic, and sadistic personality disorders. Once again, these provide support for Factor 2 of the PCL-R (but not Factor 1) and suggest it relates to generally disinhibited behaviour. These findings suggest that it is mainly Factor 2 of the PCL-R which relates to other psychometric measures, and confirm its status as a measure of antisocial personality disorder rather than psychopathy specifically.

Content and Construct validity. Hare has never disputed that Factor 2 of the PCL-R may relate to ASPD, suggesting that psychopaths are, in effect, a subgroup of those with ASPD. In order to achieve a score of over 30 and meet Hare's criterion for psychopathy one would have to score highly on both factors of the PCL-R.

Furthermore, the concept of psychopathy involves much more than just unruly behaviour.

Hare (2003b) has reported a large number of studies demonstrating that PCL-R psychopaths have difficulties in processing emotion. For example, they react with equal speed to emotional and neutral words, whereas non-psychopaths react more quickly to emotional ones. In another study, psychopaths who read stories with emotional content were able to attribute emotions to the characters, except for the emotion of guilt (which they theoretically should not feel themselves). Studies like this do suggest that the PCL-R may be tapping some of the affective deficits expected by Hare. However, Hare (2003b) states that offenders should be rated according to their clinical presentation, regardless of how it arose. Some PCL-R items (e.g., shallow affect) occur in other conditions, such as posttraumatic stress disorder and Asperger's syndrome (American Psychiatric Association, 1995; Comer, 1998), which are unrelated to psychopathy, and this could reduce the validity of ratings for some individuals.

Others have attempted to confirm these effects in more recent studies. For example, Bagley, Abramowitz, and Kosson (2009) showed that psychopaths were less able than others to classify correctly the emotion being expressed in recorded spoken sentences. However, although results were in the expected direction, correlations with PCL-R scores were low and few were statistically significant. Verona, Patrick, Curtin, Bradley, and Lang (2004) examined the physiological responses (heart rate and skin conductance) of psychopathic and non-psychopathic offenders to pleasant and unpleasant sounds. They found that those scoring highly only on Factor 1 of the PCL-R showed reduced skin conductance to sounds with emotional connotations (whether positive or negative) and those scoring highly only on Factor 2 showed a slower heart

rate reaction. Interpretation of these findings is difficult if both Factors are measures of the same construct but correlate differently with physiological measures.

Lösel and Schmucker (2004) assessed 49 German prisoners using the PCL-R, and tested their performance on a measure of attention, a gambling task, and a measure of IQ. They hypothesised that psychopaths would take riskier decisions on the gambling task. This was not supported, but they did find that psychopaths with poor attention took worse gambling decisions than those with normal attention (there was little difference in non-psychopaths).

There are many more studies of this type. Unfortunately, the results are often contradictory, and they tend to have been carried out on small samples, which are notoriously prone to extreme effects (Copas & Jackson, 2004). Copas and Jackson outlined a "bound for publication bias" hypothesis, suggesting that studies which fail to show the expected results are often left unpublished, making published ones look stronger. This risk was highlighted by Edens and Campbell (2007), whose meta-analysis of PCL-R effects in young offenders showed that published studies reported effects more than twice as powerful as unpublished ones, and that this could not be attributed to methodological superiority.

Finally, Hare (2003b) has suggested that the PCL-R is derived from Cleckley's (1976) descriptions of psychopathy, but it is not clear that all aspects of Cleckley's model are represented in the PCL-R. In particular, Cleckley and others have suggested that failure to learn from experience and/or punishment is a defining characteristic of the condition. This is not represented anywhere in the PCL-R, although it is arguable that some of its consequences (such as repeated imprisonment) might be proxy measures.

Predictive validity. There is little doubt that PCL-R scores correlate significantly with subsequent reconvictions. Hare (2003b) presents evidence for this, as have subsequent studies, e.g., Quinsey et al. (2006), Coid et al. (2007), Hilton, Harris, Rice, Houghton, and Eke (2008). This does not necessarily support the PCL-R as a measure of psychopathy. As Gendreau, Goggin, and Smith (2002) pointed out, predicting reoffending does not make the PCL-R more than a risk assessment instrument. Hare (2003b) has responded by saying that the researchers are comparing "apples and fruit salad" or being "disingenuous" and "parochial and myopic" (p. 147). However, this does not answer the criticisms.

A significant study concerning predictive validity was reported by Kroner et al. (2005), and severely undermines the PCL-R as anything more than a predictor. The authors took four instruments (the PCL-R, the VRAG, the GSIR and the HCR-20, all roughly equal in predictive power) and drew items from them at random to construct four pseudo-instruments. Each of the pseudo-instruments predicted as effectively as the original four. The authors concluded that each of the four simply measures criminal risk, with higher risk being reflected in higher scores.

Other studies have tested the idea that psychopaths should find it difficult to keep to the disciplinary code in prisons. Kroner and Mills (2001) found that the PCL-R score was correlated with the number of minor infractions, but not major ones. A similar result was obtained by Buffington-Vollum, Edens, Johnson, and Johnson (2002), who found that the PCL-R correlated with verbally aggressive infractions, but not physical aggression, in sex offenders. Hare (2003b) interprets these studies as support for the PCL-R, but the correlations were mainly with Factor 2. Given Hare's unitary model of psychopathy one might have expected a relationship with Factor 1.

Norms

Hare (2003b) presents norms for the PCL-R, based on North American male offenders, female offenders, and male forensic psychiatric patients. He also presents norms for English male prisoners, based on data supplied by HM Prison Service. For a British practitioner these data enable comparisons to be made between an individual prisoner and a prison reference group, but they do not enable such comparisons to be made for offenders in the community, or non-offenders. Indeed, there are no norms for the general population, which limits the range of comparisons available to practitioners. This underlines the degree to which the PCL-R has become associated with criminal risk assessment rather than personality assessment in general. Hare (2003b) cites research suggesting that the average PCL-R score in the general population is about eight points, and DeMatteo, Heilbrun, and Marczyk (2006) found it to be 14, but these findings do not constitute norms.

Controversy and an alternative formulation

A number of criticisms of the PCL-R have been made by Skeem, Polaschek, Patrick, and Lilienfeld (2011), following a public dispute (Hare & Neumann, 2010; Skeem & Cooke, 2010a, 2010b). A major concern of Skeem and Cooke is that the PCL-R exhibits criterion contamination: its score is partially, but significantly, derived from criminal behaviour, and this is used to predict criminal behaviour. They cite copious research demonstrating that it is mainly Factor 2 of the PCL-R which does this, and especially facet 4, which relates to such things as failure on parole and criminal versatility. In other words, past criminal behaviour predicts future criminal behaviour, which is an unsurprising conclusion for four decades of research. They question whether criminal

measures should be used to measure psychopathy, as there may well be psychopaths who do not have a criminal record.

Hare and Neumann (2010) have replied with counterarguments reaffirming the role of antisocial/criminal behaviour in psychopathy. However, these appear weak in light of the fact that Hare has written a popular book containing a chapter on how to deal with noncriminal psychopaths in one's personal life (Hare, 1999), and co-authored another on noncriminal psychopaths in business (Babiak & Hare, 2007). If many psychopaths do not have a criminal record it follows that such data cannot be regarded as fundamental to the construct. Yet on Hare's formulation, Factor 2 (including the criminal information) is a necessary part of the score (though not sufficient, as the true psychopath must score highly on Factor 1 as well).

Skeem et al. (2011) suggest that the PCL-R has become the accepted *definition* of psychopathy, because of its almost universal use, and that this hinders refinement of the construct. They suggest that researchers have ignored many studies demonstrating that psychopathy consists of more than one dimension, and that there may be different types of psychopath, corresponding to different positions on these dimensions. They propose a "triarchic" model in which three factors (disinhibition, boldness, and meanness) interact.

Disinhibition is seen as an impulse control impairment, comprising lack of foresight, poor emotional regulation, a desire for immediate gratification, and poor behavioural controls. Skeem et al. (2011) envisage consequences including irresponsibility, untrustworthiness, reactive aggression, and substance abuse problems. They are seen as being similar to some of the items on the PCL-R Factor 2. Boldness is manifested in low emotional reactivity, thrill seeking, and social assertiveness. Although some of

these might be seen as positive, they also present evidence that these characteristics are linked to narcissism and lack of empathy. Meanness is related to lack of empathy, along with a disdain for close emotional attachments to others, exploitativeness, and arrogance. Some of these characteristics are similar to Factor 1 items of the PCL-R, but Skeem et al. (2011) believe that each of these three dimensions can be manifested in ways which are unlikely to lead to involvement with criminal justice. This marks a clear distinction between their model and the PCL-R.

According to Skeem et al., this model has the advantage of accommodating multiple definitions of psychopathy. They suggest that Hare (2003b) has viewed psychopathy as a single dimension, whereas classic descriptions suggest more than one type. Even so, he has had to concede the existence of two separate, albeit correlated, factors. Skeem et al. suggest that the PCL-R describes a disinhibited aggressive and "mean" type of psychopath, but loses some characteristics of the other types. By mapping these different types onto three dimensions they hope to provide a more inclusive model.

Skeem et al. also propose the use of a questionnaire measure, the Psychopathic Personality Inventory (PPI) (Lilienfeld & Widows, 2005). They believe that the three dimensions of their model map onto the PPI reasonably well, and that it is particularly important in measuring the dimension of boldness. Psychopaths are notoriously deceitful, which may limit the use of questionnaire measures (Vien & Beech, 2006). On the other hand, recent research also suggests limitations to the methods of the PCL-R, such as a tendency for raters to obtain higher or lower scores depending on whether they appear for the defence or prosecution in cases (Edens & Campbell, 2007; Murrie et al., 2009) and effects of rater personality on scores (Miller et al., 2011). It is possible to include multiple checks for social desirability responding and manipulation in a

questionnaire, as is already done quite successfully by the Personality Assessment Inventory (Morey, 1996; Rogers, 2008).

The Skeem et al. (2011) review is recent, long, and covers a great deal of research which has attracted little attention. It contains too much material to review here. Clearly it represents a different approach to the construct of psychopathy and its measurement. Part of the authors' stated intention is to inject some new thinking into a field which they believe has become reified by an overreliance on the PCL-R and the unidimensional conception of psychopathy. At present their model appears to be the only serious challenger to the PCL-R's dominant position.

Conclusion

The PCL-R has become the leading measure of psychopathy. When it was begun there was a need for more systematic thinking about an ill-defined construct, and Hare has performed a valuable service in attempting to standardise both the construct and the method of its assessment. However, the dominance of the PCL-R has tended to stifle the further development of the construct of psychopathy and its measurement. To most researchers in this field, the PCL-R *is* psychopathy; other measures are rarely used nowadays.

One reason may be an appearance of scientific rigour. Goldacre (2008) points out that people like numbers because they seem precise. In fact, with a standard error of measurement of about three points the PCL-R is not especially precise. To be certain that there was a significant difference between two raters' assessments they would have to be about six points apart, which is a large amount on a 40-point scale. However, the PCL-R has been intensively (and lucratively) marketed. There are approved scoring sheets, structured interview booklets, and training courses. The package comes ready-

made, which makes it easy to adopt. There is an illusion of standardisation, which contributes to the scientific appearance.

The fact is that the psychometric properties of the PCL-R are not all that they could be. The test-retest reliability has not been properly established, and the little research evidence which exists is mixed. Likewise, few studies of the PCL-R's internal consistency have been carried out. Recent studies have raised serious doubts about the interrater reliability, which — along with test-retest reliability — is absolutely crucial to the viability of the PCL-R.

The validity of the PCL-R may be doubted because of its content, which may leave some aspects of classical psychopathy untapped, and because other research questions the unidimensional nature of psychopathy, which is also crucial to Hare's model. That has never been altogether comfortable, given the existence of two factors. Furthermore, many of the predictive validity studies have only established substantial correlations between criterion measures (such as reconviction) and Factor 2. This criterion contamination introduces circularity into the measurement of the construct: criminal behaviour predicts criminal behaviour. Despite misgivings by Hare and his colleagues (Hare, 1998; Zinger & Forth, 1998), this has served to push the PCL-R into the field of prediction and risk assessment, for which it was not designed.

The PCL-R has been in use for two decades, but limited norms have been established, and only for imprisoned offenders. This limits the application of the PCL-R and entrenches it further in the field of criminality, despite Hare's own acknowledgment that psychopaths need not necessarily be criminals. The alternative view of psychopathy proposed by Skeem et al. (2011) offers a new conception of the construct and how it may be measured. It is too early to say whether it is *the* way forward, but it is certainly a

possible way forward. It has the advantage that it promises to include types of psychopathy which are not fully addressed in the PCL-R.

The history of the development of the PCL-R illustrates well how a standardised instrument can attain the position of being the "gold standard" in a particular field without necessarily fulfilling all the basic requirements of a psychometrically sound instrument. In part, this has been due to commercial demands once the instrument has been published, rather than the demands of scientific research and practice.

CHAPTER 5: GENERAL DISCUSSION

Indeterminate sentences can result in prisoners serving extremely long periods of detention. This is costly to the prison system, and arguably inhumane if the detention is more prolonged than necessary. This is particularly important in the UK, where determinate-sentence prisoners are now released automatically halfway through their sentences, and parole is almost exclusively available to prisoners on indeterminate sentences. In most jurisdictions, and specifically in England and Wales, parole panels are expected to parole life sentence prisoners at a certain point provided they are satisfied that the risk to the public is low enough. This implies that adequate means of assessing that risk are available, and that parole panels will be capable of using the results of such assessments to improve their decision-making. However, this implication has largely gone untested.

In Chapter 2, the results of a systematic literature review on the determinants of parole decisions were reported. Most of the studies found were American, but there were also some from Canada, the UK, and Israel. The review showed that there was no consistent practice in parole procedures. In some jurisdictions cases were subject to prolonged deliberation, and in others parole applications received only a few minutes' consideration, or even less. Practice was not always determined by those within the parole system. For example, in some jurisdictions legal rulings had affected the process. This meant that practices were not always consistent within the same jurisdiction at different time periods. Since many of the studies were carried out some years ago, practices could easily have changed since, and the generalisability of findings was limited.

Furthermore, different factors were found to be predictive of parole decisions in different jurisdictions. In some cases parole boards claimed to take decisions on the basis of factors which were objectively not related to those decisions. Equally, in others they claimed not to be influenced by some factors which were indeed related to their decisions. However, although there was little consistency between jurisdictions, within each one parole decisions were usually predictable from a small number of factors. Most of these factors were not objectively related to the risk of reconviction or parole failure. The few studies of the parole process for UK lifers suggested that there was no systematic risk assessment (Bradford & Cowell, 2012; Padfield & Liebling, 2000), with the Parole Board making its decisions on the basis of "clinical" or subjective opinions of risk. These have long been known to be unreliable (Meehl, 1954), and may also provide scope for the more manipulative parole applicant to present himself in such a way as to obtain parole despite being high in risk (Porter et al., 2009).

Nonetheless, essentially subjective methods of risk assessment predominated. Some parole boards refused to make use of objective methods even when they were offered, and the desire to maintain parole boards' discretion to decide as they wished appeared to be widespread. One consequence was that prisoners whose offences provoked anxiety, such as sex offenders, had higher levels of risk attributed to them than was warranted by the evidence. Furthermore, parole boards need to maintain public confidence by keeping down the serious reoffending rate among parolees. In the absence of systematic and objective risk assessment the only available strategy for achieving this is to keep down the level of all releases, ensuring that many low-risk individuals who could in fact be safely released are retained in custody.

In the period since many of the studies were conducted, structured anchored clinical judgements (also called structured professional judgements) have been developed. These include the HCR-20 for violent offenders, the SVR-20 for sex offenders, and the PCL-R. The latter was originally intended to be a personality assessment instrument for the measurement of psychopathy, but has come to be used as a risk assessment measure. This raised the question as to whether modern practice in England and Wales might have improved with the advent of these instruments, which are often used nowadays to present risk estimates to the Parole Board. An additional consideration was that many of the published studies have excluded life sentence prisoners, or mixed them with other types, and it was not clear whether similar results would be obtained with a pure lifer sample. Since a source of data was available to the candidate in the course of his work, it was decided to conduct a study of life sentence prisoners in England and Wales, focusing on the use of risk assessment instruments and the impact which they might have on parole decisions.

Chapter 3 reported on this study, which examined a sample of 100 life sentence prisoners. Of these, 84 had been assessed in connection with parole proceedings and 16 for other reasons. The latter were included because they had undergone the same assessments and their data could contribute to study of the relationships between assessment instruments. Rufino et al. (2011) had established that the subscales of risk assessment instruments could be rated for their objectivity by professionals, and that these ratings correlated with their accuracy in predicting reconviction. The literature review in Chapter 2 had shown a preference on the part of parole boards for subjective risk assessment and a desire to maintain their discretion in parole decision-making. It was therefore predicted that the Parole Board for England and Wales would make

decisions reflecting parole applicants' scores on the more subjective subscales of these instruments rather than the more objective ones. This prediction was largely upheld for the HCR-20 and the PCL-R. However, the SVR-20 scores did not relate to parole decisions. This was not reflected in a lower rate of positive decisions by the Parole Board.

Typically, a parole panel will receive professional reports from a prison psychologist, an offender manager (external probation officer) and an offender supervisor (probation officer seconded to the prison). Consistent with what had been found in the literature review, this study found that parole decisions were largely predictable from the report of the external probation officer, with some additional influence attributable to the other two professionals. A combined measure of these expert opinions was in turn largely predictable from the Clinical and Risk Management subscales of the HCR-20. This applied regardless of whether the index offence contained a sexual element or not. Since these subscales of the HCR-20 were those which had been identified by Rufino et al. (2011) as the most subjective, this was consistent with the experimental hypothesis. However, since the professional report writers (unlike the Parole Board) had not seen these assessments before writing their own reports, the correlation observed could not represent a causal connection. Instead, it appeared that both the professional reports and the Clinical and Risk Management subscales reflected the prisoner's presentation in interview with those report writers. This again is consistent with the literature which was reviewed, and the general hypothesis examined in the study.

With these findings in mind, Chapter 4 examined the Psychopathy Checklist-Revised (PCL-R). Despite being primarily a measure of personality (Hare, 2003b), it has become widely used in risk assessment. It was one of the instruments examined by Rufino et al.

(2011), who concluded that PCL-R Facets 3 and 4, and therefore Factor 2, were among the less subjective measures available. Consistent with the general view taken in this thesis, these were the PCL-R measures least predictive of parole decisions. Consideration of the published literature found that they were also the PCL-R measures most predictive of reconviction. However, despite the PCL-R's prominent position in risk assessment, it was concluded that its reliability was insufficiently established, and that its commercial success had tended to stifle the development of alternative approaches. In addition, it was clear that when used "in the field", rather than in a tightknit academic research group, interrater reliabilities were poor to non-existent for Factor 1. When combined with the finding that this Factor is the one most predictive of parole decisions, this is very concerning.

Conclusions

These findings suggest that modern parole decision-making in England and Wales is not based on factors which are objectively related to the risk of reconviction. Consistent with literature ranging across different jurisdictions and different time spans, parole decisions appear to be related to subjective assessments of risk. These are known to be unreliable, performing essentially at the chance level. Nonetheless, the Parole Board is under great pressure to keep down the numbers of high profile parole failures, especially those involving serious offences. In the absence of effective risk assessment, the only strategy for doing this is to limit the releases of all prisoners. Inevitably, but in keeping with earlier research, this must result in many low-risk prisoners being kept in custody when they could be safely released.

REFERENCES

Alterman, A. I., Cacciola, J. S., & Rutherford, M. J. (1993). Reliability of the revised Psychopathy Checklist in substance abuse patients. *Psychological Assessment, 5*(4), 442-448.

American Psychiatric Association. (1995). *Diagnostic and statistical manual of mental disorders* (fourth ed.). Washington, DC: American Psychiatric Association.

Annison, J. (2013). Change and the Probation Service in England and Wales: A gendered lens. *European Journal of Probation, 5*, 44-64.

Austin, J., Coleman, C., Peyton, J., & Johnson, K. D. (2003). Reliability and validity study of the LSI-R risk assessment instrument, final report. Retrieved from http://pacrimestats.info/PCCDReports/EvaluationResearch/Completed%20Research/Corrections%20and%20Alternative%20Sanctions/Assessment%20Instruments/LSI_R_Final_Report.pdf

Babiak, P., & Hare, R. D. (2007). *Snakes in suits: when psychopaths go to work*. London: HarperCollins.

Bagley, A. D., Abramowitz, C. S., & Kosson, D. S. (2009). Vocal affect recognition and psychopathy: Converging findings across traditional and cluster analytic approaches to assessing the construct. *Journal of Abnormal Psychology, 118*(2), 388-398. doi: http://dx.doi.org/10.1037/a0015372

Barbaree, H. E., Blanchard, R., & Langton, C. M. (2003). The development of sexual aggression through the lifespan: The effect of age on sexual arousal and recidivism among sex offenders. *Annals of the New York Academy of Sciences, 989*, 59-71.

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*(4), 422-438.

Beech, A. R., Fisher, D. D., & Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology - Research & Practice, 34*(4), 339-352.

Bishopp, D., & Hare, R. D. (2008). A multidimensional scaling analysis of the Hare PCL-R: Unfolding the structure of psychopathy. *Psychology Crime & Law, 14*(2), 117-132.

Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on psychopathy checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law & Human Behavior, 32*(4), 352-362.

Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). Do some evaluators report consistently higher or lower PCL-R scores than others?: Findings from a statewide sample of sexually violent predator evaluations. *Psychology, Public Policy, & Law, 14*(4), 262-283.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: professional guidelines for assessing risk of sexual violence*. Vancouver, BC: The British Columbia Institute Against Family Violence.

Bolt, D. M., Hare, R. D., Vitale, J. E., & Newman, J. P. (2004). A multigroup item response theory and analysis of the Psychopathy Checklist-Revised. *Psychological Assessment, 16*(2), 155-168.

Bonham, G., Janeksela, G., & Bardo, J. (1986). Predicting parole decision in Kansas via discriminant analysis. *Journal of Criminal Justice, 14*(2), 123-133.

Bradford, S., & Cowell, P. (2012). *The decision making process at parole interviews (indeterminate imprisonment for public protection sentences). Research summary 1/12.* Retrieved from doi:http://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/decision-making-process-parole-reviews-ipp.pdf

Buss, A. H. (1966). *Psychopathology*. New York: John Wiley & Sons, Inc.

Campbell, T. (2011). Predictive accuracy of Static-99R and Static-2002R. *Open Access Journal of Forensic Psychology, 3*, 82-106.

Campbell, T. W., & DeClue, G. (2010). Flying blind with naked factors: Problems and pitfalls in adjusted-actuarial sex-offender risk assessment. *Open Access Journal of Forensic Psychology, 2*, 75-101.

Caplan, J. M. (2007). What factors affect parole: A review of empirical research. *Federal Probation: A Journal of Correctional Philosophy and Practice, 71*(1), 16-19.

Cleckley, H. (1976). *The mask of sanity* (5th ed.). St Louis, MO: Mosby.

Coid, J., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., . . . Farrington, D. (2007). Predicting and understanding risk of reoffending: The prisoner cohort study. Research summary no. 6. doi: http://www.personalitydisorder.org.uk/news/wp-content/uploads/MofJResearchSummary6.pdf

Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *Journal of Forensic Psychiatry & Psychology, 22*(1), 3-21. doi: http://dx.doi.org/10.1080/14789949.2010.495990

Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., . . . Rogers, R. D. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting & Clinical Psychology, 77*(2), 337-348.

Comer, R. (1998). *Abnormal Psychology* (third ed.). New York: W. H. Freeman and Company.

Cooke, D. J., & Michie, C. (1999). Psychopathy across cultures: North America and Scotland compared. *Journal of Abnormal Psychology, 108*(1), 56-68.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*(2), 171-188.

Copas, J., & Jackson, D. (2004). A bound for publication bias based on the fraction of unpublished studies. *Biometrics, 60*(1), 146-153.

Craft, M. J. (1965). *Ten studies into psychopathic personality*. Bristol, England: John Wright.

Craig, L. A., & Beech, A. (2009). Best practice in conducting actuarial risk assessments with adult sexual offenders. *Journal of Sexual Aggression, 15*(2), 193-211.

Crew, B. K. (1991). Sex differences in criminal sentencing: Chivalry or patriarchy? *Justice Quarterly, 8*, 59-83. doi: 10.1080/07418829100090911

Crighton, D. A., & Towl, G. J. (2008). *Psychology in prisons* (2nd ed.). Leicester, England: Blackwell Publishing; British Psychological Society; England.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011a). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences USA, 108*(17), 6889-6892. Retrieved from doi:http://www.pnas.org/cgi/doi/10.1073/pnas.1110910108

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011b). Letter: Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist. *Proceedings of the National Academy of Sciences USA, 108*(42), E834. Retrieved from doi:http://www.pnas.org/cgi/doi/10.1073/pnas.1112190108

DeClue, G. (2013). Years of predicting dangerously. *Open Access Journal of Forensic Psychology, 5*, 16-28. Retrieved from doi:http://forensicpsychologyunbound.ws/OAJFP/Sex_Offenders_files/DeClue%202013.pdf

DeMatteo, D., Heilbrun, K., & Marczyk, M. (2006). An empirical investigation of psychopathy in a noninstitutionalized and noncriminal sample. *Behavioral Sciences & the Law, 24*, 133-146.

Dernevik, M. (2004). *Assessment and management of risk of violent recidivism in mentally disordered offenders.* Unpublished Ph.D. thesis, Karolinska Institutet, Stockholm. Retrieved from http://publications.ki.se/xmlui/bitstream/handle/10616/39264/thesis.pdf?sequence=1

Dietrich, A. (1994). *Assessing Risk of Violence in Mentally Disordered Offenders with the HCR-20.* Unpublished MA thesis, Simon Fraser University.

Edens, J. F., Boccaccini, M. T., & Johnson, D. W. (2010). Interrater reliability of the PCL-R total and factor scores among psychopathic sex offenders: Are personality features more

prone to disagreement than behavioral features? *Behavioral Sciences and the Law, 28*, 106-119.

Edens, J. F., & Campbell, J. S. (2007). Identifying youths at risk for institutional misconduct: A meta-analytic investigation of the psychopathy checklist measures. *Psychological Services, 4*(1), 13-27. doi: http://dx.doi.org/10.1037/1541-1559.4.1.13

Edens, J. F., Hart, S. D., Johnson, D. W., Johnson, J. K., & Olver, M. E. (2000). Use of the Personality Assessment Inventory to assess psychopathy in offender populations. *Psychological Assessment, 12*(2), 132-139.

Falshaw, L., Friendship, C., Travers, R., & Nugent, F. (2003). *Searching for "What Works": an evaluation of cognitive skills programmes. Home Office Research Findings 206.* Retrieved from doi:http://webarchive.nationalarchives.gov.uk/20110218135832/rds.homeoffice.gov.uk/rds/pdfs2/r206.pdf

Farrington, D. P., Gottfredson, D., Sherman, L. W., & Welsh, B. C. (2006). The Maryland scientific methods scale. In D. P. Farrington, D. L. MacKenzie, L. W. Sherman & B. C. Welsh (Eds.), *Evidence-based crime prevention* (2nd ed., pp. 13-21). London: Routledge.

Feder, L., & Dugan, L. (2002). A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward experiment. *Justice Quarterly, 19*(2), 343-375.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE.

Freeman, J., Palk, G., & Davey, J. (2010). Sex offenders in denial: A study into a group of forensic psychologists' attitudes regarding the corresponding impact upon risk assessment calculations and parole eligibility. *Journal of Forensic Psychiatry and Psychology, 21*(1), 39-51.

Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review, 11*, 303-327.

Garber, R. M., & Maslach, C. (1977). The parole hearing: Decision or justification? *Law & Human Behavior, 1*(3), 261-281.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29*(4), 397-426.

Goldacre, B. (2008). *Bad science*. London: Fourth Estate.

Gottfredson, D., & Ballard, K. (1966). Differences in parole decisions associated with decision-makers. *Journal of Research in Crime and Delinquency, 3*(2), 112-119.

Grann, M., Danesh, J., & Fazel, S. (2008). The association between psychiatric diagnosis and violent re-offending in adult offenders in the community. *BMC Psychiatry, 8*(1), 92.

Grann, M., Långström, N., Tenström, A., & Stålenheim, G. E. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment, 70*(3), 416-426.

Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL:SV, and OGRS. *Journal of Consulting & Clinical Psychology, 72*(3), 523-530.

Gray, N. S., Taylor, J., & Snowden, R. J. (2008). Predicting violent reconvictions using the HCR-20. *British Journal of Psychiatry, 192*(5), 384-387.

Greenland, S. (1994). Quality scores are useless and potentially misleading. Reply to "re: a critical look at some popular analytic methods". *American Journal of Epidemiology, 140*(3), 300-301.

Greenland, S., & O'Rourke, K. (2001). On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics, 2*(4), 463-471.

Griffin, D., & O'Donnell, I. (2012). The life sentence and parole. *British Journal of Criminology, 52*(3), 611-629.

Grubin, D. (2011). A large-scale evaluation of Risk Matrix 2000 in Scotland. *Sexual Abuse: A Journal of Research and Treatment, 23*(4), 419-433.

Hanson, R. K. (2001). Age and sexual recidivism: A comparison of rapists and child molesters. *Public Safety Canada*. Retrieved from doi:http://www.publicsafety.gc.ca/res/cor/rep/asrcrcm-arscvae-eng.aspx

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). A meta-analysis of the effectiveness of treatment for sexual offenders: Risk, need, and responsivity. *Public Safety Canada*. Retrieved from doi:http://www.publicsafety.gc.ca/res/cor/rep/_fl/2009-01-trt-so-eng.pdf

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348-362.

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154-1163.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*(1), 1-21.

Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. *Annals of the New York Academy of Sciences, 989*, 154-166.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1*(2), 111-119.

Hare, R. D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology, 3*(1), 99-119.

Hare, R. D. (1999). *Without conscience: The disturbing world of the psychopaths among us*. New York, NY, US: Guilford Press.

Hare, R. D. (2003a). *Hare PCL-R interview guide* (2nd ed.). Toronto, ON: Multi-Health Systems.

Hare, R. D. (2003b). *Hare PCL-R technical manual* (2nd ed.). Toronto, ON: Multi-Health Systems.

Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences & the Law, 18*(5), 623-645.

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The revised Psychopathy Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting & Clinical Psychology, 2*(3), 338-341.

Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology Special Issue on Diagnoses, Dimensions, and DSM-IV: The Science of Classification, 100*(3), 391-398.

Hare, R. D., & Neumann, C. S. (2010). The role of antisociality in the psychopathy construct: Comment on Skeem and Cooke (2010). *Psychological Assessment, 22*(2), 446-454.

Harkins, L., Beech, R., & Goodwill, A. (2010). Examining the influence of denial, motivation, and risk on sexual recidivism. *Sexual Abuse: A Journal of Research and Treatment, 22*(1), 78-94.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*(3), 413-425.

Hart, S. D. (2005). *The "state of the art" in risk assessment.* Paper presented at the International Conference on the Management and Treatment of Dangerous Offenders, York, England.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the "margins of error" of group v. individual predictions of violence. *British Journal of Psychiatry, 190*(suppl.49), 60-65.

Heinz, A. M., Heinz, J. P., Senderowitz, S. J., & Vance, M. A. (1976). Sentencing by parole board: An evaluation. *The Journal of Criminal Law and Criminology, 67*(1), 1-31.

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39*(9), 1148-1171.

Hildebrandt, M., de Ruiter, C., de Vogel, V., & van der Wolf, P. (2002). Reliability and factor structure of the Dutch language version of Hare's Psychopathy Checklist-Revised. *International Journal of Forensic Mental Health, 1*(2), 139-154.

Hill, A. (2010). Many offenders in jail too long, says parole chief, *The Guardian*. Retrieved from http://www.guardian.co.uk/society/2010/jul/19/offenders-jail-too-long-parole-chief

Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The Domestic Violence Risk Appraisal Guide. *Law and Human Behavior, 32*(2), 150.

Hobson, J., & Shine, J. (1998). Measurement of psychopathy in a UK prison population referred for long-term psychotherapy. *British Journal of Criminology, 38*(3), 504-515.

Hoffman, P. B. (1994). Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's salient factor score. *Journal of Criminal Justice, 22*(6), 477-494.

Holland, T. R., Holt, N., & Brewer, D. L. (1978). Social roles and information utilization in parole decision-making. *Journal of Social Psychology, 106*(1), 111-120.

Hood, R., & Shute, S. (1999). Parole decision making: Weighing the risk to the public. Home Office Research Findings 114. Retrieved from

doi:http://webarchive.nationalarchives.gov.uk/20110218135832/http://rds.homeoffice.gov.uk/rds/pdfs/r114.pdf

Hood, R., & Shute, S. (2000). The parole system at work: A study of risk-based decision-making. Home Office Research Study 202. Retrieved from doi:http://rds.homeoffice.gov.uk/rds/pdfs/hors202.pdf

Hood, R., Shute, S., Feilzer, M., & Wilcox, A. (2002). Sex offenders emerging from long term imprisonment. *British Journal of Criminology, 42*(2), 371-394.

Howard, P., & Dixon, L. (2012). The construction and validation of the OASys violence predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior, 39*(3), 287-307.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Huebner, B. M., & Bynum, T. S. (2006). An analysis of parole decision making using a sample of sex offenders: A focal concerns perspective. *Criminology: An Interdisciplinary Journal, 44*(4), 961-991.

Huebner, B. M., & Bynum, T. S. (2008). The role of race and ethnicity in parole decisions. *Criminology, 46*(4), 907-938.

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*(11), 1054-1060.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.

Khan, K. S., Daya, S., & Jadad, A. R. (1996). The importance of quality of primary studies in producing unbiased systematic reviews. *Archives of Internal Medicine, 156*(6), 661-666.

Kline, P. (2000). *A psychometrics primer*. London: Free Association Books.

Krauss, D. A. (2004). Adjusting risk of recidivism: Do judicial departures worsen or improve recidivism prediction under the federal sentencing guidelines? *Behavioral Sciences and the Law, 22*, 731-750.

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in

predicting institutional misconduct and new convictions. *Criminal Justice and*

Behavior, 28(4), 471-489.

Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction

of antisocial behaviour: The structure of criminal risk. *International Journal of Law and*

Psychiatry, 28, 360-374.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault*

Risk Assessment Guide (2nd ed.). Vancouver, BC: The British Columbia Institute

Against Family Violence.

Levav, J., Heitmann, H., Herrmann, A., & Iyengar, S. S. (2010). Order in product customization

decisions: Evidence from field experiments. *Journal of Political Economy, 118*, 274-

299.

Liem, M. (2013). Homicide offender recidivism: A review of the literature. *Aggression and*

Violent Behavior, 18(1), 19-25.

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic Personality Inventory-Revised (PPI-*

R) professional manual. Odessa, FL: Psychological Assessment Resources.

Lösel, F., & Schmucker, M. (2004). Psychopathy, risk taking, and attention: A differentiated

test of the somatic marker hypothesis. *Journal of Abnormal Psychology, 113*(4), 522-

529.

Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects

of a relapse prevention program on sexual recidivism: Final results from California's

sex offender treatment and evaluation project (SOTEP). *Sexual Abuse : A Journal of*

Research and Treatment, 17(1), 79-107.

Matejkowski, J., Caplan, J. M., & Cullen, S. W. (2010). The impact of severe mental illness on

parole decisions: Social integration within a prison setting. *Criminal Justice and*

Behavior, 37(9), 1005-1029.

Matejkowski, J., Draine, J., Solomon, P., & Salzer, M. S. (2011). Mental illness, criminal risk

factors and parole release decisions. *Behavioral Sciences and the Law, 29*, 528-553.

McCord, W., & McCord, J. (1964). *The psychopath: An essay on the criminal mind*. Princeton,

NJ: Van Nostrand.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of

the evidence*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. (1973). Why I do not attend case conferences. In P. Meehl (Ed.), *Psychodiagnosis:

Selected papers*. Minneapolis, MN: University of Minnesota Press.

Miller, A. K., Rufino, K. A., Boccaccini, M. T., Jackson, R. L., & Murrie, D. C. (2011).

Differences in person perception: Raters' personality traits relate to their Psychopathy

Checklist-Revised scoring tendencies. *Assessment, 18*(2), 253-260.

Ministry of Justice. (2012). Statistics on women and the criminal justice system.   Retrieved

from https://www.gov.uk/government/statistics/women-in-the-criminal-justice-system-

2009-10 doi:https://www.gov.uk/government/statistics/women-in-the-criminal-justice-

system-2009-10

Ministry of Justice. (2013a). The Parole Board, from http://www.justice.gov.uk/about/parole-

board

Ministry of Justice. (2013b). Story of the prison population 1993-2013 England and Wales.

doi:https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/1631

44/story-prison-population.pdf.pdf

Mokros, A., Stadtland, C., Osterheider, M., & Nedopil, N. (2010). Assessment of risk for

violent recidivism through multivariate Bayesian classification. *Psychology, Public

Policy, and Law, 16*, 418-450.

Morey, L. C. (1991). *Personality Assessment Inventory — Professional Manual*. Odessa,

Florida: Psychological Assessment Resources, Inc.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*.

Odessa, FL.: Psychological Assessment Resources, Inc.

Morgan, K., & Smith, B. (2005a). Parole release decisions revisited: An analysis of parole release decisions for violent inmates in a southeastern state. *Journal of Criminal Justice, 33*(3), 277-287.

Morgan, K., & Smith, B. (2005b). Victims, punishment, and parole: The effect of victim participation on parole hearings. *Criminology and Public Policy, 4*(2), 333-360.

Morgan, K., & Smith, B. (2008). The impact of race on parole decision-making. *Justice Quarterly, 25*(2), 411-435.

Morrissey, C., Mooney, P., Hogue, T. E., Lindsay, W. R., & Taylor, J. L. (2007). Predictive validity of the PCL-R for offenders with intellectual disability in a high security hospital: treatment progress. *Journal of Intellectual & Developmental Disability, 32*(2), 125-133.

Muraven M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*, 247-259.

Murrie, D. C., Boccaccini, M. T., Caperton, J., & Rufino, K. A. (2012). Field validity of the Psychopathy Checklist-Revised in sex offender risk assessment. *Psychological Assessment, 24*(2), 524-529.

Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2009). Does interrater (dis)agreement on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior, 32*(4), 352-362.

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, & Law* (preprint publication online). doi: http://dx.doi.org/10.1037/a0035824

Nugent, P., & Zamble, E. (2001). Affecting detention referrals through proper selection. *Forum on Corrections Research, 13*(1), 27-30.

Nuttall, C. P. (1977). *Parole in England and Wales. Home Office Research Study 38.* London: Home Office.

Padfield, N., & Liebling, A. (2000). An exploration of decision-making at discretionary lifer panels. Retrieved from doi:http://library.npia.police.uk/docs/hors/hors213.pdf

Parent, G., Guay, J.-P., & Knight, R. A. (2011). An assessment of long-term risk of recidivism by adult sex offenders: One size doesn't fit all. *Criminal Justice and Behaviour, 38*, 188-209.

Parole Board for England and Wales. (2012). Oral Hearings Guide doi:http://www.justice.gov.uk/downloads/offenders/parole-board/oral-hearing-guide.pdf

Parsonage, W. H., Bernat, F. P., & Helfgott, J. (1994). Victim impact testimony and Pennsylvania's parole decision-making process: A pilot study. *Criminal Justice Policy Review, 6*(3), 187-206.

Petersilia, J. (2003). *When prisoners come home: Parole and prisoner reentry*. Oxford: Oxford University Press.

Philipse, M. W., Koeter, M. W., van der Staak, C. P., & van den Brink, W. (2006). Static and dynamic patient characteristics as predictors of criminal recidivism: A prospective study in a Dutch forensic psychiatric sample. *Law and Human Behavior, 30*(3), 309-327. doi: http://dx.doi.org/10.1007/s10979-006-9013-4

Pogrebin, M. R., Poole, E. D., & Regoli, R. M. (1986). Parole decision making in Colorado. *Journal of Criminal Justice, 14*(2), 147-155.

Porter, S., Birt, A. R., & Boer, D. P. (2001). Investigation of the criminal and conditional release profiles of Canadian federal offenders as a function of psychopathy and age. *Law and Human Behavior, 25*(6), 647-661.

Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal & Criminological Psychology, 14*(1), 119-134.

Porter, S., ten Brinke, L., & Wilson, K. (2009). Crime profiles and conditional release performance of psychopathic and non-psychopathic sexual offenders. *Legal and Criminological Psychology, 14*(1), 109-118.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington DC: American Psychological Association.

Quinsey, V. L., & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence, 1*(2), 143-171.

Rice, M. E., & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. *Annals of the New York Academy of Sciences, 989*, 428-440.

Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: Guilford Press; US.

Ronson, J. (2011). *The psychopath test*. London: Picador.

Rufino, K. A., Boccaccini, M. T., & Guy, L. S. (2011). Scoring subjectivity and item performance on measures used to assess violence risk: The PCL-R and HCR-20 as exemplars. *Assessment, 18*(4), 453-463.

Rufino, K. A., Boccaccini, M. T., Hawes, S. W., & Murrie, D. C. (2012). When experts disagreed, who was correct? A comparison of PCL-R scores from independent raters and opposing forensic experts. *Law and Human Behavior, 36*(6), 527-537.

Rutherford, M., Cacciola, J. S., Alterman, A. I., Mackay, J. R., & Cook, T. G. (1999). The 2-year test-retest reliability of the Psychopathy Checklist-Revised in methadone patients. *Assessment, 6*(3), 285-291.

Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41*(3), 301-339.

Schroeder, M. L., Schroeder, K. G., & Hare, R. D. (1983). Generalizability of a checklist for assessment of psychopathy. *Journal of Consulting and Clinical Psychology, 51*(4), 511-516.

Scott, J. E. (1974). The use of discretion in determining the severity of punishment for incarcerated offenders. *The Journal of Criminal Law and Criminology, 65*(2), 214-224.

Seager, J. A., Jellicoe, D., & Dhaliwal, G. K. (2004). Refusers, dropouts, and completers: Measuring sex offender treatment efficacy. *International Journal of Offender Therapy and Comparative Criminology, 48*(5), 600-612.

Shine, J., & Hobson, J. (1997). Construct validity of the Hare Psychopathy Checklist, Revised, on a UK prison population. *Journal of Forensic Psychiatry, 8*(3), 546-561. doi: http://dx.doi.org/10.1080/09585189708412275

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A meta-review. *Criminal Justice and Behavior, 37*, 965-988.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*(3), 499-513.

Skeem, J. L., & Cooke, D. J. (2010a). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment, 22*(2), 433-445.

Skeem, J. L., & Cooke, D. J. (2010b). One measure does not a construct make: Directions toward reinvigorating psychopathy research—reply to Hare and Neumann (2010). *Psychological Assessment, 22*(2), 455-459. doi: 10.1037/a0014862

Skeem, J. L., Polaschek, D. L. L., Patrick, C. J., & Lilienfeld, S. O. (2011). Psychopathic personality: Bridging the gap between scientific evidence and public policy. *Psychological Science in the Public Interest, 12*(3), 95-162. doi: 10.1177/1529100611426706

Smith, B., Watkins, E., & Morgan, K. (1997). The effect of victim participation on parole decisions: Results from a south-eastern state. *Criminal Justice Policy Review, 8*(1), 57-74.

Stone-Meierhoefer, B., & Hoffman, P. B. (1982). The effects of presumptive parole dates on institutional behavior: A preliminary assessment. *Journal of Criminal Justice, 10*(4), 283-297.

Thornton, D., Mann, R. E., Webster, S. D., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of the New York Academy of Sciences, 989*(1), 225-235.

Townsend, M. (2012). Anders Behring Breivik's 21-year jail term closes Norway's darkest chapter, *The Guardian*. Retrieved from http://www.theguardian.com/world/2012/aug/24/anders-behring-breivik-verdict-norway-utoya

Turpin-Petrosino, C. (1999). Are limiting enactments effective? An experimental test of decision making in a presumptive parole state. *Journal of Criminal Justice, 27*(4), 321-332.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

University of York Centre for Reviews and Dissemination. (2009). Systematic reviews: CRD's guidance for undertaking reviews in health care. Retrieved from http://www.york.ac.edu/inst/crd/pdf/Systematic_Reviews.pdf

Urbaniok, F., Endrass, J., Rossegger, A., Noll, T., Gallo, W., & Angst, J. (2007). The prediction of criminal recidivism: The implication of sampling in prognostic models. *European Archives of Psychiatry and Clinical Neuroscience, 257*, 129-134.

Verona, E., Patrick, C. J., Curtin, J. J., Bradley, M. M., & Lang, P. J. (2004). Psychopathy and physiological response to emotionally evocative sounds. *Journal of Abnormal Psychology, 113*(1), 99-108. doi: http://dx.doi.org/10.1037/0021-843X.113.1.99

Vien, A., & Beech, A. R. (2006). Psychopathy: Theory, measurement, and treatment. *Trauma, Violence, & Abuse, 7*(3), 155-174.

Vitale, J. E., Smith, S. S., Brinkley, C. A., & Newman, J. P. (2002). The reliability and validity of the Psychopathy Checklist-Revised in a sample of female offenders. *Criminal Justice and Behaviour, 29*, 202-231.

Vito, G. F., Higgins, G. E., & Tewksbury, R. (2012). Characteristics of parole violators in Kentucky. *Federal Probation, 76*(1), 19-23.

Wakeling, H. C., Freemantle, N., Beech, A. R., & Elliott, I. A. (2011). Identifying predictors of recidivism in a large sample of United Kingdom sexual offenders: A prognostic model. *Psychological Services, 8*(4), 307-318.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: assessing risk for violence, version 2*. British Columbia: Simon Fraser University.

Weinshall-Margel, K., & Shapard, J. (2011). Letter: Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences USA, 108*, E833. Retrieved from doi:www.pnas.org/cgi/doi/10.1073/pnas.1110910108

Welsh, A., & Ogloff, J. (2000). Full parole and the aboriginal experience: Accounting for the racial discrepancies in release rates. *Canadian Journal of Criminology, 42*(4), 469-491.

Winfree, L. T., Sellers, C. S., Ballard, V. S., & Roberg, R. R. (1990). Responding to a legislated change in correctional practices: A quasi-experimental study of revocation hearings and parole board actions. *Journal of Criminal Justice, 18*(3), 195-215.

Wollert, R. (2006). Low base rates limit certainty when current actuarials are used to identify sexually violent predators: An application of Bayes' theorem. *Psychology, Public Policy, and Law, 12*(1), 56-85.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740-767.

Zinger, I. (2012). Conditional release and human rights in Canada: A commentary. *Canadian Journal of Criminology and Criminal Justice, 54*(1), 117-135.

Zinger, I., & Forth, A. E. (1998). Psychopathy and Canadian criminal proceedings: The potential for human rights abuses. *Canadian Journal of Criminology, 40*(3), 237-276.

Appendix 1a: Printout demonstrating search results (FindIt@Bham)

Appendix 1b: Printout demonstrating search results (APA PsycNet® Gold)

Appendix 1c: Printout demonstrating search results: BPS EBSCO search

Appendix 2: Studies included in the literature review, with descriptions and comments

Study, sample, and type	Outcomes	Comments
Gottfredson & Ballard (1966). 2,053 Californian male parole applicants. Derived a statistical model to predict how long men would be expected to serve, and compared with actual time served.	Parole board members did not all see the same range of prisoners. After allowing for this, there were no differences between parole board members in the decisions made.	Intended to elucidate whether individual board members made different judgements on similar cases. The use of a single board member to take parole decisions is unusual and may make the results difficult to generalise. There was no analysis of subgroups (e.g., sex offenders, violent offenders).
Scott (1974). 325 randomly-selected male and all 34 female prisoners released in 1968 in "a Midwestern US state [not specified]". Correlational study.	Parole was more likely if the index offence was not serious, the applicant had few disciplinary reports in prison, better education, higher IQ, higher socio-economic status, or was female. No relationship with ethnicity when other variables were controlled. Author pointed out the lack of relationship between many of these factors and risk, and questioned the utility of either indefinite sentences or parole boards.	Study conducted from a sociological point of view, seeing parole as a social mechanism for regulating the severity of punishment (i.e., time actually served). The parole board heard 100-150 cases a day and the median time spent on each was eight seconds, an unusually heavy case load which may make the results difficult to generalise.

Study, sample, and type	Outcomes	Comments
Heinz, Heinz, Senderowitz & Vance (1976). 294 randomly-selected parole applicants: 95% male; recruiting 38 additional females did not affect the results, but the authors admit that the female sample may not be representative. Correlational study.	Parole was more likely if the index and previous offences were less serious, if applicants were in the oldest or youngest age groups, if they had a good educational and employment record, if they were married and had dependents, if they had fewer institutional disciplinary reports, and if the "correctional sociologist" (seconded probation officer) recommended it (correlation of .42 with parole decision). Authors questioned whether these officers were sufficiently senior and well-trained. The authors also advocated actuarial rather than clinical assessment or clinical "adjustment" of actuarial risk.	If the "correctional sociologist" was against parole, or doubtful, 39% of applications succeeded, but if in favour 96% succeeded. Although very suggestive, as with all correlational studies, the results do not prove a causal relationship between variables. The authors did not investigate what contributed to the recommendation itself, although they reported some significant correlations with criminal history and institutional disciplinary reports.
Nuttall (1977). 1,682 parole decisions taken during the first five months of 1972 in England and Wales. Study using Automatic Interaction Detection to mimic the steps and they decision-making process. Essentially a correlational study.	Having two or fewer previous convictions, having a home to go to, having a shorter current sentence, and having a limited history of juvenile delinquency were related to gaining parole. In 86% of cases the parole authorities (local review committees at that time) followed the assistant governor's recommendation (92% if it was negative). Author advocated educating prisoners more about the nature of parole in an attempt to discourage them from dropping out of the parole process.	A relatively new technique at the time, Automatic Interaction Detection is essentially a correlational technique, and thus does not demonstrate a causal relationship. It presents decisions as if they were taken as a series of sequential splits on the basis of one variable at a time, but may not be the case. The study did not examine some of the factors found in other research (e.g., probation reports), but noted that in a majority of cases no probation recommendation was made. The parole system in England and Wales has since changed radically.

Study, sample, and type	Outcomes	Comments
Holland, Holt & Brewer (1978). 421 male Californian parole applicants. Correlational study, including a multiple regression analysis.	Parole board members focused mainly on the seriousness of the index offence, but this did not relate to what they claimed to consider. Their decisions were not related to subsequent success on parole. Prison caseworkers put less weight on index offence, and made recommendations influenced partly by objective recidivism risk, but even more by institutional disciplinary reports. Authors suggested that the board was too influenced by prisoner attitudes (inevitable when discretion was so wide). They suggested constraining decisions with objective risk measures.	The authors interpreted their findings in terms of a sociological view that the use of correctional information would be a function of the "social role" of the decision-makers. This was rather vague, and interpretation correspondingly speculative. It was not clear how the sample was selected, which raises the question of how representative it was. They referred to variables as "determinants" of the parole decision, when in fact they were correlates. Even multiple regression is essentially a correlational technique.
Stone-Meierhoefer & Hoffman (1982). 224 Federal prisoners, divided into equal experimental and control groups. Group comparison experiment, one receiving a "presumptive" parole date and the other going through a traditional parole procedure.	The presumptive parole date was given to prisoners early in sentence, with the proviso that their institutional behaviour should remain good. Those given a presumptive parole date did not incur more disciplinary reports, but did enrol in fewer institutional programmes (mainly educational rather than offending behaviour programmes).	There was no indication of the gender of prisoners, who may not have constituted a representative sample, but they were allocated to groups randomly by the researchers. Neither staff nor participants were blind to group membership, which could therefore have affected their expectations. More to do with results of the parole decision than its determinants.

Study, sample, and type	Outcomes	Comments
Bonham, Janeksela & Bardo (1986). 532 Kansas parole applicants (both sexes) applying between March and September 1979. Discriminant analysis.	Parole applications were more successful if the parole plan was of good quality (poorly defined) if an objective recidivism risk measure was low, if there were few institutional disciplinary reports, if offending behaviour programmes had been completed, if the prisoner had no substance abuse or mental health problems, and was more likely the longer the time already served. Previous criminal history and index offence were not significant. Discriminant analysis created a model which correctly predicted 76% of parole board decisions.	The authors state that there were "few" women in the sample, but do not say how many, or whether any attempt was made to consider them separately. Since other studies have done so, and found considerable differences, this is a weakness. There was likewise no attempt to differentiate the group according to offence type, although other studies have shown differences in parole application success rates, especially for sex offenders.
Pogrebin, Poole & Regoli (1986). 292 randomly-selected Colorado parole applicants (both sexes). Correlational study, but included the development of a statistical predictor. Also included qualitative observations of the parole panels at work.	Parole applications were more successful if applicants had fewer convictions, or previous custodial sentences, were older, had few previous parole violations or institutional disciplinary reports, and had an employment offer. The authors developed a statistical predictor of success on parole, but this was rejected by the parole board. Observations of the parole panels suggested that results could be influenced by the composition of the panel.	The authors do not say how many women were in the sample, and do not differentiate by offence type. They stated that judgements about sex offenders were disproportionately influenced (negatively) by the lone female board member, but did not present any quantitative data on this.

Study, sample, and type	Outcomes	Comments
Winfree, Sellers, Ballard & Roberg (1990). A randomly-selected sample of 114 parole revocation hearings in the year before a change in Texas law, and 94 from the year after. Discriminant analysis.	The legal change was meant to ensure that fewer parole violators were returned to prison, but there was little change. Few criminological factors were related to the parole decision, nor were ethnicity or age. A major correlate both before and after the legal change was the recommendation from the Department of Justice. After the change the DoJ recommended more cases be released, but the parole board took less notice, the end result being about the same.	An unusual study, in that it was set up specifically to examine the effects of a legislative change. However, it did not distinguish between male and female applicants, or people with different kinds of offence. Discriminant analysis is not necessarily any better than correlation for identifying causal relationships.
Hoffman (1994). Three random samples of 3,955, 2,339 and 1,092, released several years apart. Group comparison study.	Found that the predictive power of the Salient Factor Score static risk predictor held up well after having been in use for 20 years. However, some items were removed as a result of court decisions relating to perceived fairness rather than risk.	Predictor validation requires a large sample, which this study had. However, there was no distinction between male and female prisoners. Given the large numbers, a more sophisticated statistical treatment might have been possible, rather than simply comparing groups and demonstrating a correlation between a risk "bin" and subsequent recidivism. Nonetheless the study demonstrates that a static risk assessment scheme can remain stable over many years.
Parsonage, Bernat & Helfgott (1994). 200 Pennsylvania parole applicants. Discriminant analysis.	Victim testimony reduced the likelihood of parole being granted, especially if this opposed release, and was the greatest correlate of parole decision-making. Others were institutional disciplinary reports, victim injury, and previous convictions.	Did not consider whether there was any difference in outcome according to the quantity or type of testimony. Did not consider the possibility that victim input might be correlated with the severity of the offence (e.g., less traumatised victims might be more willing to attend hearings and face the perpetrator, but others might prefer to put views in writing).

Study, sample, and type	Outcomes	Comments
Smith, Watkins & Morgan (1997). 763 violent offenders in Alabama. Multiple regression analysis.	Victim presence (or representation) at the hearing decreased the likelihood of parole, and offender presence (or representation) increased it.	Did not examine non-violent offenders or sex offenders specifically. Did not study a screening process for parole applicants, which precedes the actual hearing. Many applicants are excluded at this point (but see Morgan & Smith, 2005a, 2005b).
Hood & Shute (1999, 2000). Interviews with parole board members, probation officers, 103 prisoners recently refused parole, and 340 other prisoners. Logistic regression analysis.	Recent changes in the parole system of England and Wales had reduced the numbers paroled. A single Parole Board member who did a preliminary interview made a recommendation which was followed in over 80% of cases. The Board greatly overestimated risk. Prisoners were more likely to be paroled if they had completed all the recommended offending behaviour courses. Regression analysis predicted parole board decision 87% of the time, using probation officer's recommendation.	Did not differentiate prisoners by gender, but did by offence type, demonstrating the very low rates of parole success for sex offenders. Although regression does not prove causality, there was a strong suggestion (similar to some other studies) that the Parole Board followed a firm recommendation from probation officers, though it was not clear how objective these recommendations were in turn.
Turpin-Petrosino (1999). 10 experienced New Jersey parole hearing officers (8 male); case simulations. Logistic regression analysis.	More applicants were refused if there were aggravating circumstances in the index offence. Parole officers rated some supposed risk factors as extremely important, but only when they wanted to deny parole. Sexual and violent offenders were mostly refused parole.	Genuinely experimental study, rare in this field, although simulations may be criticised precisely because they are not using real-life cases. Did not differentiate parole applicants by gender, but did by offence type.

Study, sample, and type	Outcomes	Comments
Padfield & Liebling (2000). Observational study of 52 parole cases in England and Wales; discretionary life sentences only. No quantitative analysis.	Little systematic risk assessment was used, and none of that was objective. Some known and validated risk factors were not discussed, and others only haphazardly. Decision-making may have been based on beliefs about risk which were not sound.	No quantitative analysis presented, although data would have been suitable. Therefore, some findings were rather vague and not clearly substantiated (e.g., "personal characteristics or cultural factors seem to influence decisions": it is simply not clear what this means).
Welsh & Ogloff (2000). 2,479 male Canadian parole applicants. Logistic regression analysis.	Minority ethnic groups (aboriginals) were not discriminated against by parole system: apparent discrimination was accounted for by differences in criminal history. Parole was less likely to be granted if the current sentence was long, if it involved sex, robbery, violence, drugs, if the institutional disciplinary history included assaults, and if the prisoner was assessed as having emotional and personal needs. Surprisingly, sex offenders were less likely to be paroled if they had completed the sex offender treatment programme.	Large sample (every Canadian male federal offender reaching parole eligibility in 1996). Did not consider females, but representative of the Canadian male prison population. Unlike most studies, broke institutional disciplinary offences down by type, showing only violent incidents appear to affect parole.
Hood, Shute, Feilzer & Wilcox (2002). 162 male sex offenders in England and Wales. Descriptive statistics given, but no inferential statistics.	The Parole Board greatly overestimated the risk of recidivism, especially for incest offenders and those who denied the offence. Authors suggested making actuarial information available to the Parole Board might increase accuracy.	Unclear how the sample was collected, or how representative it was of the target group (sex offenders). However, the follow-up period was longer than usual (six years for 94 of the men).

Study, sample, and type	Outcomes	Comments
Morgan & Smith (2005a). 762 parole applicants with determinate sentences for violence convictions over a 12 month period in Alabama. Logistic regression analysis.	Various factors correlated with parole success, but in the final logistic regression analysis the only significant factors were total felonies in history, current sentence length, senior prison officer's recommendation and warden's recommendation.	Complete cohort, so likely to be representative of the target group. The regression analysis eliminated index offence seriousness, but retained current sentence length, which is likely to correlate with seriousness. This was not resolved.
Morgan & Smith (2005b). 762 parole applicants with determinate sentences for violence convictions over a 12 month period in Alabama. Logistic regression analysis.	Victim impact statements were an important obstacle to parole, and more powerful if made in person than in writing. Other things being equal, the victim's presence at the parole hearing made parole less likely, and the offender's presence made it more likely. The senior prison officer's recommendation was the most influential single factor.	This study examined all three stages of the Alabama parole system, including an initial application screening system, whereas other studies did not. The authors stressed that the results might not generalise to other jurisdictions with different systems.
Huebner & Bynum (2006). 511 sex offenders in an unnamed US state. Cox proportional hazard analysis.	As sentence progressed, parole became more likely if the index offence was minor, institutional conduct good, and "parole readiness" (according to an official checklist) was high. Parole was delayed if victims were younger or offenders older.	Only sex offenders in the sample, and other research suggests sex offenders are treated differently for parole. Unusual kind of analysis, normally used for studying survival times after release. The authors adapted it to identify factors which speeded or slowed progress towards release. They tried various different statistical models, but there was general agreement between them.

Study, sample, and type	Outcomes	Comments
Caplan (2007). Review of literature.	Concluded that much of the research was old and might not be relevant to contemporary practice. Also identified institutional behaviour, sentence length, criminal history, mental illness and victim input as being the most significant determinants of the parole decision.	No indication that the review was systematic, and the author did not state any criteria for including or excluding papers. Did not consider any papers published outside of North America, or any relevant simulation/experimental studies.
Huebner & Bynum (2008). 423 young adult male offenders in an unnamed US state. Cox proportional hazard analysis.	Parole was significantly delayed by non-white ethnicity, mental health problems, a serious interpersonal crime as the index offence, or institutional misconduct. Hispanic ethnicity speeded up parole, as did drug crime and good "parole readiness" (see also Huebner & Bynum, 2006).	Only young male offenders, with an average age of around 20. As with the same authors' 2006 paper, they tried several different statistical models which all agreed well. No evidence for validity of "parole readiness".
Morgan & Smith (2008). 762 parole applicants with determinate sentences for violence convictions over a 12 month period in Alabama. Logistic regression analysis.	Concluded that ethnicity (the particular focus of this paper) was not a determinant of parole decisions, either at a preliminary screening stage or at the full parole hearing for those who had passed the screening.	Apparently the fourth study of the same sample, which by this time was at least 11 years old. However, the fact that it consisted of all members of the target group for that year at least ensured it was representative of that group.

Study, sample, and type	Outcomes	Comments
Porter, ten Brinke & Wilson (2009). 310 Canadian adult male sex offenders. Group comparisons using MANOVA.	Found high PCL-R scores associated with more violent/nonsexual offences. Few child abusers were high scorers, but had more sex offences if they were. High scorers were approximately 2.5 times more likely to gain parole, regardless of offence types, but did worse after release.	Reasons for parole revocation not recorded, which might have clarified the influence of psychopathy. No female participants.
Matjekowski, Caplan & Cullen (2010). Random sample of 407 New Jersey parole applicants in 2007. Logistic regression analysis	Found that a diagnosis of serious mental illness did not result in a lower likelihood of gaining parole, once the effect of violent prison disciplinary reports (higher in diagnosed prisoners) was taken into account.	Did not differentiate on the basis of gender, although rates of mental illness were higher among female participants.
Matjekowski, Draine, Solomon & Salzer (2011). Similar sample to the authors' other paper in 2010. Logistic regression analysis.	Similar finding to the earlier paper: mental illness was not associated with the parole decision. However, the authors also examined the relationship between mental illness and recidivism risk, as assessed by the LSI-R, finding that there was little association between them.	No separate consideration of gender. This study examined less serious mental illness than Matjekowski et al. (2010).

Study, sample, and type	Outcomes	Comments
Danziger, Levav & Avnaim-Pesso (2011a). 1112 Israeli Jewish and Arab offenders (both sexes). Logistic regression analysis.	Found that favourable parole decisions were more likely early in the day, or following a food break, and every judge was affected by this bias. The only legally relevant variables influencing decisions were the number of prior offences and the availability of a rehabilitation programme during supervision.	One of the few studies which had a theoretical rationale for the biases found, in this case "mental depletion" resulting from rapidly taking many sequential decisions.
Weinshall-Margel & Shaphard (2011). Letter commenting on Danziger et al (2011).	Suggested supposed "mental depletion" found by Danziger et al (2011a) was an artefact of the way cases were listed administratively.	Little real data, more of a hypothetical objection prepared by people who work within the Israeli justice system.
Danziger, Levav & Avnaim-Pesso (2011b).	Reply to Weinshall-Margel and Shaphard (2011). Included a reanalysis of some of the authors' own data to examine the possibility that the artefact mentioned was present. Biases persisted.	Supplemented this analysis by interviewing prison staff, who confirmed that ordering effects of the kind suggested did not occur.

Study, sample, and type	Outcomes	Comments
Bradford & Cowell (2012). 255 prisoners serving "indeterminate sentences for public protection". Descriptive statistics of cases, and a qualitative analysis of interviews with Parole Board members.	Parole decisions coincided strongly with the recommendations of professional witnesses, especially the offender manager (external probation officer) although resources limited the meetings these could have with the prisoners concerned. No correlation between objective statistical risk scores and parole decisions. Parole was more likely if there were no drug or accommodation issues and if there was a "robust risk management plan" (not formally defined).	This paper presents preliminary results from research which will be fully reported at a later date. At this point, however, the statistical analysis of a fairly large sample seems unsophisticated. Although females were deliberately over-sampled, no separate results were presented for them.
Griffin & O'Donnell (2012). Irish life sentence prisoners. Discussion paper with some analysis of national statistics.	In the Republic of Ireland the Parole Board recommends parole decisions to the Ministry of Justice, which approves them in 87% of cases. Little or no information was available on what determines Parole Board decisions in the first place.	Minimal statistical analysis, and no study of determinants of parole decision.

Study, sample, and type	Outcomes	Comments
Zinger (2012). No sample: a commentary and review paper.	The Parole Board of Canada considers all cases on the basis of reports by the Correctional Service. The latter has no obligation to make a recommendation, but mostly does. The Parole Board agrees with that recommendation in 89.5% of cases. The author suggested that the Correctional Service probably should not make a recommendation, as this enables the Parole Board to shift responsibility to them. He also questioned the value of the Board if it functions mainly as a rubber stamp.	One of a number of papers suggesting that the parole process is so difficult that, in effect, it may not be possible.

Appendix 3: Consent form

I, [*participant's name*], have been asked if I am willing to take part in a research project to be conducted by Mr Robert A Forde, who is a Chartered Psychologist and Registered Forensic Practitioner. I understand that this will involve an interview, psychological testing, and ratings on other psychometric instruments.

I also understand that all of this information may be used for a research project which will examine the influence of prisoner characteristics upon parole decisions concerning life sentence prisoners. This information will be used anonymously and no details which might identify my individual case will be included in any report of the research. I also understand that my identity and the details of the assessment and any private information (for example, details of my life which I may choose to disclose during the interview) will be known only to Mr Forde, and that I will be identified in research data files only by a code number.

I also understand that if I agree to participate in the research today I may withdraw consent to this at any time up to 11th January 2013 through my legal representatives. If I do so all the information relating to my case will be withdrawn from the research project and destroyed by a licensed confidential data destruction service.

I understand that this project has been given ethical approval by the University of Birmingham. Information about me will be kept securely in accordance with the Rules of the British Psychological Society, and the Data Protection Act, and after use will be destroyed by a licensed confidential data destruction service.

The research project has been discussed with me by Mr Forde, and I have been given a copy of this form to keep for myself. I have been told that I may raise any questions

which I have at any time during our interview and testing, or later through my legal representatives.

I consent to take part in the research, subject to these conditions.

Signed …………………………………………………………

Date ………………………………………..

Appendix 4: 2x2 Contingency tables showing agreement (Cohen's κ) among professional recommendations and the Parole Board decision

A: Seconded and external probation officers: 68% agreement; κ=.35 (p<.05)

Seconded	External probation officer		
probation officer	No progress	Progress	Total
No Progress	18	8	26
Progress	6	12	18
Total	24	20	44

B: External probation officer and prison psychologist: 68% agreement; κ=.30 (p<.05)

External	Prison psychologist		
probation officer	No progress	Progress	Total
No Progress	20	2	22
Progress	10	6	16
Total	30	8	38

C: Seconded probation officer and prison psychologist: 72% agreement; κ=.35 (p<.05)

Seconded probation officer	Prison psychologist		
	No progress	Progress	Total
No Progress	22	2	24
Progress	9	6	15
Total	31	8	39

D: Parole Board and external probation officer: 68% agreement; κ=.44 (p<.01)

Parole Board decision	External probation officer		
	No progress	Progress	Total
No Progress	19	10	29
Progress	6	15	21
Total	25	25	50

E: Parole Board and seconded probation officer: 67% agreement; κ=.33 (p<.05)

Parole Board decision	Seconded probation officer		
	No progress	Progress	Total
No Progress	21	9	30
Progress	8	13	21
Total	29	22	51

F: Parole Board and prison psychologist: 66% agreement; κ=.25 (not

significant, p<.08)

Parole Board decision	Prison psychologist		
	No progress	Progress	Total
No Progress	22	3	25
Progress	12	7	19
Total	34	10	44

Appendix 5: Boxplots showing the distribution of Years Over in three different index offence groups