

An Examination of the Predictive Validity of the Risk Matrix 2000 in England and Wales

Sexual Abuse: A Journal of
Research and Treatment
22(4) 443–470
© The Author(s) 2010
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1079063210384274
<http://sajrt.sagepub.com>



Georgia D. Barnett¹, Helen C. Wakeling¹,
and Philip D. Howard¹

Abstract

This study examined the predictive validity of an actuarial risk-assessment tool with convicted sexual offenders in England and Wales. A modified version of the RM2000/s scale and the RM2000 v and c scales (Thornton et al., 2003) were examined for accuracy in predicting proven sexual violent, nonsexual violent, and combined sexual and/or nonsexual violent reoffending in a sample of sexual offenders who had either started a community sentence or been released from prison into the community by March 2007. Rates of proven reoffending were examined at 2 years for the majority of the sample ($n = 4,946$), and 4 years ($n = 578$) for those for whom these data were available. The predictive validity of the RM2000 scales was also explored for different subgroups of sexual offenders to assess the robustness of the tool. Both the modified RM2000/s and the complete v and c scales effectively classified offenders into distinct risk categories that differed significantly in rates of proven sexual and/or nonsexual violent reoffending. Survival analyses on the RM2000/s and v scales ($N = 9,284$) indicated that the higher risk groups offended more quickly and at a higher rate than lower risk groups. The relative predictive validity of the RM2000/s, v, and c, as calculated using Receiver Operating Characteristics (ROC) analyses, were moderate (.68) for RM2000/s and large for both the RM2000/c (.73) and RM2000/v (.80), at the 2-year follow-up. RM2000/s was moderately accurate in predicting relative risk of proven sexual reoffending for a variety of subgroups of sexual offenders.

Keywords

sexual offender, actuarial risk assessment, Risk Matrix 2000

¹National Offender Management Service, London

Corresponding Author:

Georgia D. Barnett, Rehabilitation Services Group, Room 130 Abell House, John Islip Street, London, SW1P 4LH

Email: georgia.barnett@noms.gsi.gov.uk

Introduction

The past 20 years has seen a surge of research interest in the development of methods to accurately assess risk of reoffending. A number of actuarial risk-assessment tools have emerged, which code factors that research has reliably linked to risk of reoffending to produce a score that indicates the probability of reconviction for a certain type of offence over a specified period of time (Beech, Fisher, & Thornton, 2003). Actuarial assessments of risk tend to be based on static, historical factors, and are, therefore, generally easier to use than structured assessments of risk based on clinical judgments. A number of studies have reported that actuarial assessments can both be reliable and have good predictive accuracy (e.g., Bonta, Law, & Hanson, 1996; Hanson & Morton-Bourgon, 2009; Hanson, Morton, & Harris, 2003; McNeil, Sandberg, & Binder, 1998). The most recent meta-analytic study of the accuracy of recidivism risk assessments for sexual offenders found that actuarial assessments outperformed unstructured clinical judgment and fared as well or better than structured clinical judgment (Hanson & Morton-Bourgon, 2009). For these reasons, the use of actuarial risk-assessment tools has become common practice among professionals in criminal justice systems who deal with serious offenders, such as those who have committed crimes of a sexual nature. Such assessments represent a cost-effective way of allocating limited resources to those offenders at higher risk of reoffending.

However, this form of risk assessment has its limitations; it fails to consider individually relevant risk factors and cannot be used to individualize treatment or risk-management plans. Although there remains some debate about the use of actuarial tools in individualized risk assessment (e.g., Cooke & Michie, 2009), there is overwhelming support for use of these as part of a wider risk assessment (e.g., Hanson & Morton-Bourgon, 2009; Harris, Rice, & Quinsey, 2008; Mossman & Sellke, 2007). As a result, a combination of actuarial and empirically guided clinical judgment assessments is commonly used to assess sexual offenders (e.g., Beech et al., 2003). This practice aims to provide a more comprehensive assessment of risk that incorporates both static and dynamic risk factors. The most widely used actuarial risk-assessment tool for sexual offenders in the English and Welsh Prison and Probation Services is the Risk Matrix 2000 (Thornton et al., 2003). This tool is used to make decisions about treatment pathways for sexual offenders as well as informing decisions about parole and the management of offenders serving sentences in the community.

The Risk Matrix 2000 (RM2000) is an empirically derived actuarial risk-assessment tool that uses historical information about offenders to divide them into categories that should differ substantially in their rates of reconviction for sexual or other violent offences. It was developed for use in the United Kingdom with males aged 18 and above who have been convicted of a sexual offence (at least one of which must have been committed after the age of 16). The RM2000 has three scales: the RM2000/s is a prediction scale for proven sexual offending, the RM2000/v is a prediction scale for proven nonsexual violent offending, and the RM2000/c is a combination of the first two scales and predicts proven sexual or nonsexual violent offending. The RM2000

was developed using a construction sample of 1,910 untreated convicted sexual offenders who had been discharged from prison in England and Wales and who had been followed for 2 years (Thornton et al., 2003). The RM2000/s was originally intended to be a revision of the SACJ-Min, another sexual offending risk-assessment tool (described in Hanson & Thornton, 2000). The RM2000/s was constructed using existing research knowledge to identify individual factors predictive of recidivism to incorporate in the tool and what weight to assign to each of these factors. Thornton and Travers (1991) conducted a 10-year follow-up of a sample of convicted sexual offenders, which suggested that sexual offenders equally presented a risk of nonsexual violent reoffending as they did a risk of sexual reoffending. This led to development of the RM2000/v and c scales.

All three scales have been cross-validated using further samples of sexual offenders (Thornton et al., 2003). The first of these samples consisted of sexual offenders who had engaged in treatment in custody and had been followed for at least 2 years after their release ($N = 647$). The second sample consisted of 429 convicted sexual offenders who were discharged from prisons in England and Wales in 1979, for whom 19-year follow-up data were available. The third sample consisted of 311 sexual offenders that were discharged from a prison in England and Wales in 1980 following a sentence of at least 4 years. Ten-year follow-up data were available for this sample. The area under the receiver operating characteristics (ROC) curve provides a measure of predictive accuracy and can range from .5, indicating that prediction is no better than chance, to 1, indicating perfect prediction. The scales yielded ROC Area Under the Curve (AUC) statistics between .74 and .85 (Thornton et al., 2003). However, these studies have been criticized (e.g., Grubin, 2008) for, among other things, having limited descriptions of the samples used, making it unclear how representative they were of the general sexual offender population.

Few independent cross validations of RM2000 have been conducted and they have often been small-scaled. For example, Craissati and Beech (2005) were unable to estimate the tool's predictive validity as their sample of offenders on probation in London ($N = 310$) included too few reoffenders to allow meaningful evaluation. Craig, Beech, and Browne (2006) reported AUCs between .59 and .87 for the RM2000 scales among offenders released from a regional secure unit in the United Kingdom ($N = 85$). Again, however, the small sample size limits the confidence we can place in these conclusions. Grubin (2008) examined the accuracy of both the s and v scales (AUC = .73 for the s scale and .76 for the v) using a large sample of offenders in Scotland ($N = 1,029$, mean follow-up 8.5 years, minimum follow-up 5 years). The long-term follow-up of Canadian offenders ($N = 351$) by Kingston, Yates, Firestone, Babchishin, and Bradford (2008) reported AUCs of between .65 and .70 for the three scales, depending on the recidivism outcome (sexual, nonsexual violent, or any recidivism). Finally, Hanson and Morton-Bourgon's (2009) meta-analysis found the RM2000/s and v scales to have good predictive accuracy for sexual reconviction ($d = 0.67$, which equates to an AUC of .68; Rice & Harris, 2005) and violent reconviction ($d = 0.62$, AUC = .67), respectively. Despite these encouraging findings a thorough cross-validation with a substantial English and

Welsh sample is still needed to establish whether the tool is valid and robust with this population.

It is imperative that the assessments we are using with sexual offenders are as accurate as possible with regard to assessing risk of recidivism so that actions can be taken and strategies put in place to manage those at a higher risk of reoffending. This will facilitate appropriate and effective allocation of resources and will allow us to use better and more informed assessments to guide both restrictive and constructive interventions, benefiting both the public and the offender. As such, this study represents an attempt to examine the predictive accuracy of the RM2000 with a large sample of English and Welsh sexual offenders.

Previous research into the predictive accuracy of actuarial tools has called into question the validity of applying the same tool to what is a very heterogeneous group of offenders. More specifically, researchers have questioned whether one tool can be accurate for both rapists and child molesters (e.g., Craissati, 2003). Kingston et al. (2008) reported lower rates of sexual reconviction for intrafamilial child molesters than the rapists in their sample, whereas Hanson, Helmus, and Thornton (2010), who examined the predictive accuracy of another actuarial assessment across eight varied samples, found minimal differences in rates between these groups after controlling for static, historical risk factors. Committing a noncontact offence is one of four "risk-raising" items in the RM2000; various researchers have suggested that noncontact offenders are qualitatively different to contact offenders (for a summary, see Murphy & Page, 2008). Some evidence suggests that the recidivism rates and psychological characteristics of those whose offences relate to the Internet may be different to those with more prototypical offences (e.g., Elliot & Beech, 2009; Laulik, Allam, & Sheridan, 2007; Seto & Eke, 2005; Webb, Craissati, & Keen, 2007). Given potential differences in reoffending patterns between rapists, child molesters, noncontact offenders, and Internet offenders, this study examines whether the predictive accuracy of an actuarial tool differs according to offence type.

In addition, a number of studies reported that rates of recidivism decrease with age, which has led to speculation that recidivism estimates should be adjusted for older offenders (e.g., Fazel, Sjostedt, Langstrom, & Grann, 2006; Hanson, 2008; Thornton, 2006). However, Doren (2006) questioned the generalizability of these findings and suggested that no firm conclusions could be drawn on the basis of current studies. Like most actuarial risk assessments, RM2000/s scores age as a risk factor, but the lowest-risk age band contains all offenders aged 35 or above. As some evidence suggests that offenders aged above 50 may differ from those in their 30s and 40s, we might expect the predictive validity of the RM2000/s within this age band to be reduced by its failure to reflect any resultant differences in recidivism risk. This study, therefore, also explores how well this actuarial tool predicts reconviction for offenders in different age bands, to add to current debate.

This study will aim to test the validity of the RM2000/s, v, and c scales in predicting proven sexual violent, nonsexual violent, and combined sexual and/or nonsexual violent recidivism, respectively, over 2- and 4-year follow-up periods. The focus is on

relative risk; the ability of the scales to distinguish groups of offenders that differ markedly in their rates of proven reoffending. We also aim to examine relative risk among subgroups of offenders including contact offenders with child victims and those with adult victims, noncontact offenders (excluding Internet offenders), Internet offenders, sexual offenders in different age bands, and those with a history of sexual offending but whose current offence is not sexual. Given that research suggests that such subgroups of sexual offenders may differ in their absolute risk (rates of recidivism), we will also examine rates of proven reoffending among these subgroups to establish whether such differences exist.

Method

Sample

The sample used for survival analysis consisted of 9,824 adult males convicted, either currently, or previously, of a sexual offence who were in the community, either after release from prison or serving a community sentence, by the end of March 2007 and for whom follow-up proven reoffending data were available. Of these offenders, 4,946 had been "at risk" for at least 2 years (had 2-year follow-up data). This smaller sample was used for the majority of the analysis. Offenders were identified from two main databases:

The Offender Assessment System (OASys; Home Office, 2002) database holds information on all of those offenders assessed in custody and community settings. OASys is a structured assessment of static and dynamic risk factors used to aid offender management. This database holds demographic and risk-related information but does not hold reconviction data.

The treatment database held by the National Offender Management Service (NOMS) Interventions and Substance Misuse Group (ISMG), referred to from hereon in as the ISMG database. This holds information on offenders who have undertaken a sex offender treatment programme in custody in England and Wales since 2001. The database also holds information regarding the static risk of all but a small minority of these offenders, for whom records were missing due to administrative errors.

The final sample used for survival analysis comprised 9,824 offenders of which 4,946 were able to be followed for at least 2 years. Of the 9,824, 1,488 were from the ISMG database, and 8,336 were from the OASys database. The overall sample did not include those sexual offenders who were in the community by the cutoff date of 31st March 2007 but had neither received treatment in prison nor been assessed using OASys. The sample used in this study may not, therefore, be entirely representative of the national convicted sexual offender population. However, omission from the OASys database would require the offender to have both received only a limited report

prior to sentence and been sentenced to a custodial sentence of less than 12 months or a community sentence involving neither supervision nor treatment. This is likely to be relatively rare for those with histories of sexual offending, though we do not have access to quantitative data that could determine exactly how many offenders this affects. This also means that the sample is not preselected on the basis of risk as nearly all sexual offenders should be assessed using OASys; lack of an OASys assessment before March 2007 would have been dependent on how quickly and comprehensively OASys was rolled out regionally and therefore would have been largely determined by geographical location rather than by characteristics of the offender. Similarly, those in the treatment sample in custody would not have been heavily preselected for treatment based on risk level only. In NOMS, sex offender treatment is allocated on the basis of a combination of static risk level, time left to serve/to order expiry and willingness to engage in treatment. As such, both the ISMG and the OASys samples are likely to be what Hanson et al. (2010) term *routine* correctional samples rather than preselected high-risk samples.

The 8,336 offenders in the OASys sample came from a larger OASys data set of 210,601 offenders, used previously for studies of general and violent reoffending. This larger sample already had follow-up proven reoffending (akin to reconviction) data as sought from the Home Office Police National Computer (HOPNC) database. Of this OASys sample, 12,420 had HOPNC records that included current or prior cautions (an alternative to prosecution issued for minor offences) and/or convictions for sexual offending, identifying them as sexual offenders. Of these 12,420, 488 offenders were duplicates of cases included in the ISMG sample (the treated sample data were preferred as it included RM2000/s item-level information); 111 had incomplete information to score RM2000 (see Procedure for details of missing data that led to exclusion); 2,250 were removed because the OASys assessment was not completed within 3 months of discharge from custody or commencement of community sentence; and 1,235 offenders were ineligible for RM2000 because their sexual offending history was confined to offences committed before the age of 16. The final OASys sample therefore comprised 8,336 offenders.

The 1,488 comprising the ISMG database was reached following a series of steps. First, a total of 2,351 offenders identified from the ISMG sample were submitted to the HOPNC to obtain follow-up proven reoffending data. This consisted of all offenders who had undertaken a sex offender treatment programme in custody and for whom static risk information was available. Reoffending data were received for 1,806 of these offenders, representing a 77% matching rate. The unsuccessful matches resulted from incorrect or (more frequently) missing demographic information on the offenders. Of these 1,806, 114 offenders were removed as information on their release date was not available; 27 were removed as their HOPNC data indicated no sexual conviction history; 165 had missing data required to calculate OGRS or RM2000; and for 12 offenders all their sexual offences were committed before the age of 16, rendering them unsuitable for assessment with the RM2000. Therefore, the final ISMG sample comprised 1,488 offenders.

Table 1. Demographic Characteristics of the Whole Sample and the Sample for With 2-Year Follow-Up Data

Variable	Whole sample		2-year follow-up sample	
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
Age	9,824	39.1 (13.1) years	4,946	40.0 (13.1) years
	<i>n</i>	%	<i>n</i>	%
Ethnicity				
White	8,114	82.6	4,184	84.6
Black	289	2.9	143	2.9
Asian	358	3.6	119	2.4
Mixed	81	0.8	32	0.6
Other	58	0.6	33	0.7
Missing	924	9.4	435	8.8
Index offence type				
Indecent assault	2,423	24.7	1,308	26.4
Indecent images of children	866	8.8	513	10.4
Rape/buggery	1,050	10.7	567	11.5
Unlawful sexual intercourse (USI)	460	4.7	190	3.8
Gross indecency	310	3.2	180	3.6
Indecent exposure	343	3.5	152	3.1
Incest	76	0.8	42	0.8
Other sexual offences	248	2.5	145	2.9
Violent offences (according to RM2000v definition)	1,370	13.9	548	11.1
Acquisitive offences	821	8.4	408	8.2
Compliance with requirements of sex offender orders	103	1.0	43	0.9
Other	1,754	17.9	850	17.2

Note: USI refers to nonforced sex with a child under 16. RM2000/v definition of a violent offence is any nonsexual offence whose legal charge implies threat or use of force against the person.

Table 1 depicts the demographic characteristics of the 9,824 offenders used for survival analysis and the 4,946 used for the main analyses. The age of the sample ranged from 18 years to 85 years, whereas determinate sentence lengths ranged from 1 month to 24 years. Data on the sample’s ethnicity were available for 91% of both samples. Of those for whom these data were available, the majority were White. Forty-seven percent of the larger sample had custodial sentences, with an average length of 3.8 years, whereas only 39% of the sample with at least a 2-year follow-up period and for whom we had this information (*n* = 3,434) were sentenced to a period in custody, the average length of which was 2.2 years. The community sentenced offenders in both samples had an average sentence length of 1.9 years. Sentence type information

is shown in Table 1; where convicted of multiple offences, offenders are classified on the basis of the principal (most serious) offence for which they received their sentence. Fifty-nine percent of the whole sample had a sexual conviction as their index offence, compared with 63% of the smaller sample. In both samples, for those whose index offence was sexual, the most common offence was Indecent Assault. The most common nonsexual index offences were indictable assaults and common assaults. However, clinical judgment of the motivation for these offences, as recorded as part of the OASys assessment, indicated that around 11% of all nonsexual index offences were considered to have a sexual motivation.

Measures

The RM2000 is a static risk-assessment tool for use with adult males who have ever been convicted of a sexual offence committed when the offender was age 16 or above. The RM2000/s predicts proven sexual recidivism and is made up of seven items divided into two scoring steps. Step 1 comprises three items: age of the offender on release, number of sentencing occasions for a sexual offence, and number of sentencing occasions for any criminal offence. The scores assigned to each of these items are summed and translated into one of four preliminary risk categories: low, medium, high, or very high. The second scoring step has a further four risk-raising items: Any or all of the victims of sexual offending have been male, any or all of the victims of sexual offending have been strangers, the offender has never had a stable live in relationship for more than 2 years (termed the *single* item), and any or all of their sexual offences have been noncontact (excluding Internet offences for those whose only sexual offences relate to the Internet). These items are scored on a dichotomous scale as either present or not. If two or three of these items are present, the initial risk category is raised one level (e.g., from low risk to medium). If all four of these aggravating factors are present, the initial risk category is raised by two risk levels (e.g., from low to high).

The RM2000/v, which predicts nonsexual violent reoffending, is composed of three items: age, number of sentencing occasions for a nonsexual violent offence, and whether the offender has ever been convicted of a burglary. The items have different scoring rules, as defined in the scoring manual (Thornton, 2007), and are summed and then translated into one of the four risk categories described above. The RM2000/s and v risk categories are assigned points (0 = *low*, 1 = *medium*, 2 = *high*, 3 = *very high*), which are summed to produce the RM2000/c scale score. This score is then translated into one of the four risk categories (Thornton, 2007). The RM2000/c risk classification predicts sexual *and* nonsexual violent reoffending.

Procedure

RM2000/s risk categories were available for those in the sample who engaged in treatment in prison (the ISMG sample). For these 1,488 offenders, the RM2000/s was coded by trained staff at treatment sites, prior to treatment. To deal with the inevitable

issue of missing data, offenders in the treated sample were excluded if they were missing scores for more than one aggravating factor (items in Step 2 of the RM2000/s scoring procedure), as this would affect their overall risk category. In addition, cases in the ISMG sample with one missing aggravating factor score were removed if that score was required to determine their overall risk category (i.e., when they were assessed as low, medium, or high risk at Step 1 and possessed one of the three non-missing aggravating factors, or were low or medium risk at Step 1 with all three of the nonmissing aggravating factors).

For the offenders in the OASys sample ($n = 8,336$), modified RM2000/s risk categories were computed retrospectively by the researchers using information gathered on the OASys assessments and HOPNC data. The *age* item was computed using the age of the offender at the time they had their OASys assessment, or at the time of their discharge from prison, as this marked the start of their time at risk in the community. The *sexual appearances* and *criminal appearances* items were scored using the criminal history information held on the PNC. Formal cautions were included in scoring these items, as suggested in the RM2000/s scoring guide. When combined, the scores on these three items produced an initial risk category. Whether the offender had ever been convicted of offending against a *male* was scored using HOPNC data, which use offence codes that, in the majority of cases, indicate the gender of the victim of the current and previous offences. The *noncontact* offence item was scored on the basis of OASys offence codes for current and previous sexual offences, which indicate whether the sexual offence involved physical contact with the victim(s). We were unable to score the *single* item and the *stranger* item for those in the OASys sample. In line with other similar studies (e.g., Langton, Barbaree, Hansen, Harkins, & Peacock, 2007), RM2000 risk level was still calculated for those in the sample for whom these were the only two missing items. As this was likely to result in consistent underscoring of the OASys sample, those who were scored as having one or both of the two aggravating factors that could be scored (the *male* or the *noncontact* item), were raised a risk category.

To estimate the magnitude of the effect this revised scoring procedure was likely to have on the sample's risk levels, we calculated the RM2000/s risk categories of those in the ISMG sample ($n = 1,488$) using the revised procedure and compared this to the categories this sample were assigned using the full version of the tool. Using the revised scoring, 79% of the treated sample remained in the same final risk category. Of the 21% that changed risk category as a result of the removal of the single and stranger items, the majority (16%) moved up one risk category, while the remaining 5% moved down one risk category (e.g., from medium to low risk). The revised scoring procedure is likely, therefore, to slightly overestimate the RM2000/s risk level of the sample.

The current RM2000 scoring manual indicates that for those offenders whose only sexual offences (current and previous) relate to Internet offending, the noncontact and stranger items should not be scored (Thornton, 2007). Adhering to these guidelines, we used this revised scoring procedure for all those in our sample whose only sexual

offences related to possessing or making indecent images of children, which are those offences relating to the Internet. As it was not possible to score the *single* or *male* item for the OASys sample (because the detailed relationship information required for scoring was not available on OASys and because the offence codes relating to indecent images do not indicate the gender of the victims in the images), all Internet-only offenders in the OASys sample remained at their Step 1 risk category.

RM2000/v risk categories were calculated for the entire sample using HOPNC data to determine age at release or start of community sentence, number of sentencing occasions for a violent offence and whether the offender had ever been convicted of burglary. If any of these data were missing for offenders in the sample, they were not included in the RM2000/v analyses. In accordance with the scoring rules for RM2000/v a violent offence was defined as any offence whose legal charge implied threat or use of nonsexual violence against a person.

To prevent any researcher bias in scoring risk assessments, scorers coded the variables blind to recidivism outcomes. Equally, recidivism outcomes were coded blind to risk-assessment scores. The measure of proven reoffending was restricted to any conviction or caution for a new offence. Those offenders who were convicted or cautioned of an offence committed within 4 years of their community punishment or release from prison were identified using information from the HOPNC. A buffer period of at least 6 months was satisfied for data from each sample. This buffer period meant that each offender's recidivism follow-up ended at least 6 months before the HOPNC data were extracted, allowing time for all cautions and convictions within the follow-up period to be entered onto the PNC. When an offender was recalled to prison but did not receive a new sentence, the follow-up period was terminated at the point of recall, as when recalled he is no longer at risk of reoffending for the purposes of this study. Recall is different to proven reoffending, in that no criminal offence need be committed to induce a recall to prison and, unless followed by conviction or caution, recall will not result in a new sentence.

Proven sexual reoffending (from hereon in termed *proven sexual reoffending*) was defined as any conviction or caution for a new offence whose statutory definition meets the definition of sexual offences in the Risk Matrix 2000 scoring guide (Thornton, 2007). This does not, therefore, provide a true reflection of the rates of sexual reoffending at any given time, as this method only counts those offences that were reported and then resulted in caution or conviction under the above statutes. It excludes all convictions or cautions for offences under nonsexual statutes, even where the offence had a sexual motivation or direct sexual content but was charged differently as a result of plea bargaining or prosecutorial discretion.

Proven offending of a nonsexual, but violent nature (referred to as *proven violent reoffending* from hereon in) was defined as any new conviction or caution for an offence whose statutory definition corresponds with the set of nonsexual violent offences listed in the Risk Matrix 2000 scoring guide (Thornton, 2007). It is likely that some of the offence in this category will either have been sexually motivated, or will represent a sexual offence, that, as a result of plea bargaining, has resulted in a conviction for a

violent offence. As with proven sexual reoffending, this measure of proven violent reoffending will be an underestimate of true levels of violent reoffending.

Analyses also used a combined sexual and violent proven reoffending outcome, which consisted of any new conviction or caution for a sexual and/or a violent reoffence, as defined above. This will be referred to as *combined reoffending* from hereon in.

We conducted analyses on the 2-year data, and 4-year data where possible. Some analyses were modelled on Grubin's (2008) validation of the RM2000/s and v with Scottish offenders.

Results

Differences between those offenders who had 2-year follow-up data ($n = 4,946$) and those who could not be included in the 2-year follow-up but were used in the survival analysis ($n = 4,878$) were tested. The only significant difference between the two was that a greater proportion of the 2-year sample (61% vs. 55%; $\chi^2 = 34.6, p < .001$) had an index offence that was sexual in nature.

Rates of Proven Reoffending Over Fixed Follow-Up Periods

Table 2 shows the rates of proven reoffending for 2- and 4-year follow-up periods for differing types of reoffending. The overall rate of proven sexual reoffending increased from 2.2% at Year 2 to 5.5% at Year 4, whereas the overall rate of proven violent reoffending rose more slowly, from 9.0% to 10.2%, and the overall rate of combined reoffending rose from 10.9% to 15.1%. Table 2 also shows the proven reoffending rates of those in different RM2000/s, v, and c risk categories for different types of offences and indicates that rates of proven reoffending increased by risk category for each type of offending in all of the follow-up periods. At the 4-year stage, just more than a quarter (27.3%) of very high-risk offenders on the RM2000/s scale had a proven sexual reoffence, whereas just more than a third (35.6%) of very high-risk offenders on the RM2000/v scale had a proven violent reoffence.

The 2-year proven reoffending rates of the present sample were compared with the 2-year reconviction rates reported in Thornton et al.'s (2003) and Grubin's (2008) studies (Table 3). Although there was not a significant difference between the proven sexual recidivism rates of the RM2000/s low-risk offenders in the current study and those in Grubin (2008), the 2-year rates in the Grubin study were significantly greater than those in the present study for the RM2000/s medium-, high-, and very high-risk categories and for all offenders ($p < .002$ for all comparisons). There were no significant differences in any proven sexual recidivism rates between the current study and the Thornton et al. study. Differences in proven violent reoffending rates between the current sample and Grubin's (2008) sample were not significant within any RM2000/v category. Overall proven violent reoffending rates, however, were significantly higher in the present sample compared with Grubin's sample ($\chi^2 = 5.67, p = .02$). This overall difference is partially due to differences in the risk distributions of the two samples, as

Table 2. Proven Reoffending Rates of Offenders in the Different RM2000/s, v, or c Risk Categories by Offence Type, for 2- and 4-Year Follow-Up Periods, and Odds Ratios Comparing any Reconviction or Caution for a Sexual (RM2000/s), Violent (RM2000/v), or Sexual and/or Violent Reoffence (RM2000/c) Between Risk Categories for the 2-Year Follow-Up

Type of reoffence	RM2000 category	2-year follow-up			4-year follow-up			
		N	% (n) reoffending	95% CI	Odds ratio	95% CI	N	% (n) reoffending
Sexual	RM2000/s							
	All	4,946	2.2 (110)	1.8-2.7			578	5.5 (32)
	Low	843	0.6 (5)	0.2-1.4			151	0.7 (1)
	Medium	2,489	1.5 (38)	1.1-2.1	Medium vs. low	2.60*	249	2.8 (7)
	High	1,266	3.3 (42)	2.4-4.5	High vs. medium	2.21***	134	9.0 (12)
	Very high	348	7.2 (25)	4.7-10.4	Very high vs. high	2.26**	44	27.3 (12)
Nonsexual violent	RM2000/v							
	All	4,946	9.0 (446)	8.2-9.9			578	10.2 (59)
	Low	1,815	1.1 (20)	0.7-1.7			253	0.8 (2)
	Medium	1,539	5.4 (83)	4.3-6.6	Medium vs. low	5.12***	183	11.5 (21)
	High	999	15.6 (156)	13.4-18.0	High vs. medium	3.25***	97	20.6 (20)
	Very high	593	31.5 (187)	27.8-35.4	Very high vs. high	2.49***	45	35.6 (16)
Sexual and/or nonsexual violent	RM2000/c							
	All	4,946	10.9 (539)	10.1-11.8			578	15.1 (87)
	Low	773	1.7 (13)	0.8-2.6			118	0 (0)
	Medium	1,884	4.8 (91)	3.9-5.8	Medium vs. low	2.97***	225	8.9 (20)
	High	1,936	16.3 (316)	14.7-18.0	High vs. medium	3.84***	207	25.1 (52)
	Very high	353	33.7 (119)	28.9-38.7	Very high vs. high	2.60***	28	53.6 (15)

Note: CI = confidence interval.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3. Rates of Proven Sexual or Nonsexual Violent Reoffending at the 2-Year Follow-Up for the Current Sample, Grubin’s (2008) Sample, and Thornton et al.’s (2003) Sample

Type of reoffence	RM2000 category	Percentage reoffending at 2-year follow-up		
		Current study	Thornton et al. (2003)	Grubin (2008)
Proven sexual reoffending	RM2000/s			
	All	2.2%	2.6%	6.0%**
	Low	0.6%	0.9%	1.4%
	Medium	1.6%	1.3%	4.5%**
	High	3.4%	5.7%	11.1%**
	Very high	7.2%	17.2%	23.8%**
Proven nonsexual violent reoffending	RM2000/v			
	All	9.0%	2.3%***	6.7%***
	Low	1.1%	0.0%	1.0%
	Medium	5.4%	2.8%	4.6%
	High	15.6%	3.0%***	14.8%
	Very high	31.5%	18.8%	23.3%

Note: Comparisons are between the current study and other studies.
 p* < .05. *p* < .01. ****p* < .001.

a greater proportion of the current sample were high or very high risk (32%), compared with 27% of Grubin’s sample who fell into these categories. Rates of proven violent reoffending were higher, at the 2-year follow-up, in the current sample than those reported in Thornton et al. (2003) and were higher for those in the RM2000/v high-risk category in the current study (*p* < .001 for both comparisons).

Table 2 also shows how many of those who had a proven reoffence for a sexual or violent crime were in each RM2000/s, v, and c risk category at a 2-year follow-up. Only 2-year follow-up data are used here because of the larger sample for whom these data are available. The increase in proven sexual reoffending rates is significant between each pair of adjacent RM2000/s risk categories with only one small overlap (between low and medium risk groups) in the confidence intervals of the groups’ reoffending rates. The odds of proven sexual reoffending over 2 years increased by 160% for those in the RM2000/s medium-risk category compared with those in the low-risk category. The odds of proven sexual reoffending rose significantly at each increase in risk category.

Similar patterns were found with RM2000/v and proven violent reoffending rates. There was no overlap in reoffending rate confidence intervals for any of the risk groups,

Table 4. AUC for RM2000/s, v, and c for Proven Sexual, Nonsexual Violent, and Combined Sexual and Nonsexual Violent Reoffending, Respectively, at a 2-Year Follow-Up Period

Outcome	RM2000/s		RM2000/v		RM2000/c	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
Proven sexual reoffending	.68	.63-.73	.54	.48-.59	.62	.57-.66
Proven nonsexual violent reoffending	.60	.58-.62	.80	.78-.82	.75	.73-.77
Proven combined sexual and nonsexual violent reoffending	.62	.60-.64	.75	.73-.77	.73	.71-.75

Note: AUC = area under the curve statistics; CI = confidence interval.

and the odds ratios indicate that those in a higher risk category were significantly more likely to be reconvicted of a violent offence within 2 years of being at risk than those in the adjacent, lower risk category. Indeed, the odds of proven violent reoffending over 2 years increased by more than 400% for those in a medium-risk category compared with those in the low-risk category. These findings are replicated with the RM2000/c categories and combined sexual and/or violent reoffending rates. As with the v scale, there was no overlap in confidence intervals for any of the c scale risk groups. The odds of combined reoffending over 2 years increased by almost 300% for those in the medium-risk category in comparison with those in the low-risk group.

Comparisons with other studies were not possible for the combined reoffending outcome. Grubin (2008) did not report combined reoffending rates, whereas Thornton et al. (2003) did so only for a follow-up of variable length, which averaged 3.7 years.

Predictive Accuracy of RM2000/s and RM2000/v

Receiver operating characteristic (ROC) analysis was used to establish the predictive accuracy of the RM2000/s, v, and c for different types of proven reoffending. The ROC charts the number of people correctly classified as recidivists against the number of people falsely classified as recidivists. The resulting statistic (the AUC) indicates the accuracy of the scale's classification of recidivists and is therefore a measure of how well a tool predicts relative risk (how well it is able to put people into distinct groups that differ in their reconviction rates). Absolute risk refers to the rates of recidivism attached to each risk group. AUCs do not tell us anything about how well a tool predicts absolute probabilities of recidivism. There is some debate in the literature about how to interpret AUCs. We follow the interpretation suggested by Rice and Harris (2005), and Douglas, Epstein, and Poythress (2008), where comparisons with Cohen's *d* suggest that an AUC value of 0.71 represents a large effect size and an AUC value of 0.64 represents a moderate effect size.

Table 4 shows the AUCs for the different scales for different proven reoffending outcomes. The AUC for the RM2000/s for proven sexual reoffending over a 2-year

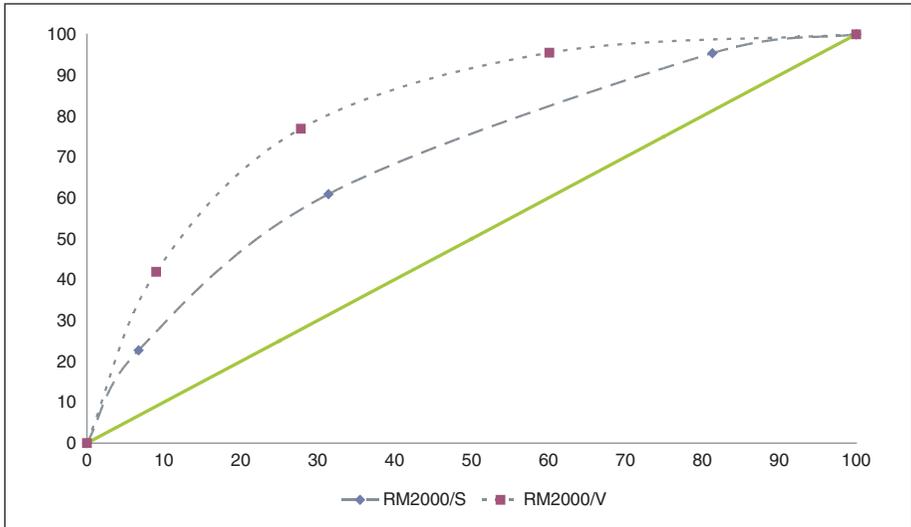


Figure 1. ROC curve based on RM2000/s and v categories and 2-year proven reoffending rates

follow-up period was .68, whereas for the combined reoffending outcome during that time, it was .62. This suggests that the RM2000/s has moderate predictive accuracy for proven sexual and combined sexual and/or violent reoffending. An additional AUC for 2-year proven sexual reoffending was also calculated for those for whom full RM2000/s scores were available. The AUC of the RM2000/s was .71 ($N = 899$) when the full scoring procedure was used and was .69 when the revised scoring procedure was used on the same sample. This suggests that the revised scoring procedure used in this study could lead to a slight underestimation of the predictive validity of the RM2000/s. AUCs for the 4-year follow-up sample are not reported for all groups due to the small number of reoffenders in this sample.

The AUC for RM2000/s for proven sexual reoffending was also calculated for a 2-year follow-up for those offenders for whom we also had 4-year follow-up data. In this case, at 2 years the AUC was .76, whereas for the same sample this increased to .78 at the 4-year follow-up.

The AUCs for the RM2000/v were large for both proven violent reoffending over a 2-year follow-up period (.80) and for combined proven reoffending (.75). We repeated the same comparison of AUCs as conducted for the RM2000/s, comparing the AUCs at a 2- and 4-year follow-up period for those offenders for whom these data were available (to control for any differences in the sample that could affect the outcome). The AUC for violent reoffending at the 2-year stage was .75, whereas for the same offenders, this increased to .79 at the 4-year follow-up stage.

The AUC for the RM2000/c for combined proven reoffending over a 2-year follow-up period was .73, whereas for proven violent reoffending alone, it was .75.

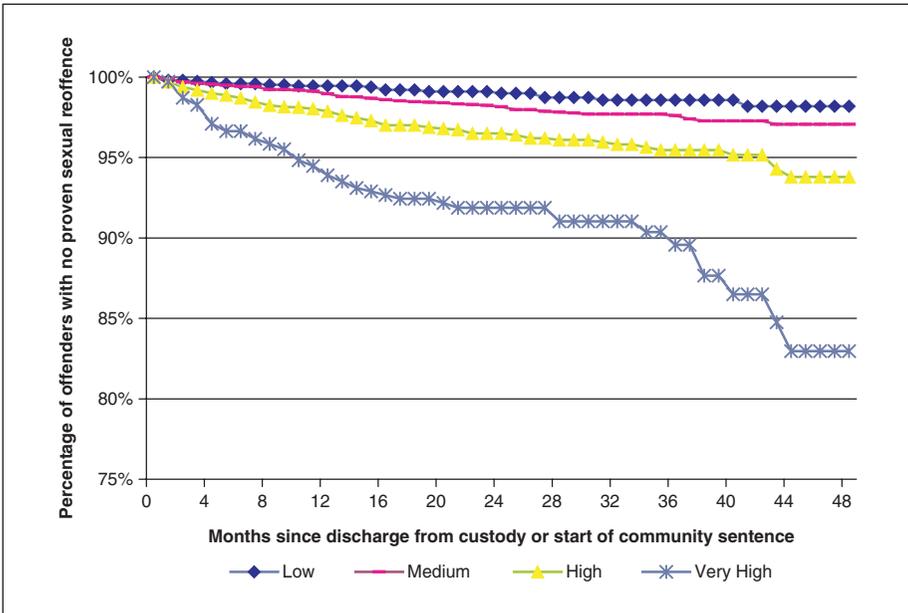


Figure 2. Kaplan-Meier survival curve of proven sexual reoffences over time by RM2000/s risk category

Figure 1 compares the sensitivity and specificity of the RM2000/s and v scales for reoffending within 2 years using data from Table 2. The low-risk RM2000/s group contained 19% of nonreoffenders, whereas for RM2000/v, it contained 40% of nonreoffenders. This indicates better specificity at the bottom end of the distribution for RM2000/v. The high- and very high-risk groups, combined, contained 61% of sexual reoffenders and 77% of nonsexual violent reoffenders, whereas the very high-risk groups only contained 23% and 42%, respectively. This indicates better sensitivity at the top end of the distribution for RM2000/v. These comparisons explain, in statistical terms, why the AUC of RM2000/v is considerably higher than that of RM2000/s. The specificity and sensitivity results may be helpful information for those responsible for targeting interventions and considering release decisions.

Survival analyses were conducted on both the RM2000/s and RM2000/v scales. Survival analysis allows a comparison of the different risk groups in relation to both their proven reoffending rate and the length of time that passed before they reoffended once they were at risk (time at risk). In survival analysis, an offender is followed until the end of the follow-up period (in this study that means until 6 months before the HOPNC data were extracted) or until (a) they commit a relevant reoffence which later leads to caution or conviction or (b) until some other significant event occurs that means the offender is no longer at risk (the earliest of the following events—reimprisonment

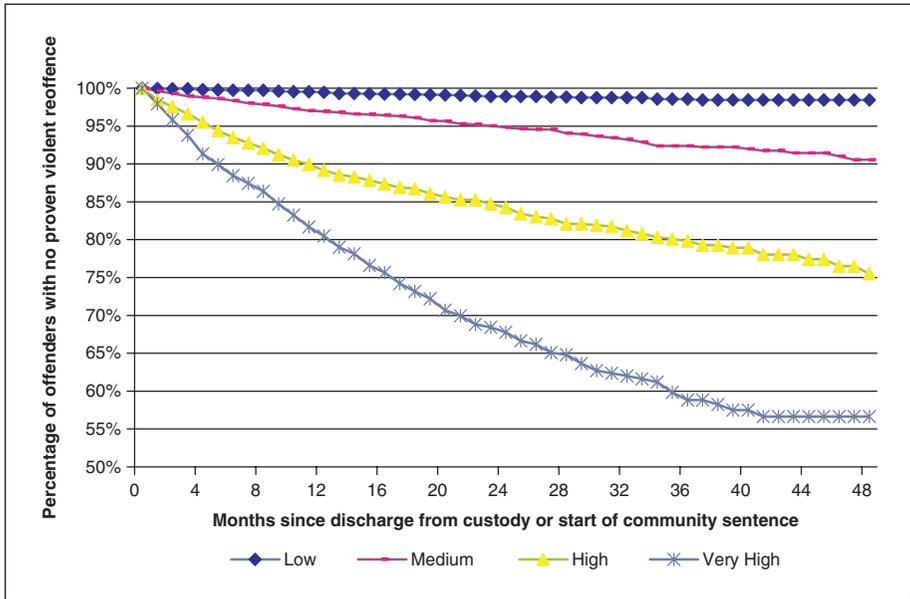


Figure 3. Kaplan-Meier survival curve of proven violent reoffences over time by RM2000/v risk category

for a nonrelevant offence, recall to prison, or imprisonment for any offence committed before the start of the follow-up period). Figure 2 shows that the rate of proven sexual reoffending varied between risk groups, with higher risk groups reoffending more quickly than lower risk groups. This difference was more pronounced in the very high-risk group, who reoffended more quickly and subsequently at a much higher rate for the entire follow-up period. Cox regression analysis was performed using RM2000/s risk level as the predictor variable and time to the perpetration of a proven sexual reoffence as the outcome. This indicated that RM2000/s risk level did significantly predict the hazard rate for proven sexual reoffending, $\beta = 0.77, SE = .07, \text{Exp}(\beta) = 2.15, p < .001$.

Similarly, Figure 3 shows that the rate of proven violent reoffending varied between risk groups, with higher risk groups reoffending more quickly than lower risk groups. Again, this difference was more pronounced in the very high-risk group, who reoffend more quickly and subsequently at a much higher rate for the entire follow-up period. Cox regression analysis indicated that RM2000/v risk category was a significant predictor of time to proven nonsexual violent reoffending, $\beta = 1.05, SE = .04, \text{Exp}(\beta) = 2.86, p < .001$. The much larger parameter estimates in the Cox regression of the RM2000/v scale suggests that the difference in proven reoffending rates between adjacent risk categories was more pronounced for the RM2000/v than RM2000/s, corresponding to the higher AUC value of RM2000/v.

Table 5. Two-Year Proven Sexual Reoffending Rates for Subgroups of Sexual Offenders, by RM2000/s Risk Category

RM2000/s category	Low		Medium		High		Very high		All	
Offender group	N	% reoffending	N	% reoffending	N	% reoffending	N	% reoffending	N	% reoffending
Adult victim (contact offences only)	170	0.6	413	1.5	135	3.0	38	5.3	756	1.7
Child victim (contact offences only)	459	0.4	670	1.5	333	2.1	95	5.3	1,557	1.5
Contact offence	629	0.5	1,083	1.5	468	2.4	133	5.3	2,313	1.6
Noncontact offence (excluding Internet offences)	3	0.0	46	2.2	86	11.6	49	16.3	184	10.3
Internet offences	138	1.5	237	0.0	113	1.8	26	11.5	513	1.4
18-24 years old	—	—	202	1.5	348	3.2	86	4.7	636	2.8
24-35 years old	—	—	767	1.7	345	3.5	74	6.8	1,186	2.5
35-49 years old	425	0.7	1,073	1.4	366	4.1	99	9.1	1,963	2.1
50-59 years old	238	0.4	309	1.6	142	1.4	56	8.9	745	1.7
60 years old or above	180	0.6	138	1.4	65	3.1	33	6.1	416	1.7
Current offence is sexual	770	0.7	1,376	1.2	667	3.5	208	8.7	3,021	2.1
Noncurrent sexual offence	73	—	1,113	1.9	599	3.2	140	5.0	1,925	2.4

Validity of the RM2000/s With Subgroups of Sexual Offenders

The validity of the RM2000/s for different groups of sexual offenders was also examined. Table 5 shows the 2-year rates of proven sexual reoffending for different subgroups of sexual offenders. Again only 2-year rates are presented here due to the small subgroup sample sizes at 4 years. Those who had committed contact offences against an adult victim were compared with those who had committed a contact offence against a child (defined as anyone below 16, plus certain 16- and 17-year-olds where the offender is aged 18 or above, as indicated by the HOPNC offence code). Those who were convicted of a noncontact offence (excluding offences relating to downloading or possessing indecent images of children, which seem to be a qualitatively different sort of offence) were compared with contact offenders. Sexual offenders in different age bands were compared, to ascertain the predictive validity of the tool with younger and older offenders, and last, those whose current offence was not sexual but who had previous sexual convictions were compared with those whose current offence was sexual. AUCs for these subgroup were also calculated, to establish whether the predictive validity (in relation to relative risk) of the RM2000/s differs between subgroups. Independent two-sample *t* tests were conducted to determine whether there were any significant differences in the AUCs of the scales between subgroups of sexual offenders and are reported where differences were found (Gönen, 2007, describes the use of these tests for AUCs).

Contact Offences Against Adults Versus Contact Offences Against Children

There was no significant difference in proven sexual reoffending rates between those who had been convicted of a contact offence against an adult and those who had committed a contact offence against a child at the 2-year follow-up ($\chi^2 = 0.1, p = .75$). Table 6 shows that the AUC of the RM2000/s for those with contact offences against adults in this time frame was moderate at .66, whereas for those with contact offences against children, it was .67. Table 5 indicates that, in each risk category, the rates of proven sexual reoffending were equivalent for those who had committed a contact offence against an adult and those who had offended in this way against children. Taken together, these findings suggest that the RM2000/s is moderately predictive of relative risk for both groups of contact offenders and that absolute rates of recidivism are the same for both groups at the 2-year follow-up point.

The same was not true when proven violent reoffending was examined. Two-year rates of proven violent reoffending were higher in the group who had committed a contact offence against an adult (6.5%) than for those whose index offence related to a contact offence against a child (3.3%), suggesting that they differ in absolute rates of recidivism ($\chi^2 = 12.6, p < .001$). However, the RM2000/v is a good measure of the relative risk of both of these groups; the AUC for this scale at the 2-year follow-up for the contact offence against an adult group was moderate to high (AUC = .73) and was even better for those whose index offence related to a contact offence against a child (AUC = .84), $t(1) = 5.44, p = .02$.

Table 6. AUC for RM2000/s, v, and c for Proven Sexual, Nonsexual Violent, and Combined Sexual and/or Nonsexual Violent Reoffending, Respectively, at a 2-Year Follow-Up Period, for Different Subgroups of Sexual Offenders

Group	RM2000/s		RM2000/v		RM2000/c	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
All	.68	.63-.73	.54	.48-.59	.62	.57-.66
Current/noncurrent						
Current sexual offence	.72	.66-.78	.80	.75-.84	.73	.69-.77
Noncurrent sexual offence	.60	.53-.68	.69	.66-.72	.64	.61-.67
Type of current offence						
All contact offenders	.67	.58-.75	.80	.76-.85	.71	.68-.76
Contact offences against adults	.66	.52-.79	.73	.66-.81	.69	.62-.76
Contact offences against children	.67	.57-.77	.84	.79-.89	.73	.68-.79
Noncontact (non-Internet) offenders	.65	.55-.75	.47	.30-.63	.58	.46-.68
Internet offenders	.70	.43-.97	—	—	.73	.52-.94
Age of offender						
18-24	.60	.49-.71	.79	.75-.83	.73	.69-.77
25-34	.62	.53-.72	.71	.67-.75	.64	.60-.68
35-49	.70	.62-.78	.80	.77-.83	.71	.68-.74
50-59	.72	.58-.85	.82	.72-.91	.75	.67-.83
60+	.72	.54-.91	.52	.39-.64	.65	.53-.78

Note: AUC = area under the curve statistics; CI = confidence interval.

Table 6 indicates moderate predictive validity for the RM2000/c for proven sexual or violent reoffending for those with a contact offence against an adult (AUC = .69) and even better for those with a contact offence against a child (AUC = .73). Due to the differences in proven violent reoffending described above, the absolute risk of combined reoffending differed between the groups, with 2-year rates of 8.1% for contact adult offenders and 4.7% for contact child offenders ($\chi^2 = 10.7, p = .001$).

Noncontact Offences Versus Contact Offences

The proven sexual reoffending rate of those convicted of a noncontact offence (excluding those convicted of possessing indecent images of children) was significantly higher than that of the contact offenders in our sample at the 2-year follow-up stage, at 10.3% versus 1.6% ($\chi^2 = 59.2, p < .001$), indicating a difference in the groups' absolute risk of recidivism.

However, the low-risk group of any other noncontact offenders comprised only three people; this is an artefact of the revised scoring system applied to those in the OASys sample. All noncontact non-Internet offenders in the OASys sample were

raised one risk category at Step 2 by virtue of their current offence. The three low-risk noncontact offenders were ISMG cases who were low at Step 1 and had none of the other three risk factors used in the full Step 2 scoring procedure. Although these two groups differed in their absolute risk of recidivism, the RM2000/s appeared to work moderately well for both groups as a measure of relative risk. The AUCs for proven sexual reoffending at the 2-year follow-up were .65 for those convicted of any other noncontact sexual offence and .67 for those with a contact offence.

However, the converse was true when proven violent reoffending was the outcome of interest. Whereas absolute rates of violent reoffending were similar for contact (4.3%) and noncontact offenders (3.3%) at the 2-year follow-up point, the relative predictive validity of RM2000/v was high for the former group (AUC = .80) and very poor and significantly worse for the latter group (AUC = .47), $t(1) = 15.26, p < .001$. As would be expected, we found a similar pattern in the AUCs for the RM2000/c for combined reoffending; with the RM2000/c obtaining an AUC of .72 with the contact offenders (who had a 5.8% rate of combined reoffending at 2 years) and .58 with the noncontact offenders (whose combined proven reoffending rate was 13.0% at 2 years). The difference in AUCs was significant— $t(1) = 5.42, p = .02$. Further examination showed that the age component of RM2000/v was negatively predictive for the noncontact group: all of this group's violent recidivists were aged 35 or above.

Those With an Internet Sexual Offence

There were no differences between the proven sexual reoffending rates of those whose index offence related to the Internet (coded here as those with an offence of making or possessing indecent images of children, as most of these offences relate to the use of the Internet), at the 2-year follow-up, and those who committed a contact offence (at 1.4% and 1.6%, respectively). The AUC for the RM2000/s for proven sexual reoffending over a 2-year follow-up period for offenders who were convicted of offences of making or possessing indecent images of children ($N = 513$) was .70, tentatively indicating moderate predictive accuracy for relative risk with this group. The equivalent AUC using the standard RM2000/s scoring rules remained at .70 (95% CI = 0.43-0.98). This group included only seven reoffenders, causing wide confidence intervals which limit the reliability of these results. There were only two (0.4%) proven violent reoffenders in the Internet offender group at the 2-year follow-up. Given this very low absolute risk of proven violent reoffending, an AUC was not calculated for the RM2000/v. Similarly, with just nine sexual and/or violent reoffenders, the ROC analysis for the RM2000/c, which indicated a large effect with Internet offenders, must be treated with caution.

Older and Younger Offender Groups

To establish whether the RM2000/s's predictive accuracy was affected by age, we calculated the AUCs of the RM2000/s for those in different age bands. At the 2-year

follow-up period, the AUC for those offenders from 18 to 24 years of age and those 25 to 34 years old indicate poor predictive accuracy with these younger groups (see Table 6). Using the same follow-up period, the AUCs increased (although not significantly) for those aged 35 to 49, 50 to 59, and those above 60, indicating that the RM2000/s had a large effect with the oldest groups of offenders in the sample. RM2000/s scoring rules ensure that all offenders below 35 are rated at least medium risk, meaning this group can only be assigned to one of three rather than four RM2000/s risk categories. Restriction of range reduces relative risk, which is why we find poorer predictive accuracy for this group (see also Hanson, 2008). The absolute risk within categories, however, was remarkably consistent across age groups.

Rates of proven violent reoffending, however, varied markedly across the age groups; the 18- to 24-year-olds had a 2-year rate of 16.5% proven violent offending compared with 11.7% for the 25- to 34-year-olds, 8.9% for the 35- to 49-year-olds, 2.7% for the 50- to 59-year-olds, and 1.7% for those aged 60 or above. Although the AUCs of the RM2000/v scale were moderate for most of these groups, it was poor for those above 60, at .52 (see Table 6). The AUCs for the 25- to 34-year and above-60 age bands were significantly poorer than the AUC for the 18- to 24-year age band— $t(1) = 6.21, p = .01$ and $t(1) = 15.64, p < .001$, respectively.

Accordingly, rates of proven combined sexual and/or violent reoffending decreased with age, from 18.4% for 18- to 24-year-olds to 3.1% for those above 60. The AUCs for the RM2000/c ranged from moderate to large for people of different ages, suggesting that the RM2000/c is a reasonable predictor of relative risk across different age groups (see Table 6). The only significant difference in AUCs for the RM2000/c was that the AUC was significantly better for the 18- to 24-year-olds than for the 25- to 34-year-olds, $t(1) = 8.83, p = .003$.

Those With an Index Sexual Offence Versus Those With Only Previous Convictions for Sexual Offences

Given that a number of the previous studies examining the predictive accuracy of the RM2000 have used samples consisting exclusively of those whose index offence was sexual in nature, we examined whether the RM2000/s, v, and c work as well for those whose current offence was not sexual but who had previous convictions for sexual offences (termed *noncurrent sexual offenders*). The two groups' absolute risks of proven sexual reoffending were similar at the 2-year follow-up as 2.1% of those whose current offence was sexual had a proven sexual reoffence in this time, compared with 2.4% of the noncurrent sexual offenders. However, RM2000/s was a poorer measure of relative risk for the latter group. At the 2-year follow-up period, its AUC for current sexual offenders was large, whereas its AUC for noncurrent sexual offenders was poor and significantly lower, $t(1) = 5.53, p = .02$.

Absolute rates of proven violent reoffending were significantly different in these groups, with 3.6% of the current sexual offenders having a proven violent reoffence within 2 years of being at risk, compared with 17.5% of the noncurrent sexual

offenders. ROC analyses indicate that the RM2000/v had a large AUC with current sexual offenders but was a moderate and significantly worse for noncurrent sexual offenders, $t(1) = 17.23, p < .001$.

A similar pattern was found when combined proven reoffending was examined. Absolute rates of combined reoffending differed significantly ($\chi^2 = 230.5, p < .001$) between current (5.5%) and noncurrent (19.3%) sexual offenders at the 2-year follow-up, and the RM2000/c was a better measure of relative risk for the current sexual offenders than for the noncurrent sexual offenders, $t(1) = 14.4, p < .001$.

Discussion

This study found that the RM2000/s had moderate predictive accuracy as a measure of relative risk, with a range of sexual offenders, suggesting that it is a robust tool that produces moderately reliable results across this heterogeneous group. The tool's ability to measure relative risk refers to its ability to classify offenders into groups that differ in their absolute rate of reoffending. We did not examine the ability of the scales to predict absolute risk, which is the actual rate of reoffending expected from those in each risk category. ROC analysis (which measures ability to determine relative risk) of the RM2000/s produced a lower AUC (.68) at 2 years than has been reported in previous studies; RM2000/s had an AUC of .75 in the Thornton et al. (2003) study and .73 with Grubin's (2008) Scottish sample. However, the current finding is in keeping with the results of Hanson and Morton-Bourgon's (2009) meta-analysis of the accuracy of risk assessments with sexual offenders, which found that this scale had a mean d of 0.67 (which equates to an AUC of .68; Rice & Harris, 2005) across 10 studies. The RM2000/v had the same level of predictive accuracy, as measured using ROC analysis, with the current sample (AUC = .80) as the Thornton et al. (2003) sample. The AUC of the RM2000/v with Grubin's (2008) Scottish sample was broadly similar (AUC = .78) and in a recent meta-analytic study was considerably worse (mean $d = 0.62, AUC = .67$; Hanson & Morton-Bourgon, 2009). A similar pattern emerges for the RM2000/c, which also performed equally well in the current study (AUC = .73) and the original validation sample (AUC = .74; Thornton et al., 2003). As with the RM2000/v, the c scale performed more poorly in the Hanson and Morton-Bourgon (2009) study, which reported the RM2000/c to have a mean d of 0.75 (AUC of .70) across five studies.

We suggest that some of these discrepancies could result from the composition of the samples used in these studies. The samples used in the original validation (Thornton et al., 2003) and in Grubin's (2008) study consisted of those whose index offence was either defined as sexual based on the legal definition of the conviction or by the underlying sexual motivation of a nonsexual criminal conviction. In contrast, as well as including those who met the criteria for the former studies, the current study included offenders whose current offence was nonsexual, either in legal definition or motive but who did have a previous caution or conviction for a sexual offence. The absolute rate of proven violent reoffending (10.2% over a 4-year follow-up) in the current study was

almost twice that of the rate of proven sexual reoffending (5.5% over the same time period). In both the Thornton et al. (2003) and Grubin (2008) studies the reported rates of violent reconviction were roughly equal to the rates of sexual reconviction, suggesting that the samples used in these two studies were qualitatively different to those in the current study. The inclusion of those whose index offence is nonsexual means that it is likely that the current sample consists of a higher proportion of “generalist” offenders, for whom sexual offending is part of a general criminal lifestyle. Conversely the earlier studies are likely to have consisted of a higher proportion of sexual “specialists”: that is, people whose every offence is sexually motivated. The differences in predictive accuracy of the RM2000/s found in these studies may, therefore, be a result of the differences in the samples used. Certainly our results indicate that the RM2000/s had better relative predictive validity with those in the sample whose current offence was sexual than for those who had a noncurrent sexual offence and may have a more general antisocial orientation. Very few noncurrent offenders could be classified as low risk because their overall criminal records increased their risk on Step 1 of the RM2000/s, and the resultant narrowing of the range of risk categories that this group could fall into can, as previously noted, lower AUCs. However, these differences in sample composition do not easily explain why those in Grubin’s (2008) study had a significantly higher 2-year rate of proven sexual reoffending than the current sample. It will be of interest to see whether this difference persists over a longer follow-up. The relationship between changes in a sample’s recidivism base rate and in the absolute rates of recidivism in each risk category has been considered (Harris & Rice, 2007; Mossman, 2006) but is not yet fully understood.

A related issue is that the RM2000/v predicted 2-year combined sexual and violent reoffending as (or slightly more) effectively as the RM2000/c (AUC = .75 for RM2000/v, and .73 for RM2000/c; confidence intervals overlap). Our results indicate that proven violent reoffending was far more frequent at the 2-year point among noncurrent (17.5%) than current sexual offenders (3.6%), whereas proven sexual reoffending rates were similar across both groups (2.4% and 2.1%, respectively). Therefore, this study’s inclusion of noncurrent sexual offenders has led to the dominance of proven violent reoffending, and thus the v scale, among all proven harmful reoffending outcomes. Again this could indicate that the RM2000/c makes a more worthwhile contribution when used with sexual specialists, whereas the RM2000/v appears to be more useful when used with those who are more generally criminal.

The study also examined whether recidivism rates, that is, absolute risk, differed between various subgroups. The absolute recidivism rates did not vary within risk category based on variation in the age of the victims (rapists, child molesters) or the age of the offender, findings which are not consistent with previous research. However, this study employed a different definition of *child victim* than has been used previously, defining a child as anyone below 16. It may be that key differences emerge between those who offend against prepubescent children and those who offend against adolescents, which would explain this finding. Further examination of these two groups of offenders would help to establish whether they are substantively different. The fact that

this study failed to find any substantive differences between the proven sexual reoffending rates of those in different age groups (a finding which is of some practical importance and is contrary to the findings of other studies) may be related to the current sample's inclusion of noncurrent sexual offenders. Rates of proven violent reoffending were far higher for the offenders in the younger age bands, with 2-year rates starting at 16.5% at age 18 to 24 and dropping with each ascending age category to just 1.7% for those above 60. This suggests that age is more influential in violent reoffending than sexual offending.

The absolute sexual recidivism rates varied significantly within risk category between contact and noncontact offenders. Noncontact offenders whose offences did not relate to possession of indecent images of children had a much higher rate of proven sexual reoffending (10.3%) than the other offender subgroups in the sample, whose sexual reoffending rates ranged from just 1.5% to 2.8% at the 2-year follow-up stage. This suggests that RM2000/s could underestimate the risk of recidivism posed by the noncontact, non-Internet group. However, this finding has come from a small sample of non-Internet noncontact offenders ($N = 184$); therefore, further examination of the patterns and rates of reoffending of this group is necessary before any firm conclusions can be drawn.

This study also briefly examined the predictive accuracy of the RM2000/s with Internet offenders, a (relatively) new group of sexual offenders. This study suggests that the RM2000/s, scored conservatively using the revised scoring procedures, as is currently advised in the manual (omitting the stranger and noncontact items from the scoring for those whose sexual offending history relates only to Internet offences), has moderate predictive accuracy for relative risk with this population. However, it must be emphasized that all Internet-only offenders in the sample were solely scored on Step 1 of the RM2000/s, not using the full RM2000/s scoring procedure, and very few of this subgroup were proven sexual reoffenders. This weakens the confidence we can place in these findings. The sample in this study included some Internet offenders who also had other sexual offences in their history, and therefore who may be more like prototypical sexual offenders than those whose offending is confined to downloading and sharing indecent images of children on the Internet. The authors have examined this latter group in more depth in a study of the validity of RM2000/s with Internet offenders (Wakeling, Howard, & Barnett, in press).

The relatively low base rates of proven sexual recidivism found in this study are in keeping with the findings of Hanson et al. (2010) who reported that routine correctional samples yielded lower base rates of sexual recidivism than preselected high-risk samples (e.g., samples consisting of those directed to civil commitment measures). The 4-year proven sexual reoffending rates ranged from 0.7% of the low-risk group to 27.3% of the very high-risk group. Given that these rates inevitably underestimate the true rates of sexual reoffending, this seems to be a meaningful distinction, and one on which decisions about resource allocations could usefully be based. The low 2-year rates, however, showed more limited discriminatory power, ranging from 0.6% to 7.2%, respectively. This suggests that the real value of RM2000/s may not be realized

immediately and that it would be used to best effect to assess risk of mid- to longer term, rather than shorter term, reoffending.

The present study found the RM2000/v scale to have very good accuracy as a predictor of relative risk of violent reoffending in this large sample of sexual offenders. It would be interesting to further examine this scale's predictive accuracy among nonsexual violent offenders. The absolute risk of proven violent reoffending, however, appears to differ across subgroups of offenders, suggesting that the RM2000/v is not a reliable measure of absolute risk. The same applies to the RM2000/c.

There are a few limitations to this study which need to be highlighted. First, in common with other validation studies, the full RM2000/s scale could not be scored for the whole sample. Our best estimate is that 21% of offenders in the OASys sample were misscored, resulting in a likely net 11% being raised a risk category. This may have reduced our estimates of the predictive accuracy of the tool. RM2000/s ratings will be collected routinely through OASys following an update of its operational IT system, scheduled for 2011; research data quality should therefore eventually improve. Second, recall information was not available for the whole of the custodial sample (this information was sought for approximately 90% of the sample), which might have had the effect of lowering the observed reoffending rates.

In summary, the RM2000/s scale demonstrates moderate AUCs when predicting the relative risk of proven sexual reoffending in a population of convicted sexual offenders in England and Wales, despite likely underestimation of its predictive validity due to the study's scoring procedure. The RM2000/v scale has large AUCs when predicting relative risk of violent reoffending with this group. The RM2000/c scale also demonstrates large AUCs when predicting sexual and violent reoffending among this group of sexual offenders. The results suggest that the RM2000 scales are valid for use with this particular population. The RM2000/s and v risk categories reliably distinguish those at higher risk of reoffending from those at lower risk, across all four risk bands, and these risk bands have good sensitivity and specificity. This study supports the use of the RM2000/s and v scales as a method of screening offenders for the purpose of directing valuable resources to those presenting a higher risk of reoffending.

Once a larger 5-year follow-up sample is available, it would be useful to compare the properties of the RM2000/s with an English population to those found with the Scottish sample in Grubin (2008). This would allow examination of how consistently the RM2000 performs across different populations of offenders. In addition, a larger sample for a longer follow-up would allow comparison of the validity of the RM2000 over time and, particularly, would allow better comparison of base rates of offending for those in different risk groups.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

- Beech, A., Fisher, D. D., & Thornton, D. (2003). Risk assessment of sexual offenders. *Professional Psychology: Research and Practice, 34*, 339-352.
- Bonta, A., Law, M., & Hanson, R. K. (1996). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123-142.
- Cooke, D. J., & Michie, C. (2009). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior, 33*, 541-542.
- Craissati, J. (2003, March). *Adjusting standard assessment and treatment models to meet the needs of sex offenders in the community*. Paper presented at the 12th annual Conference of the Division of Forensic Psychology, Churchill College, Cambridge, UK.
- Craissati, J., & Beech, A. R. (2005). Risk prediction and failure in a complete urban sample of sex offenders. *Journal of Forensic Psychiatry and Psychology, 16*, 24-40.
- Craig, L. A., Beech, A., & Browne, K. D. (2006). Cross-validation of the Risk Matrix 2000 sexual and violent scales. *Journal of Interpersonal Violence, 21*, 612-633.
- Doren, D. M. (2006). What do we know about the effect of aging on recidivism risk for sexual offenders? *Sexual Abuse: A Journal of Research and Treatment, 18*, 137-157.
- Douglas, K. S., Epstein, M. E., & Poythress, N. G. (2008). Criminal recidivism among juvenile offenders: Testing the incremental and predictive validity of three measures of psychopathic features. *Law and Human Behavior, 32*, 423-438.
- Elliot, I. A., & Beech, A. R. (2009). Understanding online child pornography use: Applying sexual offense theory to Internet offenders. *Aggression and Violent Behavior, 14*, 180-193.
- Fazel, S., Sjøstedt, G., Langstrom, N., & Grann, M. (2006). Risk factors in criminal recidivism in older sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 159-167.
- Gönen, M. (2007). *Analysing receiver operating characteristic curves using SAS*. Cary, NC: SAS Institute.
- Grubin, D. (2008). *Validation of Risk Matrix 2000 for use in Scotland*. Unpublished report prepared for the Risk Management Authority, Paisley, UK.
- Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 15*, 15-17.
- Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism amongst sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior, 34*, 198-211.
- 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*, 1-21.
- Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. In R. Prentky, E. Janus, M. Seto, & A. W. Burgess (Eds.), *Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management* (pp. 154-166). New York: New York Academy of Sciences.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119-136.
- Harris, G. T., & Rice, M. E. (2007). Characterizing the value of actuarial violence risk assessment. *Criminal Justice and Behavior, 34*, 1638-1656.

- Harris, G. T., Rice, M. E., & Quinsey, V. L. (2008). Shall evidence-based risk assessment be abandoned? *British Journal of Psychiatry, 192*, 154.
- Home Office (2002). *The Offender Assessment System OASys User Manual Version 2*. London: Home Office.
- Kingston, D. A., Yates, P. M., Firestone, P., Babchishin, K., & Bradford, J. M. (2008). Long-term predictive validity of the Risk Matrix 2000: A comparison with the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment, 20*, 466-484.
- Langton, C. M., Barbaree, H. E., Hansen, K. T., Harkins, L., & Peacock, E. J. (2007). Reliability and validity of the Static-2002 among adult sexual offenders with reference to treatment status. *Criminal Justice and Behavior, 34*, 616-640.
- Laulik, S., Allam, J., & Sheridan, L. (2007). An investigation into maladaptive personality functioning of Internet sex offenders. *Psychology, Crime & Law, 13*, 523-535.
- McNeil, D. E., Sandberg, D. A., & Binder, R. L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence. *Law and Human Behavior, 22*, 655-669.
- Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment, 18*, 41-63.
- Mossman, D., & Sellke, T. M. (2007, July 5). Avoiding errors about "margins of error" [Electronic letter]. *British Journal of Psychiatry*. Retrieved December 12, 2008, from <http://bjp.rcpsych.org/cgi/eletters/190/49/s60#5674>
- Murphy, W. D., & Page, I. J. (2008). Exhibitionism, psychopathology and theory. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment and treatment* (pp. 61-76). New York: Guildford.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's *d*, and *r*. *Law and Human Behavior, 29*, 615-620.
- Seto, M. C., & Eke, A. W. (2005). The criminal histories and later offending of child pornography offenders. *Sexual Abuse: A Journal of Research and Treatment, 17*, 201-210.
- Thornton, D. (2006). Age and sexual recidivism: A variable connection. *Sexual Abuse: Journal of Research and Treatment, 18*, 123-135.
- Thornton, D. (2007). *Scoring Guide for the Risk Matrix 2000.9/SVC: February 2007 Version*. Unpublished manuscript.
- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., et al. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of New York Academy of Sciences, 989*, 225-235.
- Thornton, D., & Travers, R. (1991). A longitudinal study of the criminal behaviour of convicted sexual offenders. In *Proceedings of the 1991 Prison Psychologist's Conference I Scarborough, England* (pp. 16-18). London: Her Majesty's Prison Service.
- Wakeling, H. C., Howard, P. D., & Barnett, G. D. (in press). Comparing the validity of the RM2000 scales and OGRS3 for predicting recidivism by internet sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*.
- Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of Internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19*, 449-465.