Improving Agreement About Intervention Plans in Probation by Decision Support

Jacqueline Bosker[1], Cilia Witteman[2], Jo Hermanns[3], and Donnalee Heij[1]

Abstract

Reliability in decision making about intervention plans is a necessary condition for evidence-based probation work and equal treatment of offenders. Structuring decision making can improve agreement between clinical decision makers. In a former study however, we found that in Dutch probation practice structured risk and needs assessment did not result in acceptable agreement about intervention plans. The Dutch probation services subsequently introduced a tool for support in decision making on intervention plans. This article addresses the question whether the use of this tool results in better agreement between probation officers. A significant and meaningful improvement in agreement was found on all domains of the intervention plan. Implications for probation practice are discussed.

Introduction

The growing knowledge base about what works in reducing recidivism has had a major influence on probation work in the last decades. Inspired by developments in

[1]University of Applied Sciences, Utrecht, The Netherlands
[2]Radboud University Nijmegen, The Netherlands
[3]University of Amsterdam, The Netherlands

Corresponding Author:
Jacqueline Bosker, Research Centre for Social Innovation, University of Applied Sciences, Heidelberglaan 7, 3584 CS Utrecht, The Netherlands.
Email: jacqueline.bosker@hu.nl

Canada and the United States, the so-called "what works movement" took flight in Europe in the 1990s, starting in Great Britain and followed by probation services in the Scandinavian countries, Ireland, and the Netherlands. One of the major findings in research about what works is that supervision and interventions should follow the risk and needs of offenders (Andrews & Bonta, 2010; Dowden & Andrews, 1999; Lowenkamp, Latessa, & Holsinger, 2006). Consequently, the risk and needs of offenders should be assessed to decide about an intervention plan.[1] Evidence that actuarial prediction of human behavior generally outperforms professional judgment stimulated the introduction of risk and needs assessment in forensic practice (Andrews & Bonta, 2010; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009). The first instruments to be introduced—so-called second-generation risk assessment, following first-generation professional judgment—were actuarial and consisted almost entirely of static risk factors such as the number of previous convictions and age (Andrews & Bonta, 2010). Although such instruments can predict recidivism satisfactorily, they are of little use for decision making about intervention plans. This limitation is solved in the third-generation offender assessments by incorporating dynamic risk factors, identifying criminogenic needs that can be changed (Bonta, 2002).

Researchers in several countries studied the use of risk and needs assessment by probation officers in daily practice. In a survey among community corrections staff across the United States, Miller and Maloney (2013) found that although nearly half of the participating staff could be characterized as "substantive compliers," well over half of the practitioners do not make full use of the assessment and often deviate from tool recommendations. Research in Sweden and Ireland also showed that some of the probation officers did not commit to the use of risk and needs assessment, but focused on social context problems, probably influenced by their background as a social worker (Fitzgibbon, Hamilton, & Richardson, 2010; Persson & Svensson, 2011). Based on research among probation officers in England and Wales, Fitzgibbon (2007) points to the danger of deskilling probation officers when the introduction of risk and needs assessment substitutes casework skills and leaves no room for continuity in the relationship between officer and offender. Moreover, other factors seem to be important for probation officers in decisions about intervention plans, such as the suitability of an intervention for the offender, and the offenders' willingness and ability to participate (Kemshall, 2010). These studies imply that the use of instruments for risk and needs assessment may not naturally result in intervention plans that match the risk and needs assessed.

Studies about intervention plans and service delivery by probation officers using tools for risk and needs assessment show similar results. In a study about community supervision in the Canadian province of Manitoba, Bonta, Rugge, Scott, Bourgon, and Yessine (2008) found that often the assessed criminogenic needs were not included in the intervention plans. Bosker, Witteman, and Hermanns (2013b) found that Dutch probation officers seemed to focus their plans on some criminogenic needs such as drug and alcohol abuse, cognitive skills, and emotional well-being, while other

criminogenic needs assessed such as relationships, accommodation, or finance often were not included in the plans. In a study among young offenders in Canada, it was found that officers did not use the risk and needs assessment to identify treatment needs (Flores, Travis, & Latessa, 2003).

To improve the match between the risks and needs assessed and service delivery, so-called fourth-generation risk assessment instruments were developed, in which assessment and case management are integrated (Andrews & Bonta, 2010). In the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) for example, after assessing the risk, needs, and responsivity of the offender, correctional practitioners must prioritize the criminogenic needs, define goals with every need, and choose an intervention to reach the goals. Kemshall (2010) states that in such tools risk prediction has taken second place and enhancing the match of offenders to interventions has become the main focus. Whether the use of fourth-generation risk assessment actually improves decision making about intervention plans has, to our knowledge, not been investigated yet.

One of the foundations for improved decision making about intervention plans is an acceptable level of agreement between probation officers when they develop an intervention plan for the same offender. Such agreement is important for two reasons. First, probation practice cannot be evidence-based if probation officers disagree in their decisions about what criminogenic needs should be addressed. Second, the equality of rights of offenders should be guaranteed. If the intervention plan has an impact on the offender's freedom, probation officers should have a certain degree of objectivity, or at least intersubjective agreement. For example, decisions about special conditions, the intensity of supervision, or means of control can have a large impact on the offender's life. Such decisions should not depend on the individual officers' opinions. Although the final decision about special conditions and electronic monitoring is usually made by the court, judges often use the advice of the probation service in their decision process (Boone, Beijer, Franken, & Kelk, 2008; Van Wingerden, Moerings, & Van Wilsem, 2011).

Decision Support for Intervention Plans

Since 2005, the probation service in the Netherlands has used a computerized instrument for structured risk and needs assessment called RISc (Recidivism Assessment Scales; Hildebrand, 2010a). The risk and needs assessment usually results in an intervention plan that contains several domains: the goals describing the changes in behavior or circumstances, the interventions[2] needed to achieve these goals, the sanction and special conditions imposed on the offender, the means of control, and the intensity of the supervision (Hildebrand, 2010b). In the initial version of RISc, intervention plans were based on a structured risk and needs assessment, but decision making about these plans was not supported. In a previous study, Bosker, Witteman, and Hermanns (2013a) described the agreement of Dutch probation officers about the intervention plans. In that study, a group of Dutch probation officers was asked to write intervention plans for four cases in which the risk and needs assessment was given. Although

results differed per domain of the intervention plan, overall agreement was poor (Bosker et al., 2013a).

The use of instruments for structured risk and needs assessment can indeed help probation officers reach an acceptable level of agreement about the risk of recidivism and the criminogenic needs in an individual case (Lowenkamp, Holsinger, Brusman-Lovins, & Latessa, 2004; Raynor, Kynch, Roberts, & Merrington, 2000). Research in other fields such as medicine and psychiatry showed that structured decision making can increase the reliability of treatment decisions (Galanter & Patel, 2005; Garb, 2005). In (mental) health care, introducing a (computerized) decision support system has improved practitioners' performance, for example, disease management, drug dosing, or transparency of the decisions (Garg et al., 2005; Witteman & Kunst, 1999). We therefore expect that agreement about intervention plans in probation can be improved by structuring the decision-making process.

In 2010, the Dutch probation service implemented a revised version of RISc that offered more support for decision making about the intervention plan (RISc3), turning it into a fourth-generation risk assessment instrument. The question addressed in this study is whether using this instrument improves agreement between probation officers about these plans compared to the previous situation where no such tool was used (RISc2).

Method

Participants

In our earlier study (RISc2-study), 44 probation officers had participated. We had planned to engage the same group of probation officers for the RISc3-study. However, some of these probation officers could not participate, either because they had moved to another job or task, or because they were not available. If possible, these probation officers were replaced by colleagues from the same location. As a result, 29 probation officers participated in the RISc3-study, 14 of whom had also participated in the RISc2-study. Table 1 shows characteristics of these participants.

Comparing the probation officers who participated in the two studies (see Table 1), the probation officers in the RISc3-study were more experienced than the officers in the RISc2-study: more years of service and a higher average number of risk assessments performed. This is partly because the officers who participated in the RISc2-study had two more years of experience when they joined the RISc3-study. In addition, the new probation officers who joined the RISc3-study were relatively experienced. To test whether the differences between the populations in the RISc2- and RISc3-study influenced the results in our analysis, we compared the results when we included all probation officers to the results when we included only the probation officers who participated in both studies. Characteristics of all Dutch probation officers are only available for the year 2009. In comparison to all probation officers, female probation officers were overrepresented in both studies, the participating probation officers were more experienced in performing risk and needs assessments but had fewer years of experience as a probation officer.

Table 1. Characteristics of Probation Officers Who Participated in the RISc2-Study and RISc3-Study in Relation to all Probation Officers in 2009.

	All probation officers (2009)	RISc2-study n = 44 (2009)	RISc3-study n = 29 (2012)	
Mean number of assessments ever (SD)	68 (56)	78 (66)	139 (105)	
Gender				
Female	64%	33 (75%)	25 (86%)	
Male	36%	11 (25%)	4 (14%)	
Mean years of service as a probation officer (SD)	9.7[a]	4.5 (4.5)	All	7.9 (5.2)
			In RISc2-study (n = 14)	7.5 (4.3)
			Not in RISc2-study (n = 15)	8.1 (6.2)

Note. RISc = Recidivism Assessment Scales.
[a]Indicative, because for some locations the information about the years of service was not available.

Instruments

Risk and needs assessment. The intervention plans are based on structured risk and needs assessment using RISc (Hildebrand, 2010a). RISc is based on the English and Welsh Offender Assessment System (OASys; Home Office, 2002). RISc contains 12 scales, corresponding to 12 criminogenic needs: (a) offending history; (b) current offence; (c) accommodation; (d) education and employment; (e) income and financial management; (f) relationships with partner, family, and relatives; (g) relationships with friends and acquaintances; (h) drug abuse; (i) alcohol abuse; (j) emotional well-being; (k) thinking and behavior; and (l) attitudes. Each scale contains risk items that are scored 0 (*no problem*), 1 (*some problem*), or 2 (*significant problem*). The total score of a scale represents the degree of criminogenic need. A criminogenic need is considered to be present when the scale score exceeds a certain threshold. The total RISc-score expresses the risk of recidivism. In addition to the risk of recidivism and the criminogenic needs, the probation officer assesses the responsivity and the risk of harm.

The psychometric qualities of RISc are considered sufficient to use it as a basis for the intervention plan. The internal consistency of the scales is adequate to good, alpha varies from .61 for Scale 6 (relationships with partner, family, and relatives) to .88 for Scale 4 (education and employment; Van der Knaap, Leenarts, & Nijssen, 2007). Inter-rater agreement about the assessed risk and needs has been reported as moderate to excellent (Van der Knaap, Leenarts, Born, & Oosterveld, 2012). The inter-rater agreement of the total score is good ($T = .68$). The predictive validity for general recidivism was examined by looking at the 2-year follow-up reconviction rates of a sample of

16,239 male and female offenders (Van der Knaap & Alberda, 2009), and was found to be sufficient (the Area Under the Curve (AUC) = .70 for male offenders and AUC = .68 for female offenders).

Intervention plan. The newly developed, computer-based decision support for the intervention plan contains several steps (Hildebrand, 2010b). As a first step, probation officers decide what criminogenic needs should change, making use of the results of the needs assessment. In addition, they prioritize these needs (options are as follows: high, moderately, a little, no importance to intervene). Second, for every criminogenic need that must be influenced, probation officers decide what goal(s) should be achieved and what interventions are needed to realize the goal(s). When an offender meets the inclusion criteria for a specific intervention, the instrument suggests this intervention as a possibility. Third, when the intervention plan is the basis of a pre-sentence report, probation officers can advise the court about the sanction and about special conditions. In line with Dutch policy that in a forensic setting all (treatment) programs should be imposed as a special condition, the interventions that were chosen in the previous steps are automatically suggested as a condition. When probation officers select a restrictive condition, a control measure is suggested (for example electronic monitoring to control the offender to stay at home during certain hours). Finally, the system gives a suggestion about the intensity of the supervision, based on the assessed risks of recidivism and harm, the risk of noncompliance, and the necessity for extra guidance by the supervising officer.

Procedure

The data collection for the RISc3-study took place in September 2012, 2½ years after the data collection for RISc2. The participating probation officers all completed four cases in one day, using the software they work with in daily practice, under supervision of a researcher. The researcher gave instructions and stayed in the room to make sure that the probation officers did not discuss the cases. Because some probation officers did not complete all four cases, the data collection resulted in 103 intervention plans: 27 for case 1, 24 for case 2, 27 for case 3, and 25 for case 4.

Cases

Participants assessed the same four cases as in the RISc2-study. The names and dates were changed for the RISc3-study so probation officers who had participated in the RISc2-study would not recognize the cases. Because of the large timespan between the two studies and the fact that probation officers who participated in both studies had performed a large number of assessments during that period, the chance of recognition is considered to be very small. The cases represent different and generally occurring offender profiles (see Table 2 for details). The first case is a 28-year-old man with an anxiety disorder who assaulted his wife. He is unemployed, has debts, and uses a lot of soft drugs. The second case is a 33-year-old high-risk male offender with a long offending record that started at the age of 16, and who is currently serving a prison

Table 2. Characteristics About the Four Cases Used in this Study.

| Case number | Risk level | Dynamic criminogenic needs | | Motivation for change |
		Significant problems	Some problems	
Case 1	Moderate to high	Relationships partner/family Emotional well-being	Education/work Finance Drug abuse Thinking/behavior Attitude	Moderate
Case 2	High	Housing Education/work Drug abuse Attitude	Finance Friends Alcohol abuse Thinking/behavior	Moderate, only regarding practical problems
Case 3	Moderate to high	Drug abuse Alcohol abuse	Education/work Thinking/behavior	Moderate
Case 4	Moderate to high	Housing	Finance Relationships partner/family Friends Drug abuse Emotional well-being Thinking/behavior	High

sentence because of serious assault and the possession of hard drugs. He is unemployed, homeless, has debts, and uses hard drugs. Earlier interventions all failed because of noncompliance. The third case is a 22-year-old man who threatened a police officer. He resisted authority more than once, leading to four earlier convictions. He left school without a qualification, is unemployed, and has no income. He has good contacts with his family and has many friends. He is a frequent soft drugs user and was drunk at the time of the offence. The fourth case is a 36-year-old woman, convicted for theft and fraud. She has a borderline personality disorder, uses cocaine, and has addicted friends who also have offending histories.

The intervention plans of the first and third case had to be made in the context of a pre-sentence report, for the second case an intervention plan had to be developed containing interventions to be delivered during detention, and in the fourth case the plan relates to a conditional release from prison. The basic assessment (the 12 scales with risk items and short descriptions of the situation on every scale) and conclusions about the risks of recidivism and harm, criminogenic needs, and responsivity were given with every case, to make sure that all probation officers started with the same information.

Categorizing Domains of the Intervention Plan

Every intervention plan contains decisions about several domains. These decisions were put in the same categories as were used in the RISc2-study. The criminogenic

needs, goals, and interventions were clustered into seven categories that match the dynamic criminogenic needs of the assessment: housing, education and work, finance, social network, addiction, personality, and other/none. For the intervention domain, we added a category "supervision" as a general intervention for support and practical aid. For details about the clustering, see Bosker et al. (2013a).

In the RISc3-study, probation officers not only decided whether a criminogenic need should be addressed or not (as in the RISc2-study), but they also prioritized that need into four categories (options are as follows: high, moderately, a little, no importance to intervene). To allow comparison with the RISc2-study, the decisions about high, moderate, and low importance to intervene are taken together as a category "criminogenic need to be addressed," resulting in a two category variable (importance to intervene: yes or no).

Decisions about the sanction domain were clustered into six categories (e.g., suspended sentence, prison sentence, treatment measure). The special conditions domain was clustered into eight categories (e.g., attend treatment, prohibition to use drugs or alcohol). Only three categories for decisions about control were distinguished (electronic monitoring, alcohol/drug test, and no control), and four levels of supervision (Levels 1, 2 or 3, and no supervision).

Most domains in the intervention plan, except for the intensity of supervision, give room for one or more decisions. Probation officers can decide to include one or more goals or one or more interventions in the intervention plan. To make a comparison possible, all decisions were dichotomized. In the domain goals for example, for every goal cluster we coded whether the intervention plan contains a goal from that cluster or not. The level of agreement was analyzed for the goal clusters altogether, resulting in a conclusion about the agreement about the domain goals per case. All intervention plans were coded by the first and fourth author. To test whether there was enough agreement about the coding, 20 plans were double coded (5 randomly selected intervention plans for every case). With a mean Cohen's kappa of .92 (range = 0.44 to 1.00), agreement was good enough to code the other files separately. The categories with a moderate agreement were double checked.

Analyses

Agreement can be measured by calculating the percentage agreement between decision makers. In this study for example, we found an average pairwise percent agreement between the probation officers about the criminogenic needs to be influenced of 70% in the RISc2-study and 88% in the RISc3-study. However, this is an overestimation because a certain amount of agreement may be expected by chance (Cohen, 1960). Therefore, agreement between the probation officers was measured using the average pairwise Cohen's kappa. Cohen's kappa is an often used statistic for inter-rater agreement of nominal variables that does correct for chance. The average pairwise Cohen's kappa is determined by first calculating the pairwise Cohen's kappa for every possible pair, and subsequently determining the mean kappa of all pairs. Calculations were performed with ReCal (Reliability Calculation), an online utility available at http://

www.dfreelon.org that computes reliability coefficients for more than two coders (Freelon, 2010). As a rule of thumb for the interpretation of kappa, agreement is considered to be poor when kappa is lower than 0.40, fair when kappa is between 0.40 and 0.59, good when kappa is between 0.60 and 0.74, and excellent when kappa is 0.75 or higher (Cichetti, 1994). We used a *t* test (95% confidence interval, two-tailed probability) to measure whether agreement on the domains in the RISc3-study differed significantly from the RISc2-study.

Results

In this section, first the results of the RISc3-study are presented and compared to the results of the RISc2-study. Then, the results are described for the selection of probation officers who participated in both studies, to test whether changes in the populations of the two studies affected the results.

Agreement Between Probation Officers About Domains of the Intervention Plan

Table 3 shows the agreement in average Cohen's kappa between the probation officers about the domains of the intervention plans in the RIS2-study and the RISc3-study. Overall, agreement between probation officers about the domains of the intervention plan is significantly better in the RISc3-study. In the RISc2-study, the average agreement about the domains was poor to fair, in the RISc3-study it is fair to excellent. Improvement differs per domain and is largest for the criminogenic needs that probation officers decided to influence, the goals probation officers formulated, and the means of control.

Looking at the decisions about the *criminogenic needs* to be influenced clustered into two categories (should be influenced or not), agreement in the RISc3-study is good and substantially better than in the RISc2-study. The mean agreement in the four cases improved significantly, $t(3604) = 32.69$, $p < .01$. In the RISc3-study, probation officers had to prioritize the criminogenic needs to be influenced. Agreement about the priority of the criminogenic needs turned out to be fair.

Agreement about the *goals* that probation officers formulated in the intervention plans improved from poor in the RISc2-study to fair in the RISc3-study. The average kappa of the four cases doubled, $t(4845) = 23.25$, $p < .01$. Improvement is largest in Cases 2 and 4, the cases where agreement between the probation officers was worst in the RISc2-study. Agreement about the *interventions* that probation officers chose to improve the offender's needs is fair in both studies, and improved slightly but significantly, $t(2749) = 7.67$, $p < .01$.

In the RISc2-study, agreement was best about the (advised) *sanction*. Agreement about this domain did not improve much in the RISc3-study, although the mean improvement in the four cases is significant, $t(2390) = 4.40$, $p < .01$. Table 3 shows that the agreement about the *conditions* improved from fair in the RISc2-study to good

Table 3. Agreement Between Probation Officers About Domains of Intervention Plan in RISc2-Study and RISc3-Study in Average Pairwise Cohen's Kappa.

Domain	Case 1		Case 2		Case 3		Case 4		Mean agreement	
	RISc2 (n = 43)	RISc3 (n = 27)	RISc2 (n = 43)	RISc3 (n = 24)	RISc2 (n = 44)	RISc3 (n = 27)	RISc2 (n = 43)	RISc3 (n = 25)	RISc2	RISc3
Needs[a]	.47	.75	.30	.83	.54	.67	.40	.72	.43	.74
Prioritized needs	—	.61	—	.41	—	.53	—	.39	—	.50
Goal	.31	.48	.17	.62	.40	.51	.09	.49	.25	.52
Intervention	.64	.73	.24	.40	.60	.64	.37	.44	.48	.57
Sanction	.81	.85	.30	.37	.79	.86	.51	.49	.60	.67
Conditions	.76	.81	.22	.32	.58	.77	.38	.66	.49	.66
Control	.31	.79	.16	.80	.25	.89	.22	.84	.23	.83

	All cases	
Intensity supervision	.20	.39[b]

Note. RISc = Recidivism Assessment Scales; *n* = number of probation officers.
[a]Criminogenic needs to be influenced, clustered in two categories.
[b]To measure agreement about the intensity of the supervision on the four cases, no missing values are accepted. Because in the RISc3-study one or more cases were missing for 11 probation officers, this agreement could only be measured for 18 of the 29 probation officers.

in the RISc3-study. Mean improvement taking the four cases together was found to be significant, $t(2621) = 14.36$, $p < .01$. Agreement about the means of *control* improved from low to excellent in all four cases. Overall, the mean kappa of the four cases improved significantly, $t(2353) = 39.39$, $p < .01$.

Finally, agreement about the *intensity of the supervision* that probation officers thought necessary improved significantly from low in the RISc2-study to nearly fair in the RISc3-study, $t(161) = 5.30$, $p < .01$. Although agreement about this domain is still poor, the mean kappa nearly doubled.

Differences Between Populations of Probation Officers

The probation officers who participated in the RISc3-study differed from the probation officers who participated in the RISc2-study. Therefore, we repeated the analyses with the selection of probation officers who participated in both studies. The results turned out to be similar to the results of the total groups that participated in the two studies, although figures differ here and there. For the group of probation officers that participated in both studies, improvement in agreement between the RISc2-study and RISc3-study is slightly lower for the domain criminogenic needs (mean improvement from .51 to .61, whereas an improvement from .43 to .74 was found for the whole group) but significant, $t(636) = -4.18$, $p < .01$. For the interventions domain, mean agreement between probation officers who participated in both studies was slightly higher and significant, $t(610) = -6.20$, $p < .01$. The mean agreement improvement from .47 to .64, against .48 to .57 for the whole group. For all other domains, mean

agreement is similar to the results for the whole group. Therefore, we conclude that the differences between the probation officers who participated in the RISc2- and RISc3-study hardly affected the results.

Discussion

We examined whether structuring the process of making decisions about intervention plans by using a decision aid increases reliability of these plans. We did so by comparing agreement between probation officers about several domains of the intervention plan in two studies. In the first study (Bosker et al., 2013a), when they used RISc2, probation officers had to decide about the domains of an intervention plan without decision support. In the second study (the study reported here), a revised version of RISc was used (RISc3) that supports decision making in different ways: by distinguishing different decision steps; by encouraging probation officers to describe the goals and interventions separately for every criminogenic need that they decided should be influenced; by presenting relevant information on screen; by suggesting possible decisions based on the risk and needs assessment and on probation policy; and by asking probation officers to justify their decisions.

It was found that agreement between probation officers about the different domains of the intervention plan was significantly improved by the introduction of a decision tool. Agreement improved from low or fair in the RISc2-study to fair or good in the RISc3-study. Improvement of agreement was largest in the domains criminogenic needs that should be influenced, goals, special conditions, means of control, and intensity of supervision. For decisions about interventions and advice about the sanction only a small improvement in agreement was found. In the introduction we described that in several countries the use of risk and needs assessment does not in itself lead to a probation practice that focuses on criminogenic needs. This study shows that using fourth-generation risk and needs assessment tools (Andrews & Bonta, 2010) can improve agreement about intervention plans in probation. That is a relevant finding for probation practice. First, because better agreement about the goals and criminogenic needs that should be influenced may improve the extent to which intervention plans focus on a positive change of dynamic risk factors that may support desistance from crime. And second, more agreement about the special conditions, means of control, and intensity of supervision improve equality of rights of offenders.

Some features of the decision support tool used in this study may have helped to improve agreement. Agreement about the criminogenic needs to be influenced, goals, and interventions may have been improved by using the needs profile as a basis for the decision making, and subsequently describing the goals and interventions for each criminogenic need separately. This helps probation officers not to overlook criminogenic needs present in a specific case. In an earlier study, we found that some probation officers seem to be focused on what they consider as the main problem related to the offence and to ignore criminogenic needs that may be very relevant to support desistance from crime (Bosker et al., 2013b). Structured decision making may prevent this. In RISc3, interventions are suggested when offenders match the inclusion

criteria. Because these suggestions were often not followed, we do not think this improved agreement between probation officers.

Agreement about the special conditions improved substantially. This may be due to the fact that in RISc3, suggestions for special conditions are given based on previous decisions about necessary interventions. In most cases, these suggestions were accepted by the probation officers. Agreement about control improved from poor to excellent. This improvement may be caused by the fact that the control options in RISc3 are limited to electronic monitoring and drug/alcohol tests. All other control measures are a standard activity of the supervisors and therefore not a separate decision in the intervention plan in RISc3. Reducing the number of options to choose from makes it easier to reach agreement. Moreover, probation officers in the RISc3-study especially agreed not to include additional control measures in the intervention plan.

Two changes may have helped improve agreement about the level of supervision. First, probation policy about offender supervision changed in the period before RISc3 was introduced: levels of intensity were described in detail (e.g., number of contacts with offender each month). Second, and following the new policy, in RISc3 a suggestion is given about the level of supervision based on the results of the risk and needs assessment. An unambiguous definition of intensity and a standardized way to decide about the intensity necessary in a specific case may improve agreement. In this light, finding a poor average agreement about the level of supervision in the RISc3-study is disappointing.

Limitations

Although we had intended to do so, it was not possible to work with the same group of probation officers in both studies. A majority of the probation officers who participated in the RISc2-study did not participate in the RISc3-study. These officers were replaced by new officers as much as possible. The new probation officers were more experienced than the officers who had dropped out. In the RISc2-study, we had already concluded that experience does not have a substantial effect on the agreement between probation officers about intervention plans (Bosker et al., 2013a), and indeed additional analyses in the RISc3-study showed that the higher number of experienced probation officers did not influence the results substantially. Because working with the same group of professionals in two studies proved to be difficult, working with two random samples might have been easier and probably would have led to similar results.

Because the same procedure was followed in the RISc2-study and the RISc3-study, the limitations that were described for the RISc2-study apply to this study too (Bosker et al., 2013a). The probation officers had to work with a paper case for which the basic assessment was given. This may limit the information about the offender. Also, contrary to the general procedure in Dutch probation work, intervention plans could not be discussed with the offender. This may have led to an overestimation of agreement because differences may already occur in the assessment of risks and needs (Van der Knaap et al., 2012) and by taking the offenders' perspective into account.

Because of practical limitations, probation officers formulated intervention plans for only four cases. These cases do not represent the whole population of cases of the probation services. Offenders who are supervised by the probation service vary in different aspects, such as offending behavior, criminogenic needs, and demographic features. To study the mean agreement about intervention plans of a representative group of probationers, more cases must be included. In this study, it was found that agreement between probation officers differed per case. It might well be that including different cases would have led to different conclusions about the mean agreement. Still, convincing improvement in agreement was found in all four cases. Therefore it is expected that the conclusion of this study, that the use of a decision aid to formulate intervention plans improves agreement, will also hold for other cases.

Agreement about the intervention plans was measured after clustering the specific decisions into general categories. Being in the same category does not mean that decisions themselves are identical. For example, some of the probation officers who formulated a goal about addiction in a specific case focused on complete abstinence from drugs while others thought a reduction in the use of drugs was more realistic. An analysis on a more detailed level might have given a somewhat different picture about the agreement, but on the whole it can be concluded that structuring the decision-making process about intervention plans can improve agreement between probation officers about the main lines of these plans.

To decide about an intervention plan in a specific case, probation officers must use and integrate a fair amount of information and considerations, both evidence-based and practice-based. In such complex decision tasks, relevant information can easily be overlooked. Instruments for structured decision making are meant to support probation officers, not to take over the decision making. Decisions about intervention plans must be made by probation officers with sufficient knowledge about effective interventions to reduce reoffending and support desistance. It is neither possible nor desirable to fully prescribe the best intervention plan in a specific case. However, decision support can facilitate this task and optimize decision making about intervention plans.

In contrast to the extensive body of research about risk assessment and behavioral interventions, research about intervention plans is scarce. We consider this a deficiency, because these plans can have far-reaching consequences for offenders. With this study we have tried to contribute to the knowledge of decision making by probation officers about intervention plans. Our conclusion is that structuring the decision-making process can improve agreement. Whether it also leads to better decisions is the next question to be answered.

Acknowledgments

Declaration of Conflicting Interests

Funding

Notes

1. An intervention plan may also be called a supervision plan, case management plan, risk management plan, or sentence plan.
2. Interventions can entail behavioral programs, psychological treatment, practical aid, and support by the probation service or local authorities.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Newark, JH: LexisNexis.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI) user's manual*. Toronto, Ontario, Canada: Multi-Health Systems.

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, *29*, 355-379.

Bonta, J., Rugge, T., Scott, T., Bourgon, G., & Yessine, A. K. (2008). Exploring the black box of community supervision. *Journal of Offender Rehabilitation*, *47*, 248-270.

Boone, M., Beijer, A., Franken, A. A., & Kelk, C. (2008). *De tenuitvoerlegging van sancties: maatwerk door de rechter?* [The execution of sanctions: Tailor-made by the judge?]. Utrecht, The Netherlands: Willem Pompe Instituut voor Strafrechtswetenschappen.

Bosker, J., Witteman, C., & Hermanns, J. (2013a). Agreement about intervention plans by probation officers. *Criminal Justice and Behavior*, *40*, 569-581.

Bosker, J., Witteman, C., & Hermanns, J. (2013b). Do intervention plans meet criteria for effective practice to reduce recidivism? How probation officers forget about social capital and basic needs. *European Journal of Probation*, *5*, 65-85.

Cichetti, D. V. (1994). Guidelines, criteria and rules of thump for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284-290.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

Dowden, C., & Andrews, D. A. (1999). What works for female offenders: A meta-analytic review. *Crime & Delinquency*, *45*, 438-452.

Fitzgibbon, D. W. M. (2007). Risk analysis and the new practitioner: Myth or reality? *Punishment & Society*, *9*, 87-97.

Fitzgibbon, W., Hamilton, C., & Richardson, M. (2010). A risky business: An examination of Irish probation officers' attitudes towards risk assessment. *Probation Journal*, *57*, 163-174.

Flores, A. W., Travis, L. F., & Latessa, E. J. (2003). *Case classification for juvenile corrections: An assessment of the Youth Level of Service/Case Management Inventory (YLS/CMI)*. Cincinnati, OH: Center for Criminal Justice Research.

Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, *5*, 20-33.

Galanter, C. A., & Patel, V. L. (2005). Medical decision making: A selective review for child psychiatrists and psychologists. *Journal of Child Psychology and Psychiatry*, *46*, 675-689.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67-89.

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., . . .Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, *293*, 1223-1238.

Grove, W. M., Zald, D. H., Lebow, Z. B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30.

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.

Hildebrand, M. (2010a). *Handleiding RISc versie 3. Deel 1: Basisdiagnostiek, verdiepingsdiagnostiek en integrale conclusie* [Manual RISc version 3. Part 1: Basic assessment, in-depth assessment and integral conclusion]. Utrecht: Reclassering Nederland.

Hildebrand, M. (2010b). *Handleiding RISc versie 3. Deel 2: Indicatiestelling* [Manual RISc version 3. Part 2: Intervention plan]. Utrecht: Reclassering Nederland.

Home Office. (2002). *Offender Assessment System OASys: User manual*. London, England:Home Office.

Kemshall, H. (2010). The role of risk, needs and strengths asessment in improving the supervision of offenders. In F. McNeill, P. Raynor, & C. Trotter (Eds.), *Offender supervision: New directions in theory, research and practice* (pp. 155-171). Oxon, UK: Willan Publishing.

Lowenkamp, C. T., Holsinger, A. M., Brusman-Lovins, L., & Latessa, E. J. (2004). Assessing the inter-rater agreement of the Level of Service Inventory Revisited. *Federal Probation*, *68*, 34-38.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned form 13,676 offender and 97 correctional programs? *Crime & Delinquency*, *52*, 77-93.

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior*, *40*, 716-736.

Persson, A., & Svensson, K. (2011). Signs of resistance? Swedish probation officers' attitudes towards risk assessments. *European Journal of Probation*, *3*, 95-107.

Raynor, P., Kynch, J., Roberts, C., & Merrington, S. (2000). *Risk and need assessment in probation services: An evaluation*. London, England: Home Office Research, Development and Statistics Directorate.

Van der Knaap, L. M., & Alberda, D. L. (2009). *De predictieve validiteit van de Recidive Inschattingsschalen (RISc)* [The predictive validity of the Recidivism Assessment Scales (RISc)]. Den Haag, The Netherlands: Wetenschappelijk Onderzoek- en Documentatiecentrum.

Van der Knaap, L. M., Leenarts, L. E. W., Born, M. P., & Oosterveld, P. (2012). Reevaluating interrater reliability in offender risk assessment. *Crime & Delinquency*, *58*, 147-163.

Van der Knaap, L. M., Leenarts, L. E. W., & Nijssen, L. T. J. (2007). *Psychometrische kwaliteiten van de Recidive InschattingsSchalen (RISc)* Interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit [Psychometric qualities of the Recidivism Assessment Scales (RISc) Inter-rater reliability, internal consistancy and concurrent validity]. Den Haag, The Netherlands: Wetenschappelijk Onderzoek- en Documentatiecentrum.

Van Wingerden, S. G. C., Moerings, M., & Van Wilsem, J. (2011). *Recidiverisico & Straf. De rol van het recidiverisico bij de straftoemeting* [Risk of recidivism & punishment. The role of risk of recidivism in sentencing]. Leiden, The Netherlands: University of Leiden.

Witteman, C. L. M., & Kunst, H. (1999). SelectCare: In aid of psychotherapists' treatment decisions. *Computers in Human Behavior*, *15*, 143-159.