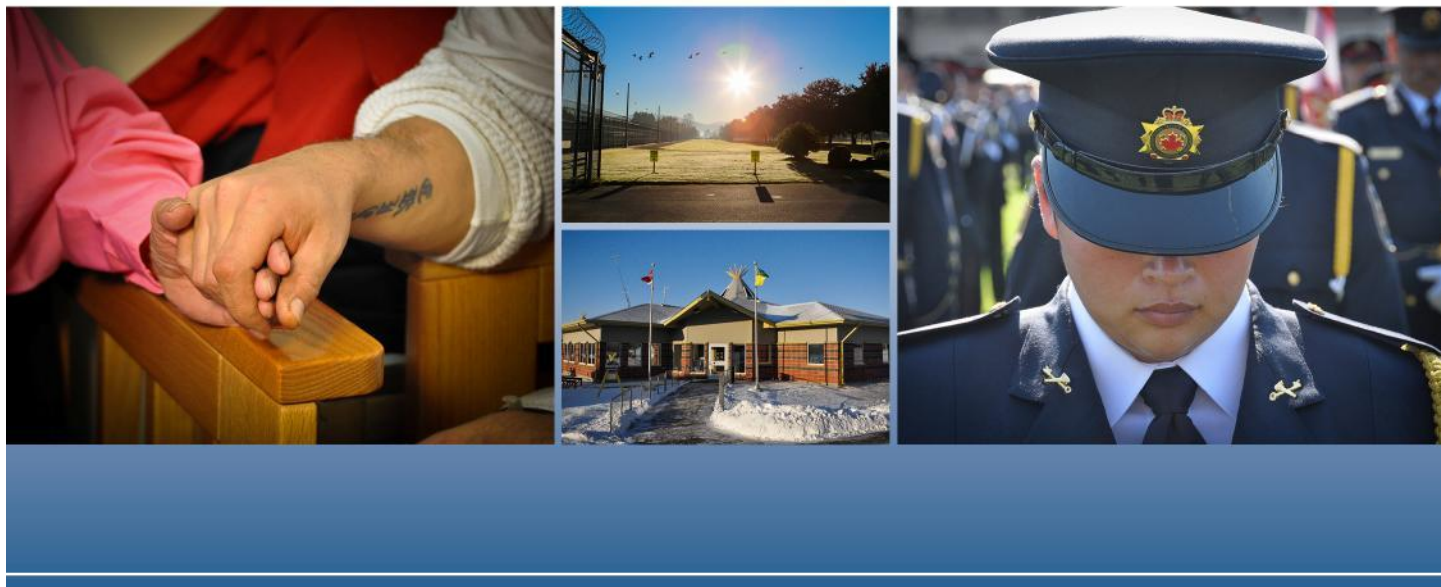


# CORRECTIONAL SERVICE CANADA

CHANGING LIVES. PROTECTING CANADIANS.



## RESEARCH REPORT

### Can the Predictive Validity of the Dynamic Factors Identification and Analysis – Revised Be Improved by Calculated Ratings?

**YEAR N° R-400**

Ce rapport est également disponible en français. Pour en obtenir un exemplaire, veuillez vous adresser à la Direction de la recherche, Service correctionnel du Canada, 340, avenue Laurier Ouest, Ottawa (Ontario) K1A 0P9.

This report is also available in French. Should additional copies be required, they can be obtained from the Research Branch, Correctional Service of Canada, 340 Laurier Ave. West, Ottawa, Ontario K1A 0P9.



Correctional Service  
Canada

Service correctionnel  
Canada

Canada



**Can the Predictive Validity of the  
Dynamic Factors Identification and Analysis – Revised  
Be Improved by Calculated Ratings?**

Geoff Wilton

Lynn A. Stewart  
&

Laurence L. Motiuk

Correctional Service of Canada

April 2017



## **Acknowledgements**

This research has been facilitated by the work of many people who developed and researched the tool and earlier versions of the measure, in particular, Shelley Brown, Craig Townson, and Dave Pisapio. In addition, the work could not have been conducted without the involvement of staff and offenders who participated in the assessment process. We would also like to thank John Weekes who provided feedback on an initial version of this report and Kaitlyn Wardrop for her assistance with formatting. Thanks also to Jennie Thompson and Dena Derkzen for their input on methodology and Kaitlyn Wardrop for her role in completing the initiative DFIA-R study that forms the basis of this current one.



## Executive Summary

**Key words:** *Dynamic Factors Identification and Analysis-Revised (DFIA-R), dynamic risk, offender needs, risk assessment, revocation, Indigenous offenders, women offenders*

The Dynamic Factors Identification and Analysis – Revised (DFIA-R) is a key component of the Correctional Service of Canada’s (CSC) Offender Intake Assessment process. Implemented in 2009, it is the current tool designed to assess dynamic risk. The DFIA-R is comprised of 100 indicators across seven domains (employment/education, marital/family, associates, substance abuse, community functioning, personal/emotional, and attitude) assessed by parole officers. Based on their professional judgement, an overall dynamic need rating is generated as well as a rating on each domain. Recent research has demonstrated that the overall rating and each of the domain ratings are predictive of revocations of conditional release across offender groups (Stewart, Wardrop, Wilton, Thompson, Derkzen & Motiuk, 2017). The present study examined whether calculated ratings could improve the predictive validity of the tool over the structured professional judgment provided by the parole officers at intake.

The study included 15,487 non-Indigenous men, 4,640 Indigenous men, and 1,195 women who had completed DFIA-R assessments (that included the assessment of indicators) and had been released. Because of the low number of revocations for women offenders we were unable to disaggregate by Indigenous status. Calculated domain ratings on the DFIA-R were based on the proportion of indicators endorsed on each domain for each offender group. Using this method, no endorsed indicators led to a rating of no need; fewer than 33% of indicators endorsed produced a low need rating; 33% to 66% of indicators endorsed was rated as moderate need, and more than 66% of indicators endorsed produced a high need rating. For the overall rating of need two calculated methods were used. The first was based on the mean of the proportions of indicators endorsed across the domains and the second, weighted each domain by its strength of association with outcome based on the survival analyses (hazard ratios) for each offender group. For the first method, offenders with fewer than 25% of indicators endorsed across the domains were rated low need, 25% to 50% endorsed were rated medium need, and greater than 50% rated high need. An adjusted calculation for women set the cut-offs at less than 33%, 33%-66%, and greater than 66% for low, moderate or high ratings respectively. We did not control for covariates given the need to maximize statistical power.

Results indicated that across all domains and for non-Indigenous men, Indigenous men, and women, ratings based on the proportion of indicators endorsed produced greater differentiation in predicting revocations (with or without an offence among offenders rated as no, low, moderate and high need than the parole officers’ ratings. Predictive validity was improved for all three study groups using these two methods. Both calculated methods of overall need ratings produced stronger predictions of revocations than ratings made by parole officers. For women offenders, the adjusted calculated method further improved the prediction of the overall ratings. Split-half reliability indicated that parole officers’ ratings of overall need had lower reliability than the calculated ratings. The results allow us to conclude that calculated ratings based on proportion of indicators endorsed provided greater predictive power and reliability than the parole officer ratings.





## Table of Contents

Acknowledgements.....	ii
Executive Summary .....	iv
List of Tables .....	vii
List of Figures .....	ix
List of Appendices .....	xi
Introduction.....	1
Method .....	5
Participants.....	5
Measures/Material.....	5
Analytic Approach.....	5
Calculated domain ratings.....	6
Overall dynamic risk ratings .....	7
Results.....	9
Distributions of Overall Ratings and Domain Ratings .....	9
Associations between Revocations and Parole Officer Ratings and Calculated Domain Ratings	17
Non-Indigenous men.....	17
Indigenous men.....	21
Women.....	25
Association between Revocations and Parole Officer Overall Intake Ratings and Calculated	
Overall Need Ratings.....	30
Non-Indigenous men.....	30
Indigenous men.....	32
Women.....	32
Split-half Reliability of Intake and Calculated Overall Need Ratings.....	35
Non-Indigenous men.....	35
Indigenous men.....	37
Women.....	38
Discussion .....	41
Associations of Domain Ratings with Revocations.....	42
Associations of Overall Ratings with Revocations.....	43
Split-Half Reliability of Overall Ratings. ....	44
Limitations .....	44
Conclusions.....	46
References.....	47

## List of Tables

Table 1 <i>Frequencies and percentages of offenders in the test sample for the overall rating methods</i> .....	10
Table 2 <i>Frequencies and percentages of parole officer intake ratings and calculated domain ratings: Federally sentenced men.</i> .....	13
Table 3 <i>Frequencies and percentages of parole officer intake ratings and calculated domain ratings: Women offenders</i> .....	16
Table 4 <i>Revocations associated within the three overall need rating methods for the test sample: Men</i> .....	31
Table 5 <i>Revocations by overall need rating methods for the test sample: Women (N = 575) ....</i>	35
Table 6 <i>Split-half reliability of associations between revocations and parole officer intake ratings and calculated overall need ratings: Non-Indigenous men</i> .....	36
Table 7 <i>Split-half reliability of associations between revocations and parole officer intake ratings and calculated overall need ratings: Indigenous men</i> .....	38
Table 8 <i>Split-half reliability of associations between revocations and parole officer intake and calculated overall need ratings: Women</i> .....	40



## List of Figures

<i>Figure 1.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Employment domain .....	18
<i>Figure 2.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Marital/Family domain.....	18
<i>Figure 3.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Associates domain.....	19
<i>Figure 4.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Substance use domain .....	19
<i>Figure 5.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Community Functioning domain .....	20
<i>Figure 6.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Personal/Emotional domain .....	20
<i>Figure 7.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Attitude domain.....	21
<i>Figure 8.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Employment domain .....	22
<i>Figure 9.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Marital/Family domain.....	22
<i>Figure 10.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Associates domain .....	23
<i>Figure 11.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Substance Abuse domain.....	23
<i>Figure 12.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Community Functioning domain.....	24
<i>Figure 13.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Personal/Emotional domain.....	24
<i>Figure 14.</i> Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Attitude domain .....	25
<i>Figure 15.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Employment/Education domain.....	26
<i>Figure 16.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Marital/Family domain .....	27
<i>Figure 17.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Associates domain.....	27
<i>Figure 18.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Substance Abuse domain .....	28
<i>Figure 19.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Community Functioning domain .....	28
<i>Figure 20.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Personal/Emotional domain .....	29
<i>Figure 21.</i> Percentage of revocations based on parole officer ratings and calculated ratings for Women: Attitude domain.....	29
<i>Figure 22.</i> Revocations associated with the three methods of overall need ratings for the test	

sample: Non-Indigenous men ( $N = 7,714$ ) .....	31
<i>Figure 23.</i> Revocations associated with the three methods of overall need ratings for the test sample: Indigenous men. ( $N = 2,319$ ).....	32
<i>Figure 24.</i> Revocations associated with the three methods of overall need ratings for the test sample using the unadjusted calculated scores: Women ( $N = 575$ ) .....	33
<i>Figure 25.</i> Revocations associated with the three methods of overall need ratings for the test sample using the adjusted calculated scores: Women ( $N = 575$ ) .....	34

## **List of Appendices**

Appendix A: Hazard Ratio Weights Applied to Calculate Overall Need Ratings .....	48
Appendix B: Survival Analyses Comparing Intake and Calculated Domain Ratings on Revocations.....	51
Appendix C: Survival Analyses Comparing Parole Officer Intake, Mean, and Weighted Mean Overall Ratings on Revocations .....	57



## **Introduction**

Accurate prediction of offender outcomes is a key component of criminal justice systems whose ultimate goal is safe offender reintegration. Throughout the criminal justice process, various decisions such as determining sentence length, security level, intensity and type of correctional interventions, candidacy for parole, and the imposition of release conditions involve at least an implicit assumption that public safety will be enhanced by accurate decisions.

The Correctional Service of Canada (CSC) uses several assessment tools designed to provide the necessary information to formulate informed decisions. Many of these assessments were designed to adhere to the Risk-Need-Responsivity (RNR) framework which guides decisions regarding the implementation of correctional interventions. Applying the RNR framework, offenders with a greater risk of recidivism are provided higher intensity interventions (the risk principle); interventions target dynamic factors associated with offending (the need principle); and, effective cognitive behaviour modification techniques tailored to the specific learning styles, motivation level, and abilities of offenders are applied in these interventions (the general and specific responsivity principles) (Andrews & Bonta, 2006; Andrews, Bonta, Wormith, 2006).

In addition to setting the course of intervention for individual offenders, assessments provide important information on the profile of the general offender population. The results of these assessments contribute to informed planning and resource allocation for the correctional agency. It is important therefore that these critical assessments be valid and reliable reflections of offenders' risk of recidivism and areas of criminogenic need.

At admission to federal custody, offenders in Canada are administered the Dynamic Factors Identification and Analysis Revised (DFIA-R) measure as one component of a comprehensive intake assessment. The DFIA-R was developed following the examination of empirically and theoretically supported dynamic factors and indicators associated with criminal behaviour (Brown & Motiuk, 2005). The tool identifies relevant criminogenic needs that should be the focus of correctional intervention during the offender's sentence. The DFIA-R was introduced in 2009 following research by Brown and Motiuk (2005) that led to the revision of the original tool. This resulted in the incorporation of additional assessments (e.g., assessments of motivation and responsivity), the reduction in the number of indicators, and the inclusion of a



possible fifth level of rating on the domains. The wording of several indicators was revised and training and instructions provided to intake parole officers were improved.

As part of the Offender Intake Assessment, the DFIA-R falls among fourth generation risk assessment strategies (Andrews, Bonta, & Wormith, 2006). The first generation consisted of unstructured clinical judgments of the probability of criminal behaviour. Second generation assessments incorporated empirically supported static items. The third generation assessments were more theoretically driven and added dynamic factors, and fourth generation assessments strengthened adherence to responsivity principles by linking outcomes with risk, need, strengths and responsivity assessments and service plans, delivery and intermediate outcomes (Andrews et al., 2006).

The DFIA-R is completed by specialized parole officers on all offenders admitted to CSC. The tool consists of 100 indicators scored as yes or no in seven domains – employment/education (12 indicators), marital/family (16 indicators), associates (11 indicators), substance abuse (18 indicators), community functioning (7 indicators), personal/emotional (25 indicators), and attitude (11 indicators). With the exceptions of the substance abuse and personal/emotional domains, each domain has five possible ratings – factor seen as an asset to community adjustment, no immediate need for improvement, low need for improvement, moderate need for improvement and high need for improvement. The substance abuse and personal/emotional domains do not have the asset rating option. Finally, an overall dynamic risk rating of low, medium or high is applied using structured professional judgement.

Recent research has examined the validity of this revised version of the tool (Stewart, Wardrop, Wilton, Thompson, Derkzen, & Motiuk, 2017). The results of the research established that the indicators were significantly associated with domain ratings, the indicators had a high degree of internal consistency, and the indicators, domain ratings, and overall ratings were generally related to both revocations and revocations with an offence.

Although the tool demonstrated good psychometric properties, it was noted that the distribution of ratings was restricted largely to moderate or high ratings. Very few offenders were assessed in the low need category, yet the level of recidivism for federal offenders on release is quite low. It appeared that parole officers were very cautious in assigning a low need rating. What is more, the overall ratings on the measure and the domains significantly predicted returns to custody across offender groups. Although some of the indicators were more predictive

for some groups than others, there were more commonalities in the indicators that predicted revocations the strongest than differences across the groups.

Validating a tool across offender groups has become a concern for CSC and other correctional agencies that have recently been challenged on the application of “generic” instruments to all offender groups. In particular, advocates have objected to standardized tools widely applied to women and offenders from diverse backgrounds. Legally, and ethically, tools applied for decision making purposes must demonstrate their relevance for specific groups as well as the population of offenders. One method of addressing this concern would be to empirically weight the components of the assessment based on their importance for each offender group. An empirical scoring system might also address the issue of the low prevalence of parole officers’ rating of low risk.

A previous study using empirically derived scores for an early version of the risk needs assessment (Community Needs Identification and Analysis; Motiuk & Porporino, 1989), though promising, was deemed too complicated at that time and it was decided that it did not provide sufficient additional predictive power over structured professional judgment to warrant making a change to an automated scoring procedure. Over the last years, however, developments in the information technology permit immediate calculations and CSC has successfully implemented other tools where parole officer assessment information is scored empirically (e.g., Custody Rating Scale), albeit allowing for professional overrides. Wormith (2017) has noted that research has been accumulating that demonstrates that calculated assessments outperform structured clinical judgements. This research therefore provided an opportunity to test possible improvements to the predictive validity and the precision of the DFIA-R by examining statistical methods to rate the domains and level of overall need. Specifically, the study’s objectives are to:

1. Assess whether calculated domain ratings and calculated overall need ratings will have stronger associations with revocations than parole officer intake domain and overall ratings on first admission to custody for non-Indigenous and Indigenous federally sentenced men and women.
2. Explore whether overall ratings calculated based on mean proportions of indicators endorsed across the domains or mean proportions of indicators endorsed across the domains weighted by hazard ratios is a stronger predictor of revocations.
3. Examine the split-half reliability of the weighted means method of calculating overall

ratings compared to the intake ratings and mean method of calculating overall need ratings.

## **Method**

### **Participants**

The population of offenders with at least one complete assessment of the DFIA-R domains and indicators was identified for this study. This population was the same one examined in recent validation study of the tool (Stewart et al., 2017). Analyses were restricted to those who were released on parole or statutory release prior to January 15<sup>th</sup> 2017 with at least 30 days of their sentence remaining following release to the community before their sentence ended or data was collected. This included 15,487 non-Indigenous men, 4,640 Indigenous men, and 1,195 women. Released offenders were followed until either their conditional release was revoked, their sentence ended, or the data collection date. The average number of days between release and the end of follow-up was 373 days for non-Indigenous men, 274 days for Indigenous men, and 389 days for women.<sup>1</sup>

### **Measures/Material**

The data for the present study were collected from the Offender Management System (OMS) databases. OMS is the official electronic record for all federal offenders in Canada.

The Dynamic Factors Identification and Analysis – Revised (DFIA-R) has been described previously in the introduction. There are five domain ratings for five of the domains: factor seen as an asset to community adjustment, no immediate need for improvement, low need for improvement, moderate need for improvement and high need for improvement and four rating categories for substance abuse and personal emotional domains. For the purposes of this research, the asset rating was combined with no need into a single category due to low frequencies in the asset category.

Revocations of conditional release was the outcome variable of this study. Revocations most often occur for technical reasons such as violation of release conditions or increased risk of criminal activity, but they may result from new criminal offences.

### **Analytic Approach**

The proportions of offenders across the parole officers' intake domain and overall ratings

---

<sup>1</sup> It should be noted that while the differences in time at risk across groups affect the interpretation of the percentages of revocation they are not a concern in comparing the relative validity of the three scoring methods.

were compared to the analogous ratings from the calculated domains and the two methods of calculated overall ratings.

The proportions of offenders with revocations across the parole officers' intake domain and overall ratings were compared to the analogous ratings from calculated domain and overall ratings. The calculated domain and overall need ratings were based on the indicators from the intake assessments. The examinations of proportions were supplemented with Cox regression survival analyses.

### **Calculated domain ratings**

Calculated domain ratings were based on the proportion of the indicators endorsed within that domain for an offender. Previous research with federal offenders found that some DFIA-R indicators were not predictive of revocation outcomes. For this analysis, indicators consistently unrelated to revocations or associated with reduced likelihood of revocations were not included in the calculated domain ratings. These included: "Belief in oneself to improve employability is low" of the employment/education domain rating; "Uses excessive force to discipline child" from the marital/family calculated domain rating; "Has deviant sexual preferences" and "Displays deviant sexual attitudes" were excluded from the calculation of the personal/emotional calculated domain, and "Denies crime or uses excuses to justify or minimize crime" was excluded from the attitudes calculated domain. "Formally investigated for suspicion of child abuse/neglect" was excluded from the marital/family calculated domain rating for men but retained for women.

Offenders with no indicators endorsed on a domain received a rating of no need. Those with fewer than one third of indicators endorsed received a domain rating of low need. Endorsement of one third to less than two thirds of the indicators on a domain resulted in a rating of moderate need, and endorsement of two thirds or more of the indicators resulted in a rating of high need for the domain. These initial cut offs were derived based on their face validity.

Ratings of "Asset to community adjustment" and no need were combined into a single category for the original intake domain ratings. Therefore, both the parole officer intake ratings and calculated ratings had four levels – no need, low, moderate, and high need.

The patterns of results observed with the proportions of revocations were confirmed by Cox Regression survival analyses.

### **Overall dynamic risk ratings**

Two methods were explored to calculate the overall dynamic risk ratings. The first method used the percentages of indicators endorsed within each domain (without exclusions) and averaged across domains. For example, an offender with half the indicators endorsed in each domain would have a score of 0.5 while an offender with half the indicators endorsed in four domains but no indicators endorsed in the other three domains would have a score of 0.29 or 2  $(0.5 + 0.5 + 0.5 + 0.5)$  divided by 7 – the number of domains. In this way, each domain received the same weight as the others regardless of the number of indicators comprising the domain. Cut-off values of 25% and 50% were used to determine the low, moderate, and high need ratings. Offenders with an average of 25% or fewer indicators endorsed across domains had a rating of low; those with an average of greater than 25% to 50% of indicators endorsed across the domains were rated as moderate, and those with an average of greater than 50% were rated as high need.

The second method used the same percentages of indicators endorsed within each domain and entered them into Cox regression models for each group – non-Indigenous men, Indigenous men, and women. The hazard ratios from these models produced for each domain were multiplied by the percentages of indicators endorsed within each domain, summed and divided by seven to achieve an overall score. The hazard ratios used to weight each domain differed for the three groups, and therefore, the strength of each domain in reaching the overall rating was customized for each group. (See Appendix A for the hazard ratio weights applied to each domain for each group of offenders.) For example, the largest hazard ratio for women offenders was associated with the substance abuse domain, and it therefore was given a greater weight than the other domains. For non-Indigenous men the attitudes domain was strongest, while the employment/education domain was strongest for Indigenous men. The average of the hazard ratio weights for each group was multiplied by the 25% and 50% cut-off values applied in the first method to determine analogous cut-off values for the weighted averages.<sup>2</sup> (Appendix A provides these formulae.)

Using hazard ratio results from prediction models in calculating ratings that are used in further prediction models can result in over-estimating the strength of the predictions associated

---

<sup>2</sup> The regression model included the scores from all the domains together; therefore, the hazard ratios represent each domain's unique contribution to predicting revocations in the overall model when all domains are considered together and not the individual impact of the domain.

with those ratings, and the ratings may not apply as well to other groups of offenders. Therefore, for the second method, a split-half reliability method was applied in which offenders were randomly assigned to either a development sample used to calculate the hazard ratios or a test sample. Prediction for the development sample would be stronger than for the test sample if the hazard ratio weights do not generalize well to other groups of offenders. Similar results for the two samples would confirm good reliability of the hazard ratio weighted method of calculating overall ratings. For consistency, the split-half reliability method was applied to both methods of calculating overall need ratings as well as the parole officers' intake ratings.

## **Results**

### **Distributions of Overall Ratings and Domain Ratings**

Table 1 shows the distributions of the overall ratings using the method conducted by parole officers at intake and the two calculated methods. For both Indigenous and non-Indigenous men the two calculated scoring methods resulted in an increase in the proportion of offenders rated low need. The proportion of low ratings according to the mean and weighted methods was at least double that produced from parole officers' intake ratings for both groups of men. Using the mean calculated method for non-Indigenous men, the increase in the proportion of low ratings was associated with a decrease in the proportion with high ratings. Using the weighted method it resulted in a decrease in medium ratings. For Indigenous men, the increase in low ratings was associated with a decrease in high ratings for the mean method and a decrease in both medium and high ratings for the weighted mean method.

Results of the distribution of overall ratings by method did not follow the same pattern for women as it did for men. The proportions of women with medium need ratings decreased for both the mean and weighted mean calculated methods compared to the parole officers' intake ratings. The proportion with low ratings also slightly decreased in the calculated overall ratings compared to the parole officers' intake ratings. This resulted in the proportion of women with calculated high ratings being higher than what is produced by the parole officers' intake ratings.

These results indicate that women have greater needs according to the number of indicators endorsed than the parole officers' overall ratings would suggest. While this reflects the extent to which federally sentenced women are a group with varied needs, their rates of revocation, and in particular, their rates of revocation with an offence, are lower than those of federally sentenced men. We were concerned that increasing the proportion of women in the high need group could potentially disadvantage women for decision making purposes. Therefore, we provided an adjusted calculated scoring formula which created more stringent criteria for designating a high need rating. The adjusted calculated scores required that less than 33% of the indicators be endorsed for a low rating, 33% 66% be endorsed for a moderate rating and greater than 66% for a high rating. Using these cut-off values, both methods of calculated ratings produced slightly greater number of women offenders rated as low need and medium need and fewer women rated as high need compared to the parole officers' intake ratings. The differences



were greater for the mean method than the weighted mean method.

Table 1

*Frequencies and percentages of offenders in the test sample for the overall rating methods*

Rating	Parole Officer Intake Rating		Mean Calculated Rating		Weighted Mean Calculated Rating	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Non-Indigenous						
Men						
Low	718	9.3	2,286	29.6	1,619	21.0
Medium	3,531	45.8	3,503	45.4	2,707	35.1
High	3,465	44.9	1,925	25.0	3,388	43.9
Indigenous Men						
Low	105	4.5	237	10.2	208	9.0
Medium	961	41.4	964	41.6	914	39.4
High	1,253	54.0	1,118	48.2	1,197	51.6
Women (unadjusted)						
Low	121	21.0	97	16.9	102	17.7
Medium	295	51.3	231	40.2	175	30.4
High	159	27.7	247	43.0	298	51.8
Women (adjusted)						
Low	121	21.0	161	28.0	143	24.9
Medium	295	51.3	337	58.6	311	54.1
High	159	27.7	77	13.4	121	21.0

Table 2 presents the parole officers' ratings and calculated domain ratings of non-Indigenous and Indigenous men. In the calculated domains, the asset or no need category was reserved only for offenders with no indicators endorsed. The substantially higher proportion of

asset or no need ratings resulting from the parole officers' intake rating method than the calculated domain ratings indicates that many non-Indigenous men receive ratings of asset or no need at intake despite having at least one indicator endorsed. For example, the parole officers' intake ratings on the community functioning domain indicate that nearly 70% of non-Indigenous men and nearly 54% of Indigenous men have an asset or no need in this domain. However, most offenders with an asset or no need rating have at least one of the seven community functioning indicators endorsed, and over 50% of non-Indigenous men and over 68% of Indigenous men have more than one third of the community functioning indicators endorsed.

In addition to discrepancies at the no need rating level, the proportions of offenders rated as high need also often differed substantially between the parole officer intake ratings and calculated ratings. For non-Indigenous men, there were more than three times more men rated as high need on the employment/education domain and more than five times more men rated as high on the community functioning domain using the calculated ratings than the parole officer intake ratings. Calculated ratings of high need in the personal/emotional domain, on the other hand, was about one third of the number of non-Indigenous men rated as high by parole officers at intake. The number of Indigenous men rated as high on the community functioning domain by parole officers at intake was less than one third of the number rated as high by the calculated method, and the number of Indigenous men rated as high on the personal/emotional domain was more than double the number rated as high by the calculated domain ratings. Together, these results suggest that for the parole officers' ratings there is a reduced concordance between the number of indicators endorsed and the domain ratings.

The calculated domain ratings based on the proportion of indicators endorsed are more likely to approximate a normal distribution than the parole officers' intake ratings for non-Indigenous and Indigenous men. For non-Indigenous men, the low and moderate ratings capture over 50% of offenders in each of the domains using the calculated method, but this only occurred in two of the domains – the employment/education and associates domains – using the parole officer ratings. For Indigenous men, the parole officers' employment/education domain was the only domain that showed an approximately normal distribution of the rating levels, while the calculated domain ratings were close to a normal distribution for all domains except the substance use domain, in which the high need rating was the most prevalent.



Table 2

*Frequencies and percentages of parole officer intake ratings and calculated domain ratings: Federally sentenced men.*

Domain and Rating	Non-Indigenous Men			
	Parole Officer Intake Rating		Calculated Rating	
	<i>n</i>	%	<i>n</i>	%
<b>Employment/Education</b>				
Asset or No Need	3,417	22.1	1,262	8.2
Low	4,248	27.4	5,466	35.5
Moderate	6,828	44.1	5,681	35.9
High	994	6.4	2,991	19.4
<b>Marital/Family</b>				
Asset or No Need	9,013	58.2	2,001	12.9
Low	2,151	13.9	8,560	55.3
Moderate	2,915	18.8	4,379	28.3
High	1,408	9.1	539	3.5
<b>Associates</b>				
Asset or No Need	4,067	26.3	1,316	8.5
Low	1,752	11.3	5,013	32.4
Moderate	6,181	39.9	7,801	50.4
High	3,487	22.5	1,353	8.7
<b>Substance Abuse</b>				
No Need	4,576	29.6	2,903	18.8
Low	2,288	14.8	3,516	22.7
Moderate	3,228	20.8	4,862	31.4
High	5,395	34.8	4,186	27.1
<b>Community Functioning</b>				
Asset or No Need	10,598	68.4	3,014	19.5
Low	1,688	10.9	4,405	28.5
Moderate	2,470	16.0	4,217	27.3
High	731	4.7	3,842	24.8
<b>Personal/Emotional</b>				
No Need	2,976	19.2	462	3.0
Low	1,331	8.6	6,795	43.9
Moderate	5,162	33.3	6,219	40.2
High	6,018	38.9	2,007	13.0
<b>Attitude</b>				
Asset or No Need	2,884	18.6	1,709	11.0
Low	1,400	9.0	5,991	38.7
Moderate	5,741	37.1	5,029	32.5
High	5,462	35.3	2,757	17.8

Table 2 *Continued*

Domain and Rating	Indigenous Men			
	Parole Officer Intake Rating		Calculated Rating	
	<i>n</i>	%	<i>n</i>	%
<b>Employment/Education</b>				
Asset or No Need	423	9.1	140	3.0
Low	838	18.1	1,221	26.4
Moderate	2,764	59.6	1,826	39.5
High	615	13.3	1,442	31.2
<b>Marital/Family</b>				
Asset or No Need	1,762	38.0	201	4.3
Low	717	15.5	1,710	36.9
Moderate	1,395	30.1	2,245	48.4
High	766	16.5	484	10.4
<b>Associates</b>				
Asset or No Need	968	20.9	121	2.6
Low	451	9.7	976	21.0
Moderate	1,925	41.5	2,476	53.4
High	1,296	27.9	1,067	23.0
<b>Substance Abuse</b>				
No Need	385	8.3	110	2.4
Low	375	8.1	469	10.1
Moderate	1,077	23.2	1,589	34.3
High	2,803	60.4	2,469	53.3
<b>Community Functioning</b>				
Asset or No Need	2,495	53.8	356	7.7
Low	635	13.7	1,128	24.3
Moderate	1,134	24.4	1,445	31.2
High	376	8.1	1,710	36.9
<b>Personal/Emotional</b>				
No Need	424	9.1	42	0.9
Low	185	4.0	1,384	29.8
Moderate	1,511	32.6	2,232	48.1
High	2,520	54.3	981	21.2
<b>Attitude</b>				
Asset or No Need	882	19.0	334	7.2
Low	369	8.0	1,441	31.1
Moderate	1,765	38.0	1,664	35.9
High	1,624	35.0	1,200	25.9

Across all the domains for women offenders, results show that the number of offenders with no need according to the calculated rating – indicating that none of the indicators were endorsed – was lower than the number of offenders with asset or no need ratings according to intake parole officers. (See Table 3.) That is, intake parole officers rate the domain as having no need despite endorsing indicators within the domain.

Ratings of high need on the other hand were much more likely for the personal/emotional domain. Fifty percent of women were rated as high need on the personal/emotional domain at intake by parole officers, but only 16% of women had a rating of high need according to the calculated domain. The community functioning domain was less likely to have a rating of high according to parole officers' intake ratings. Less than 8% of women had a rating of high on this domain according to the parole officers' intake ratings, but over 45% had high ratings according to the calculated domains. As was observed for the men, these results suggest common discrepancies between the domain ratings and the number of indicators endorsed for women offenders.

Due to the greater number of indicators endorsed for federally sentenced women than men, the calculated domains ratings of high need were more likely to have the greatest proportion of women offenders. The marital/family, associates, personal/emotional, and attitude domains were approximately normally distributed while the employment/education, substance abuse, and community functioning domains had the greatest proportion of women in the calculated high need rating. Although the calculated domain ratings were less likely to approximate a normal distribution for federally sentenced women than men, more of the calculated domain ratings were approximately normally distributed than the parole officers' intake ratings. Only the parole officers' intake ratings on the employment/education domain approximated a normal distribution. Rather than seeing greater proportions of women in the high need category, the parole officers' intake ratings often had a greater proportion in the asset or no need category than a normal distribution would suggest. The marital/family, associates, substance abuse, community functioning, and attitude domains all had more women with asset or no need than low need according to parole officers' intake ratings. Parole officers' intake ratings on the substance abuse and personal/emotional domains also placed a majority of women in the high need category.

Table 3

*Frequencies and percentages of parole officer intake ratings and calculated domain ratings:  
Women offenders*

Domain and Rating	Parole Officer Intake Rating		Calculated Rating	
	<i>n</i>	%	<i>n</i>	%
Employment/Education				
Asset or No Need	185	15.5	69	5.8
Low	250	20.9	286	24.0
Moderate	592	49.5	411	34.5
High	168	14.1	427	35.8
Marital/Family				
Asset or No Need	291	24.4	54	4.5
Low	180	15.1	323	27.1
Moderate	490	41.0	569	47.7
High	234	19.6	246	20.6
Associates				
Asset or No Need	195	16.3	47	3.9
Low	145	12.1	259	21.7
Moderate	434	36.3	592	49.7
High	421	35.2	294	24.7
Substance Abuse				
No Need	226	18.9	177	14.8
Low	60	5.0	118	9.9
Moderate	141	11.8	325	27.2
High	768	64.3	573	48.0
Community Functioning				
Asset or No Need	459	38.4	69	5.8
Low	253	21.2	242	20.4
Moderate	390	32.6	334	28.1
High	93	7.8	544	45.8
Personal/Emotional				
No Need	59	4.9	18	1.5
Low	88	7.4	413	34.6
Moderate	441	36.9	574	48.0
High	607	50.8	190	15.9
Attitude				
Asset or No Need	440	36.8	298	25.0
Low	184	15.4	514	43.2
Moderate	381	31.9	307	25.8
High	190	15.9	71	6.0

## **Associations between Revocations and Parole Officer Ratings and Calculated Domain Ratings**

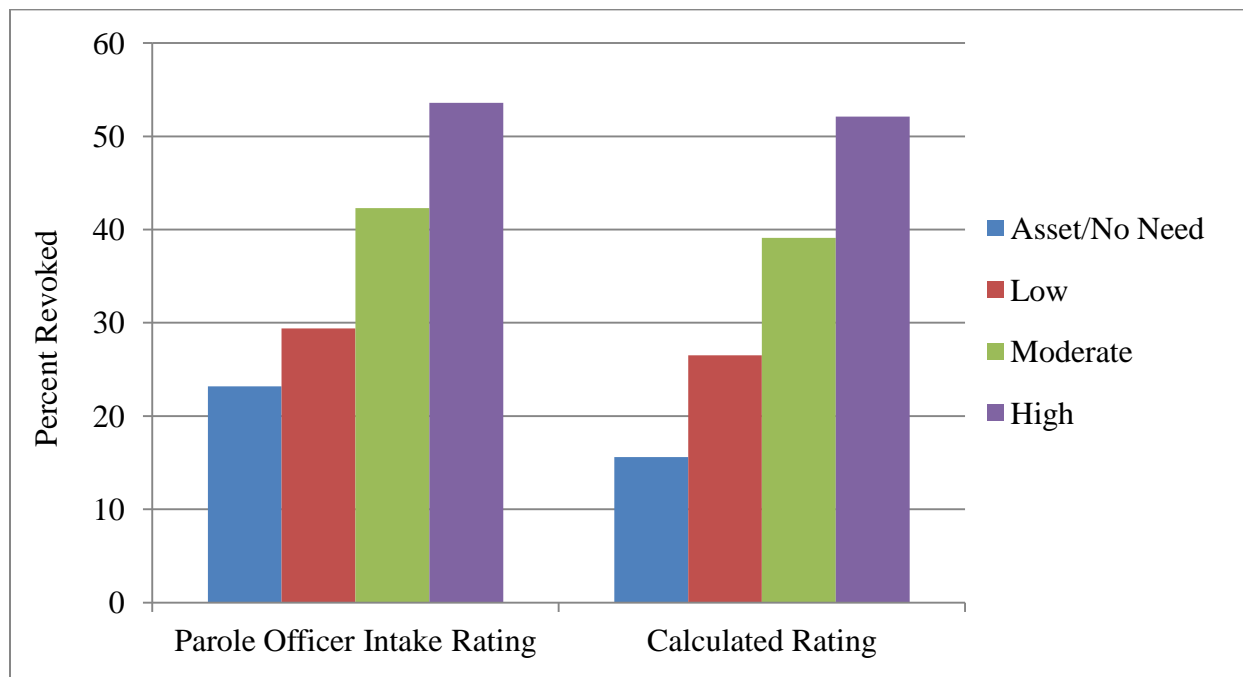
### **Non-Indigenous men**

Figures 1 through 7 show the percentages of revocations of offenders across domains for the initial parole officers' ratings and calculated ratings for non-Indigenous men. Across all the domains, the calculated ratings showed greater distinction among the ratings as they relate to revocations. That is, calculated ratings of asset or no need were associated with lower rates of revocations and calculated ratings of high need were associated with higher ratings of revocations compared to parole officers' intake ratings. For example, 52% of offenders with calculated ratings of high need and 16% with calculated ratings of asset or no need in the employment/education domain (Figure 1) had revocations – a difference of 36%. Parole officer ratings demonstrated a lower difference 31%. The patterns were similar for the associates (Figure 3), substance abuse (Figure 4), and community functioning (Figure 5) domains. Greater advantages of the calculated domain ratings over the parole officers' ratings were observed for the personal/emotional (Figure 6) and attitude (Figure 7) domains. In these domains, we again observe greater distinctions between high and asset or no need ratings for the calculated ratings compared to the parole officers' ratings. However, parole officers' intake ratings display a problematic pattern in that percentages of revocations in the asset or no need category were higher than the low need category. This problem was not observed in the calculated domain ratings. Finally, the parole officers' ratings on the marital/family domain (Figure 2) showed very little differentiation in percentages of revocations, and the high need group actually had fewer revocations than both the low and moderate need groups. Using the calculated domain ratings, the percentages of revocations increase as expected. As the need ratings increase from asset or no need to high need, the rates of revocations increase.

The improvement of the calculated ratings over the parole officers' ratings for non-Indigenous men was further supported by survival analyses which take time at risk into consideration. The survival analysis results are presented in Table B1 in Appendix B.



*Figure 1. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Employment domain*



*Figure 2. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Marital/Family domain*

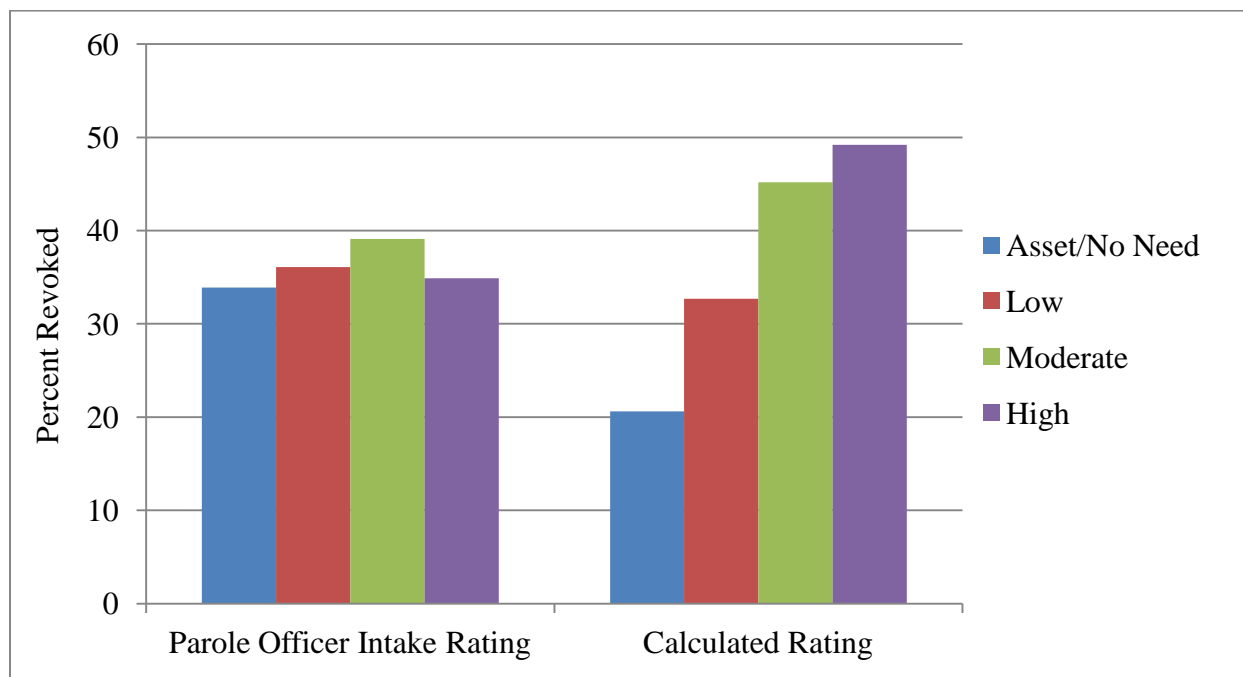


Figure 3. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Associates domain

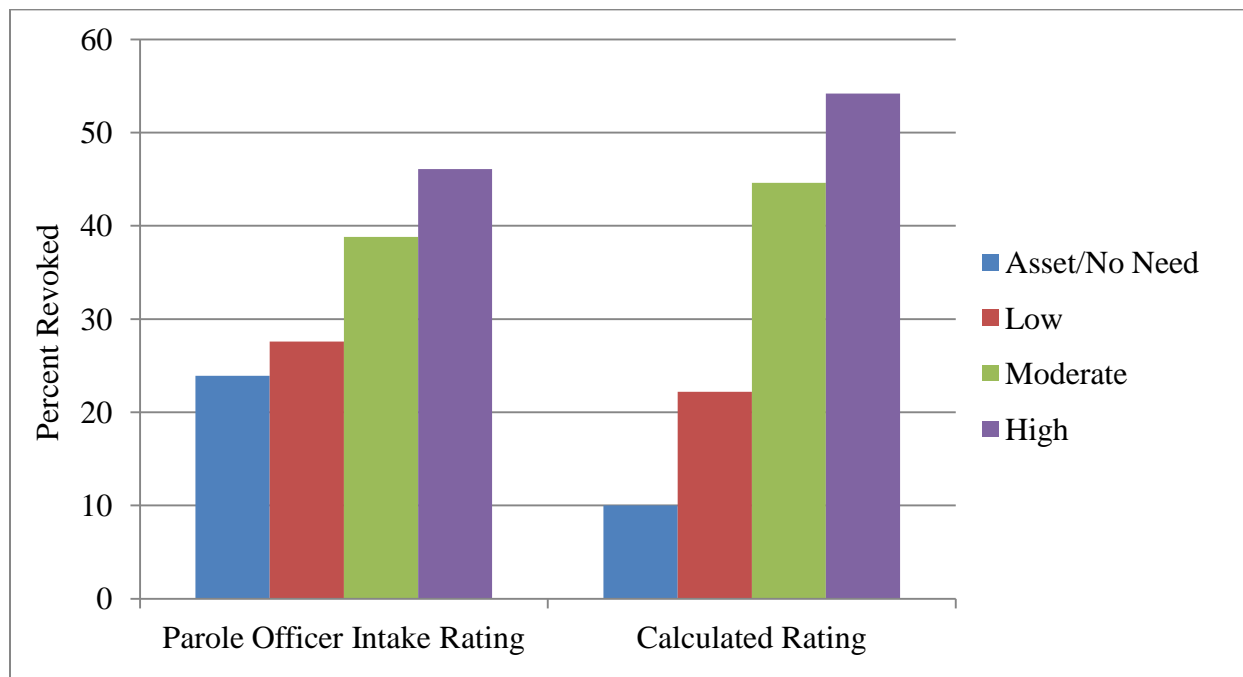


Figure 4. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Substance use domain

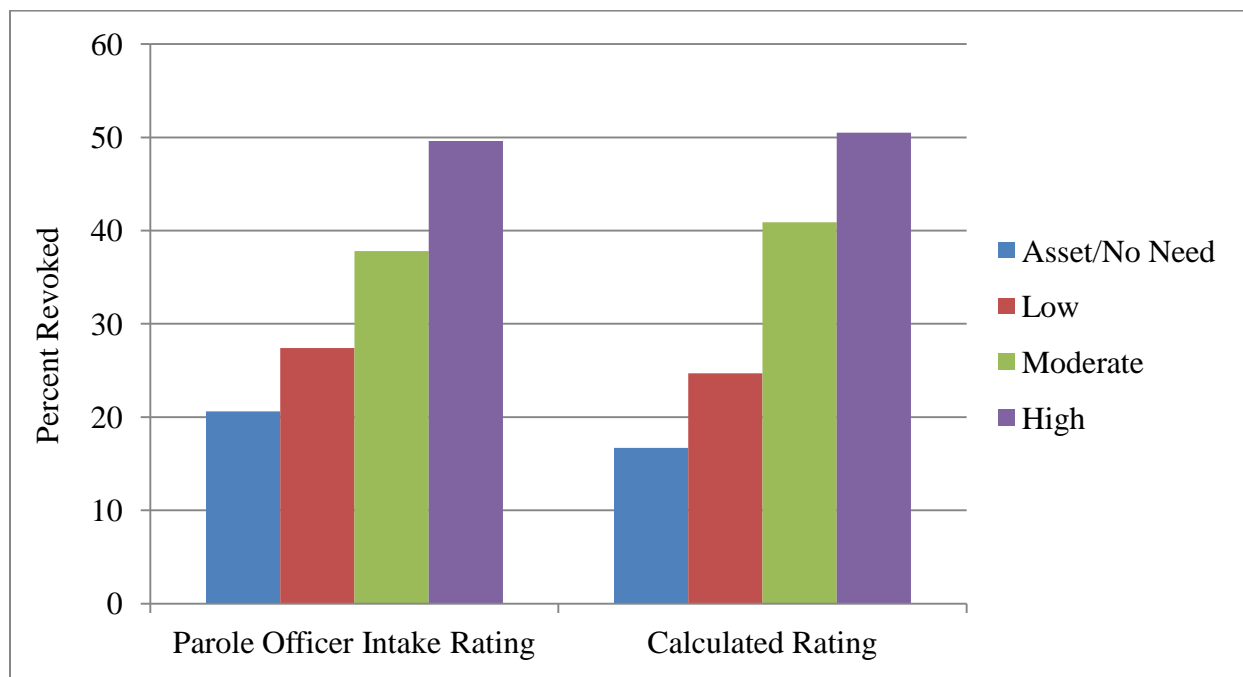


Figure 5. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Community Functioning domain

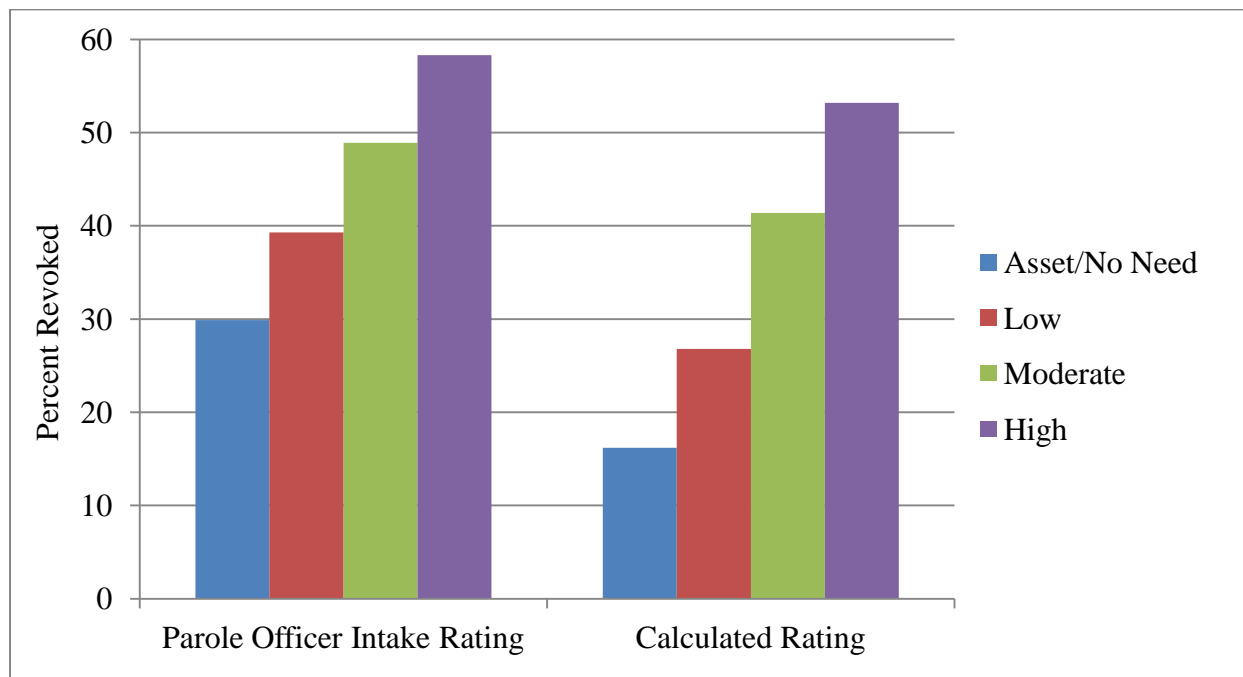


Figure 6. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Personal/Emotional domain

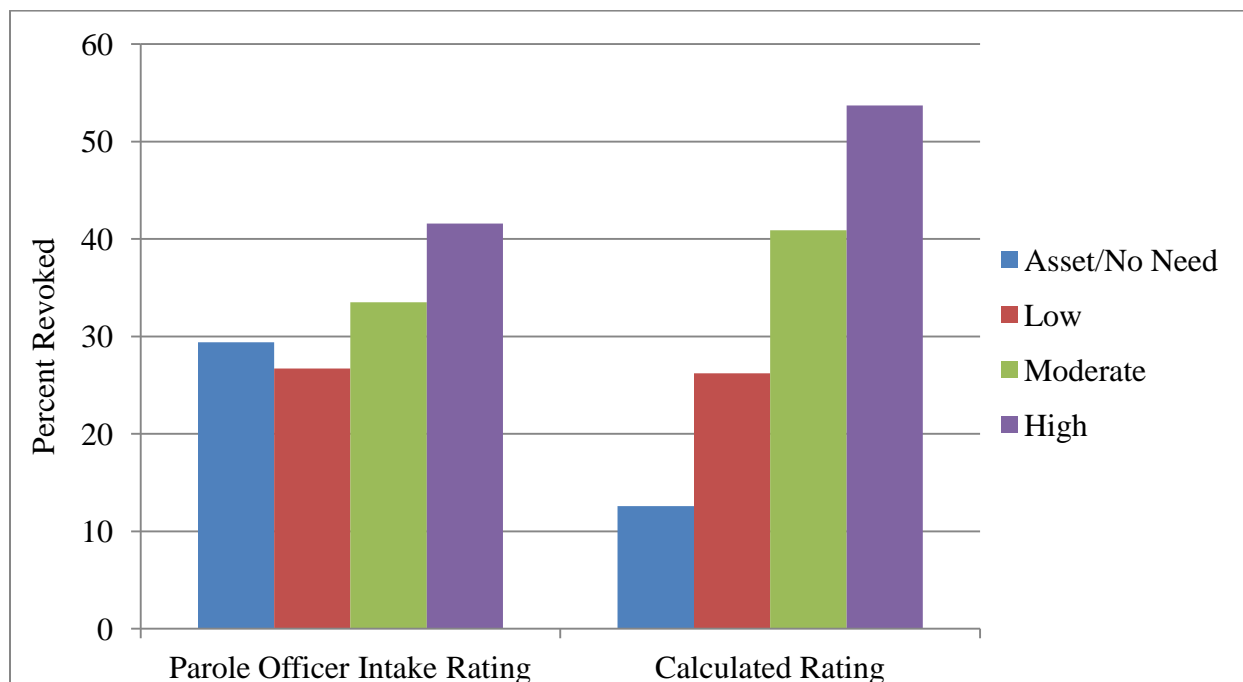
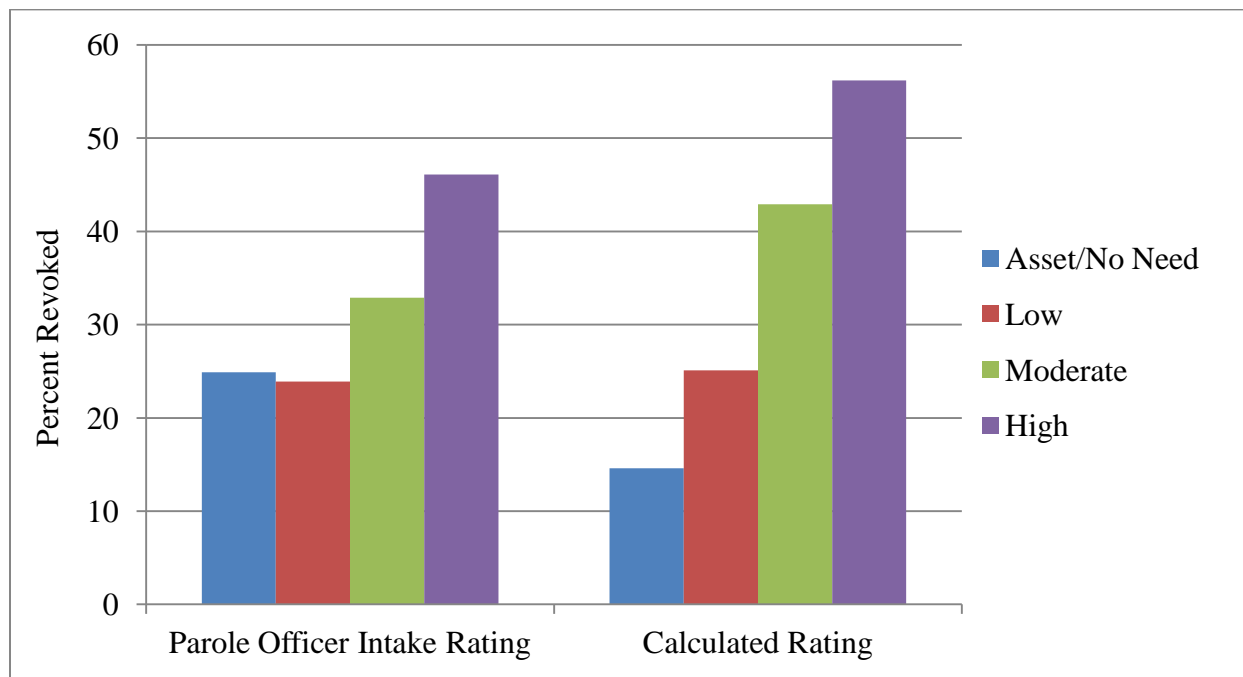


Figure 7. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for non-Indigenous men: Attitude domain



### Indigenous men

As was observed for non-Indigenous men, calculated domain ratings improved the prediction of revocations over that provided by the parole officers' ratings. As displayed in Figures 8 through 14, the calculated domain ratings showed greater differentiation among the four rating levels for all seven domains. Furthermore, the parole officers' ratings on the associates and attitude domains showed higher proportions of revocations in the asset or no need category compared to the low need category. The parole officer ratings on the marital/family domain also showed counter-intuitive results with greater proportions of revocations for offenders rated low and moderate than those rated high need. In comparison, each calculated domain showed an incremental increase in the proportion of revocations as the rating increased.

The observed frequencies of revocations and comparisons between the parole officers' ratings and calculated domain ratings for Indigenous men were supported by survival analyses which take time at risk into consideration. The survival analysis results are presented in Table B2 in Appendix B.

Figure 8. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Employment domain

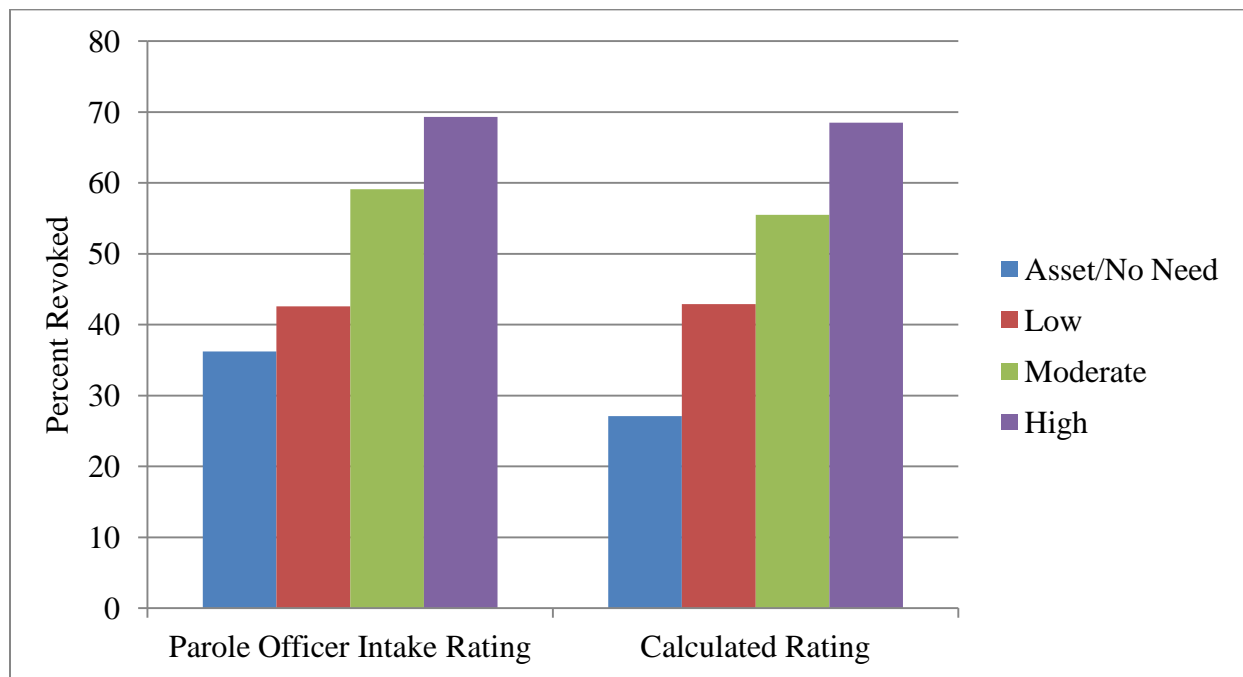


Figure 9. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Marital/Family domain

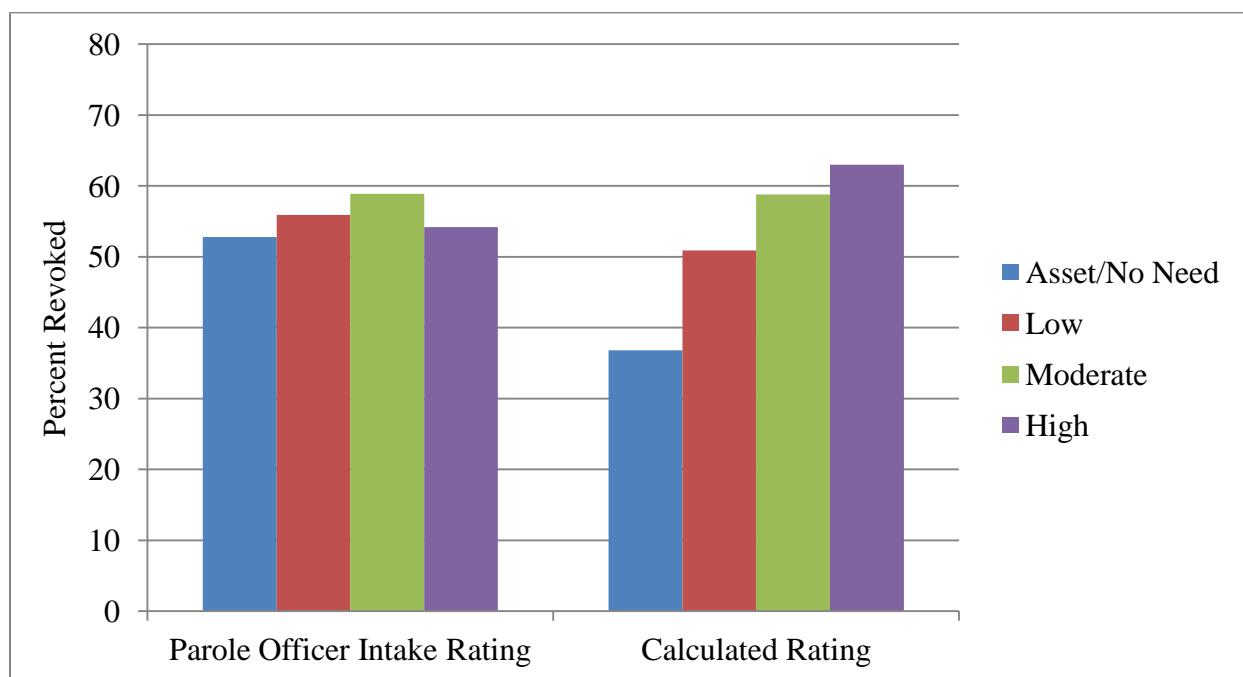


Figure 10. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Associates domain

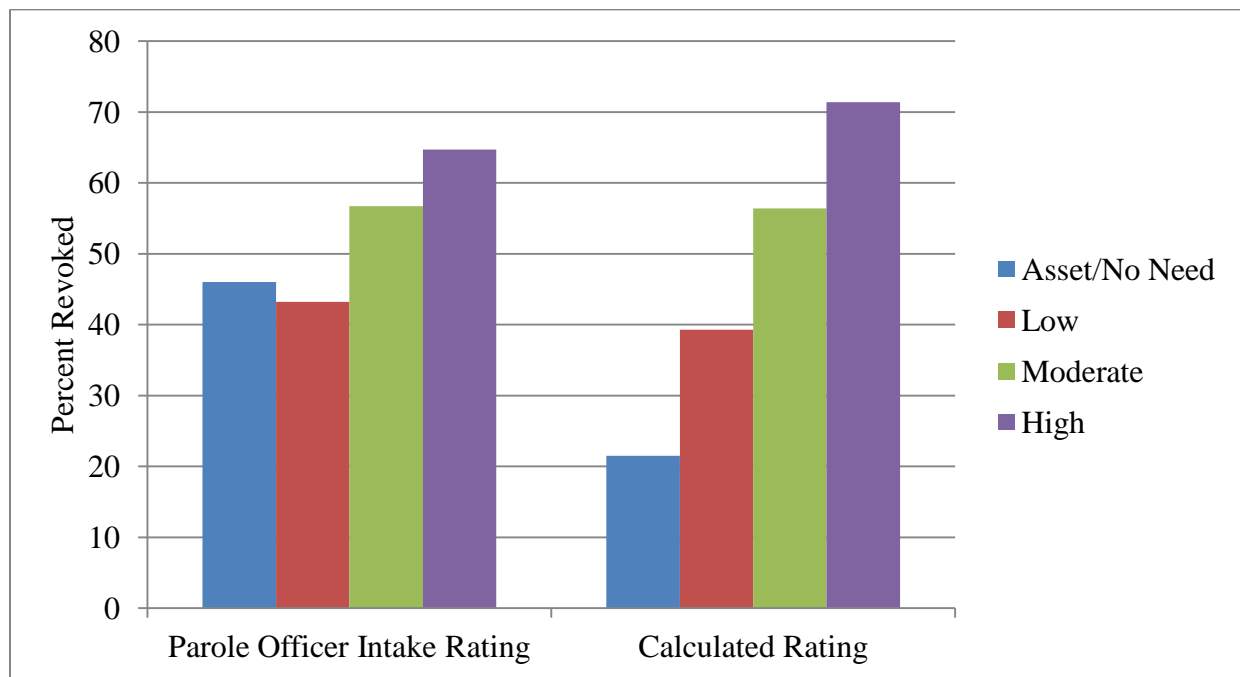


Figure 11. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Substance Abuse domain

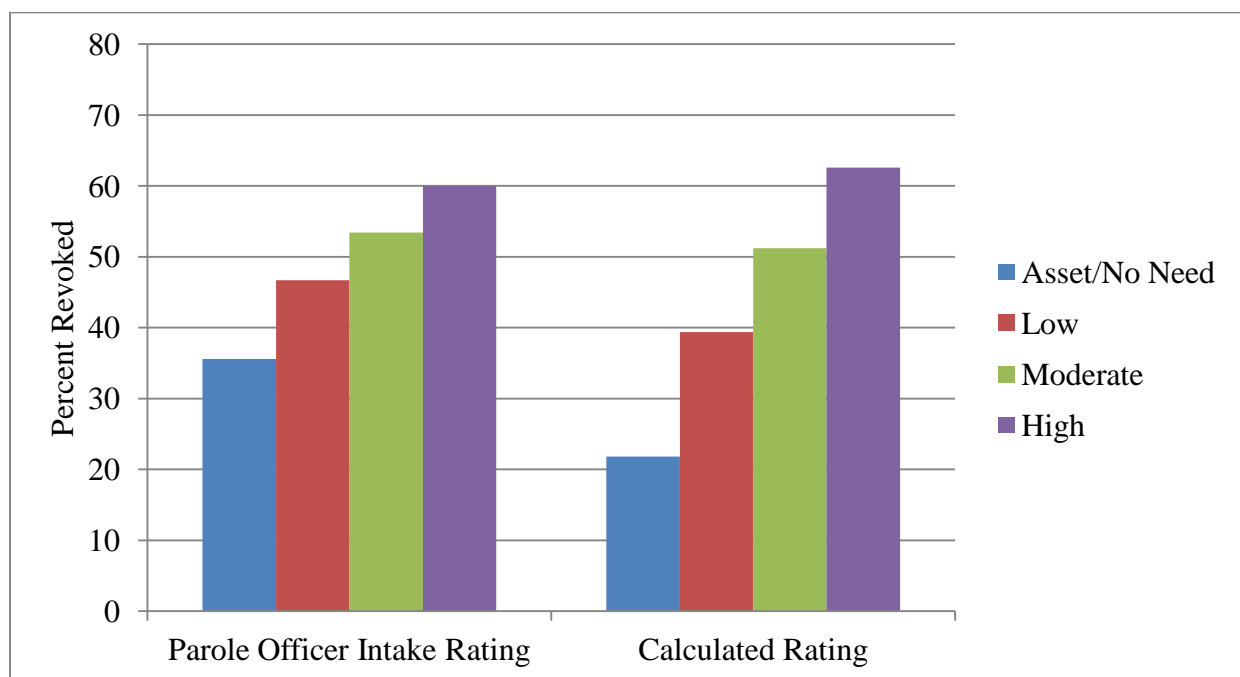


Figure 12. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Community Functioning domain

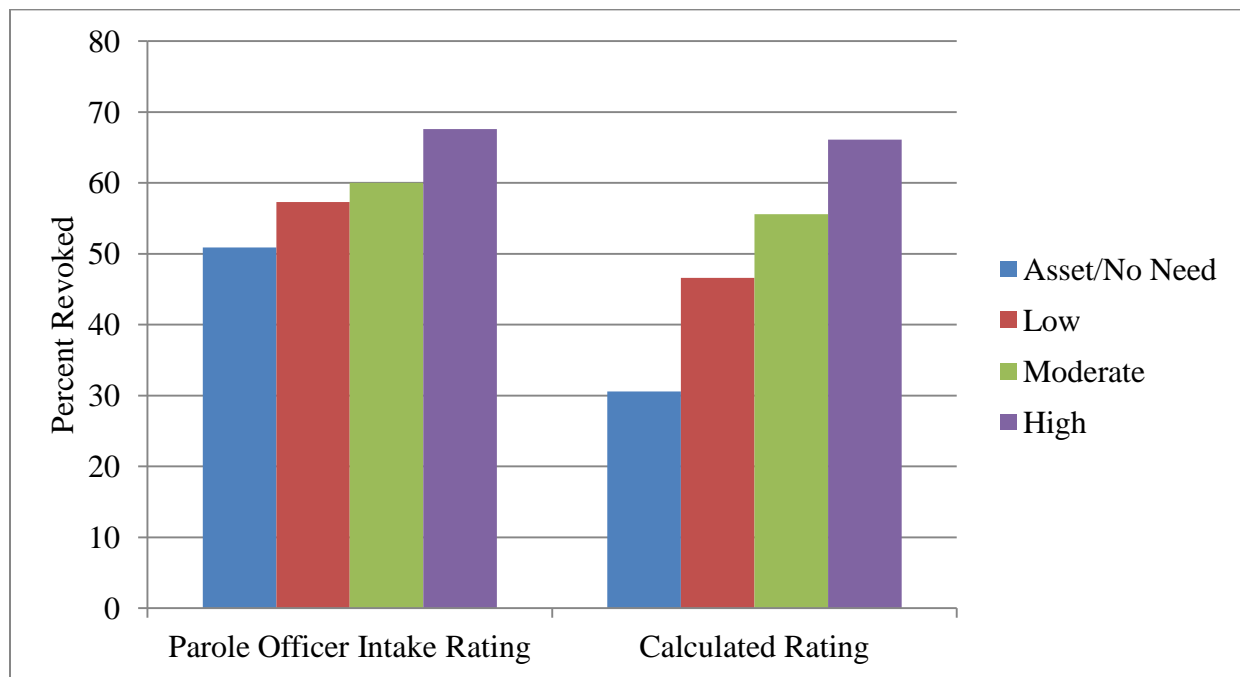


Figure 13. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Personal/Emotional domain

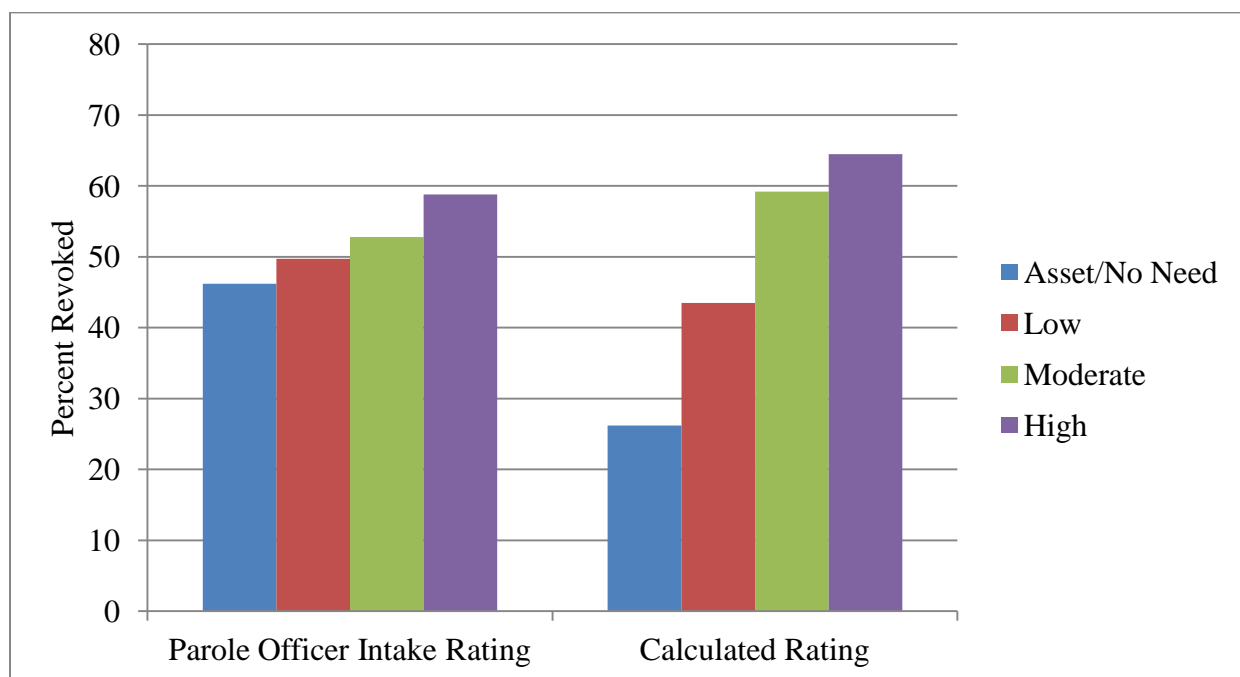
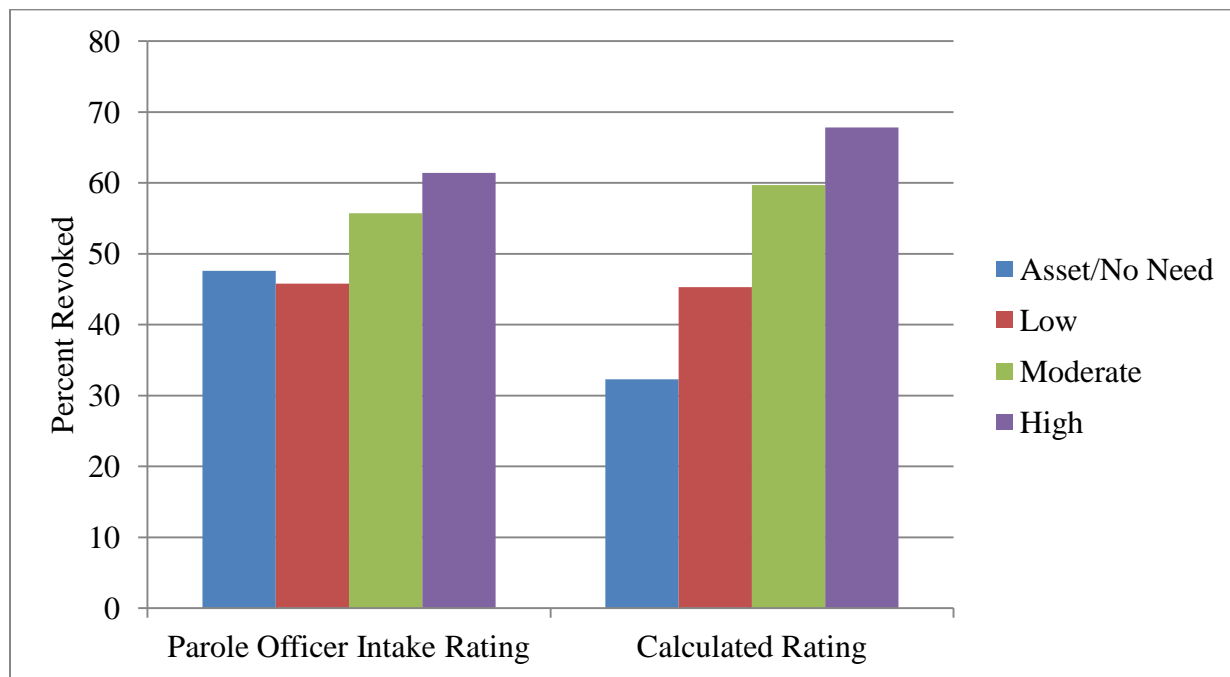


Figure 14. Percentage of revocations at different ratings of need based on parole officer ratings and calculated ratings for Indigenous men: Attitude domain



### Women

The association between parole officers' ratings and calculated domain ratings as they relate to revocations for federally sentenced women are displayed in Figures 15 through 21. As was the case for men, the calculated domains provided a greater distinction among the ratings, and the incremental increases in revocations from the asset or no need rating through to the high need rating were more consistent for the calculated ratings than the parole officers' intake ratings.

For federally sentenced women, the employment/education and substance abuse domains were the only two domains in which the parole officers' ratings demonstrated an incremental increase in percentage of revocations as need ratings increased. In each of the marital/family, associates, community functioning, personal/emotional and attitude domains, a lower rating, usually the asset or no need rating, was associated with a higher percentage of revocations than one of the higher need ratings. For example, women with a parole officers' rating of moderate need in the marital/family domain had the highest percentage of revocations, and women rated as no need in the personal/emotional domain were more likely to have a revocation than those rated



as low or moderate need. However, across the calculated domain ratings, the percentage of revocations followed the expected increase as the ratings increased. For each of the calculated domains, women with ratings of no need had the lowest percentages of revocations followed by calculated ratings of low need, then moderate need, and women with high ratings had the highest percentage of revocations.

The observed frequencies of revocations and comparisons between parole officers' intake ratings and calculated domain ratings for women are supported by survival analyses which take time at risk into consideration. The survival analysis results are presented in Table B3 in Appendix B.

*Figure 15.* Percentage of revocations based on parole officer ratings and calculated ratings for Women: Employment/Education domain

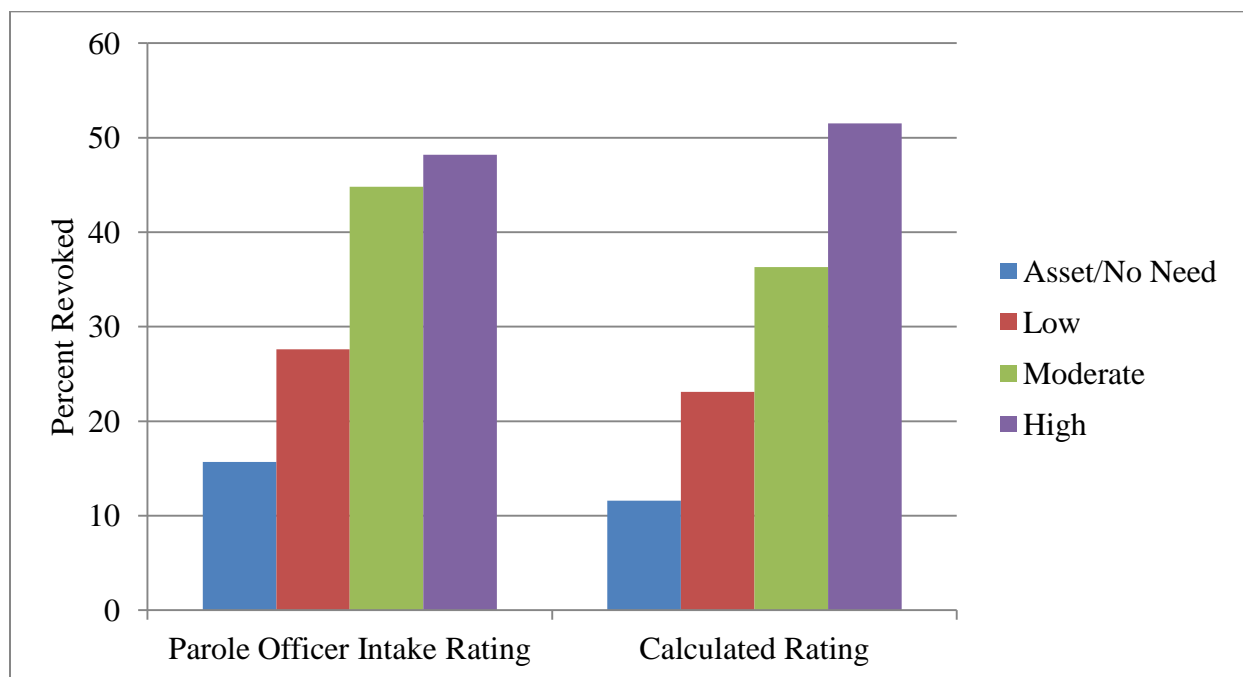


Figure 16. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Marital/Family domain

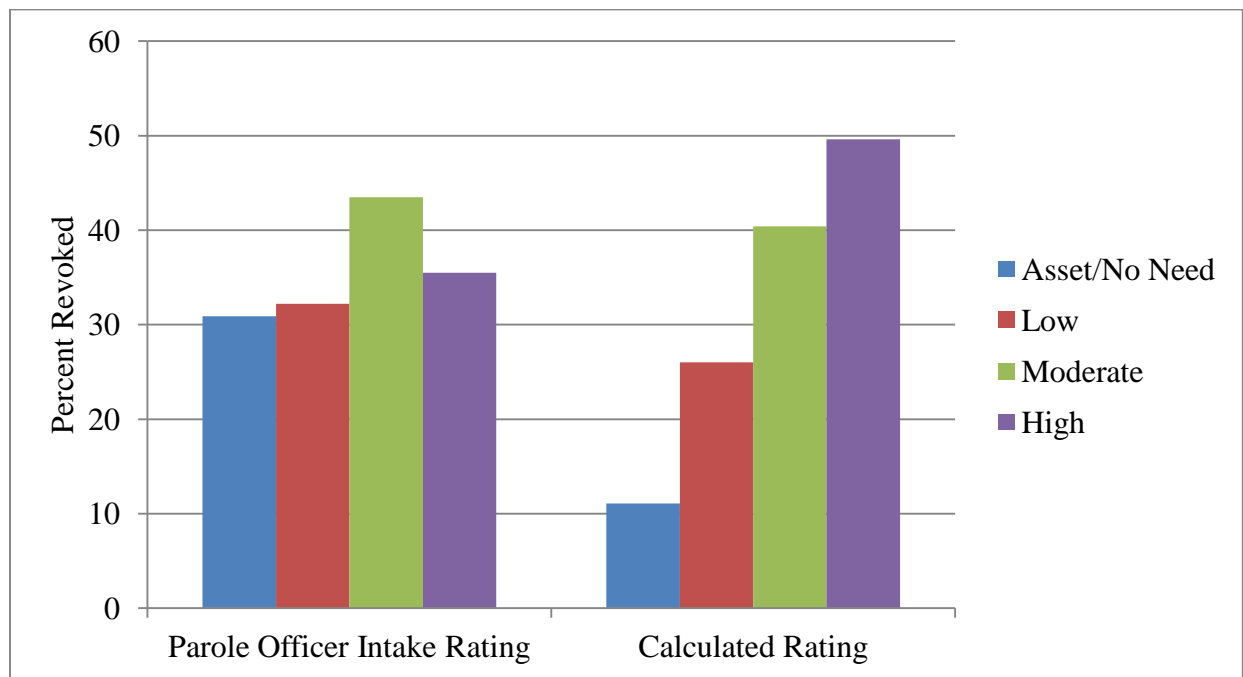


Figure 17. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Associates domain

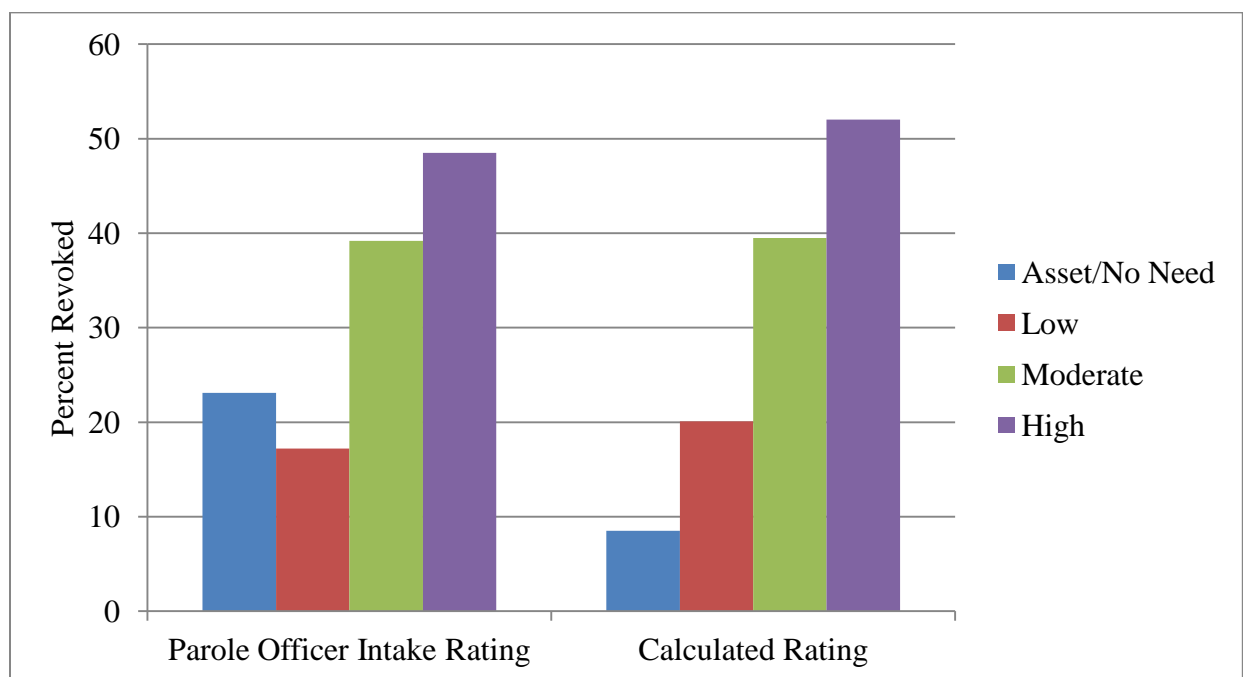


Figure 18. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Substance Abuse domain

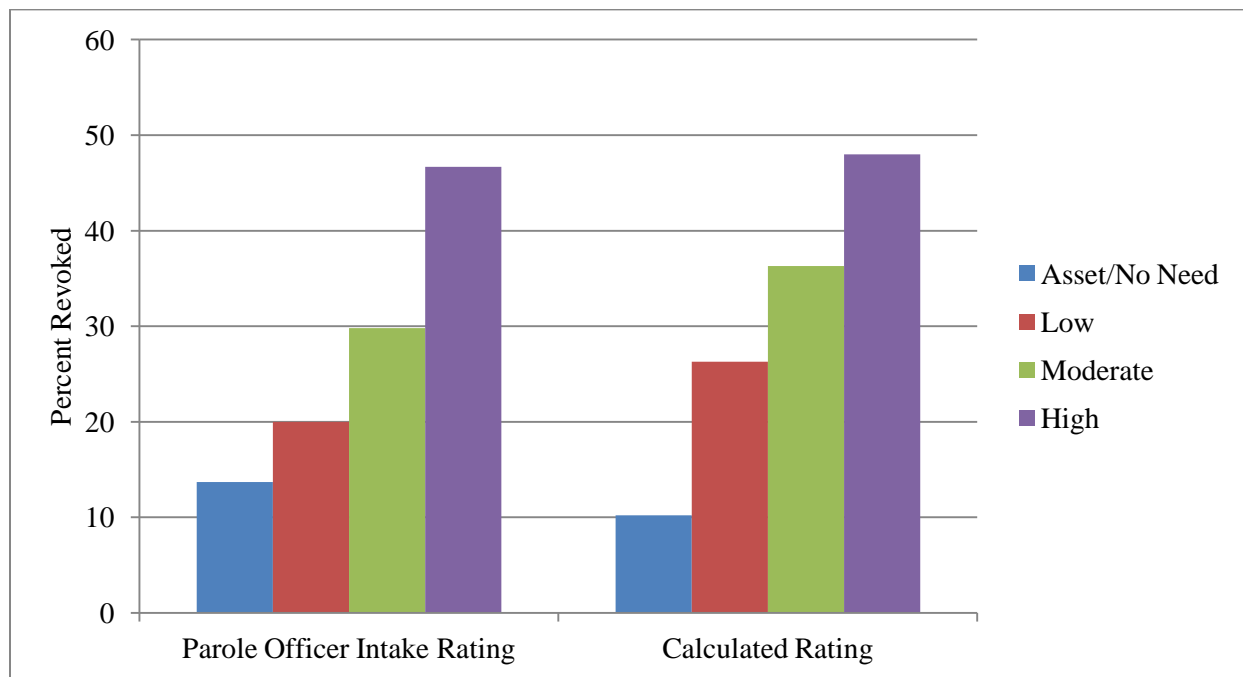


Figure 19. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Community Functioning domain

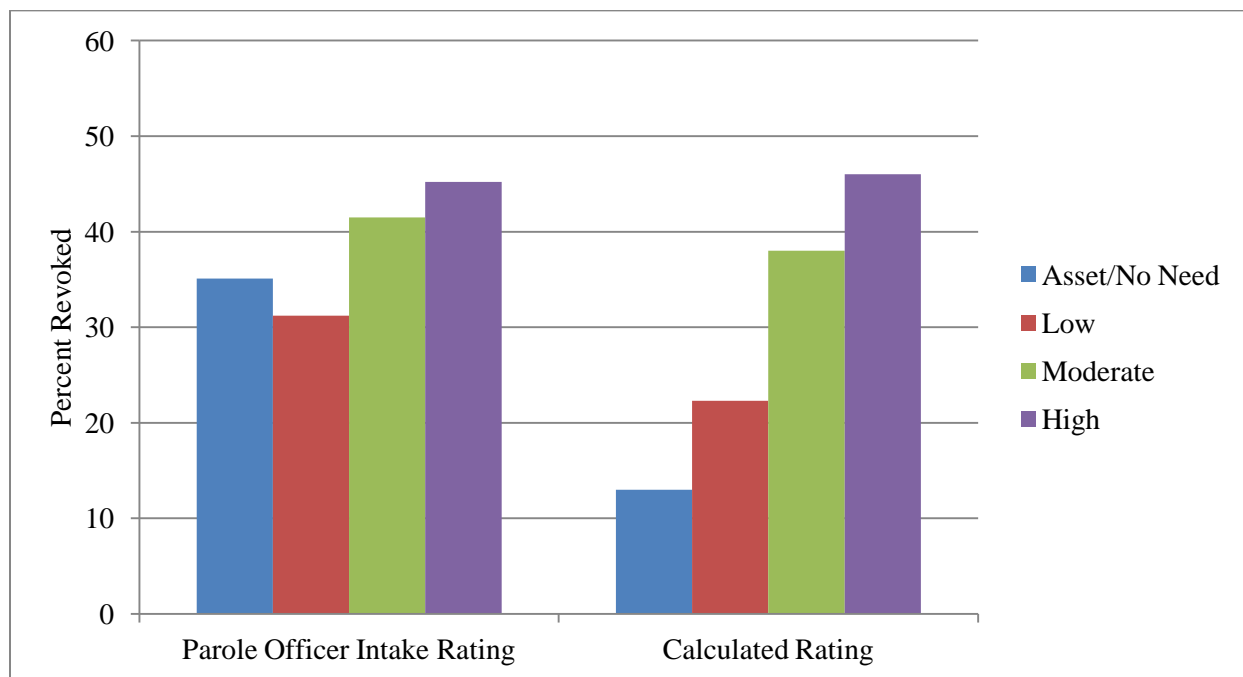


Figure 20. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Personal/Emotional domain

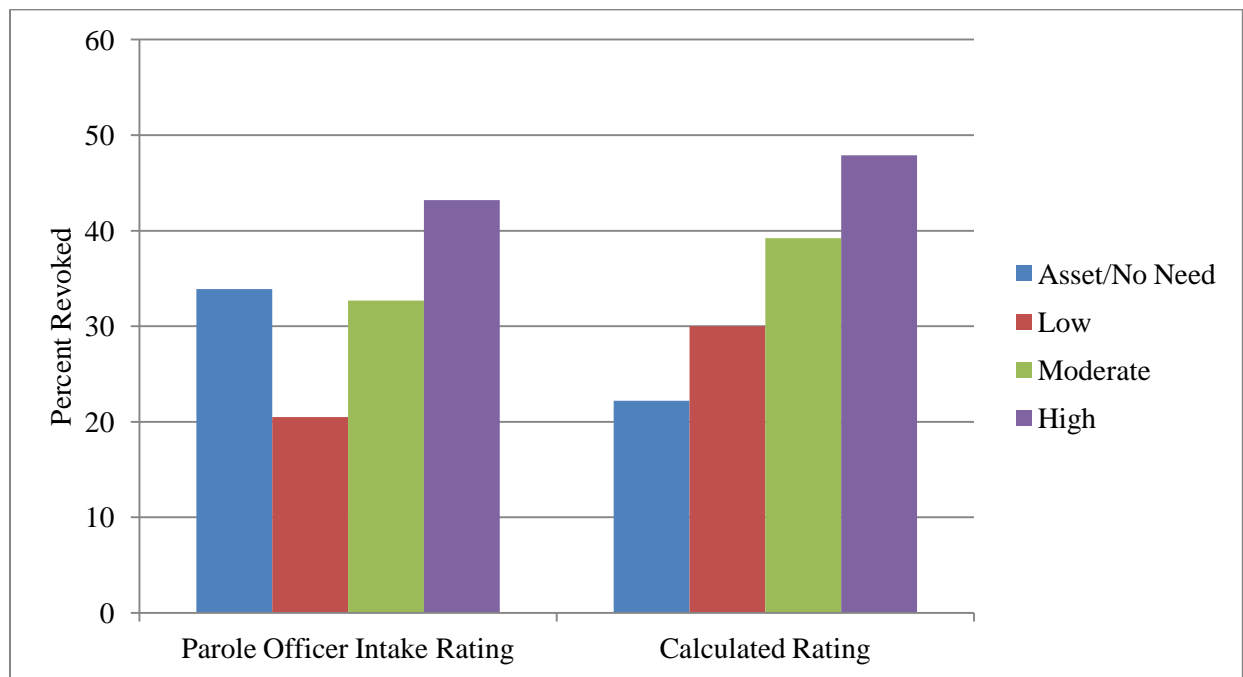
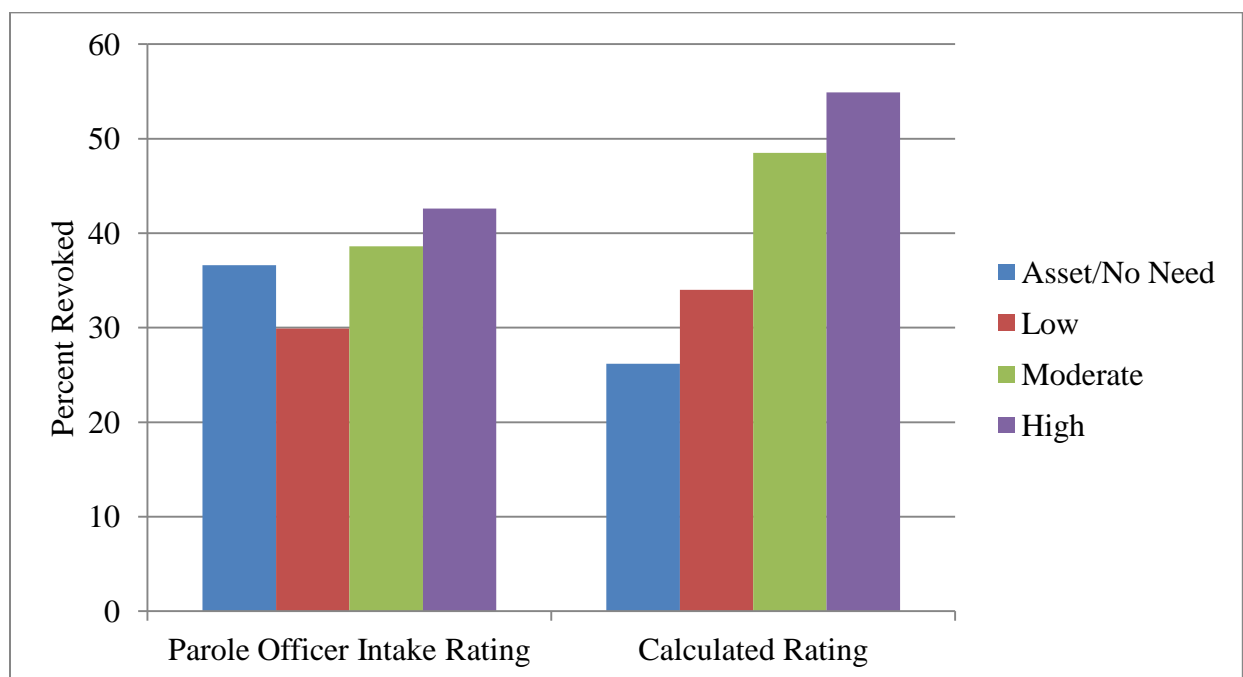


Figure 21. Percentage of revocations based on parole officer ratings and calculated ratings for Women: Attitude domain



## **Association between Revocations and Parole Officer Overall Intake Ratings and Calculated Overall Need Ratings**

Associations of revocations with overall need ratings completed by parole officers were compared to two calculated ratings. The first calculated overall need rating method was based on the percentages of indicators endorsed in each domain averaged across the domains. The second method used a weighted mean. Based on the results of survival analyses presented in Appendix A3 the percentages of indicators endorsed in each domain were multiplied by the domain's hazard ratio then averaged across domains. The cohort was randomly split into two halves – a development sample and a test sample. Figures 22 through 25 and Tables 4 and 5 present the percentages of revocations for the ratings of offenders in the test sample. Appendix C provides survival analyses to account for time at risk. Reassessing the outcomes on the test sample avoided using the weighted mean method to predict the same outcomes that contributed to its development and the split-half reliability could be examined in comparison to the other overall ratings.

### **Non-Indigenous men**

Figure 22 displays the percentages of revocations for non-Indigenous men on the parole officers' overall need ratings, the calculated mean overall ratings, and the weighted mean overall ratings. All three methods showed the desired increase in percentage of revocations as the ratings increased. Non-Indigenous federally sentenced men with ratings of low need were least likely, and those with a high rating were most likely to have a revocation. However, the two calculated methods showed greater differentiation among the ratings with the mean method showing the greatest differentiation. That is, the offenders with ratings of low need on the mean method were less likely to have a revocation than those rated as low on the intake ratings, and those with high ratings on the mean method were more likely to have a revocation than those rated as high on the intake ratings. This pattern of results demonstrates that the mean method had the greatest predictive validity of the three methods. The weightings of the domains developed specifically for non-Indigenous men did not provide an advantage over simply weighting each domain

---

<sup>3</sup> The survival analysis produced hazard ratios that were used in this study as means of determining the relative impact of each domain on outcome when all domains were considered together. For example, the substance abuse domain for women produced a hazard ratio of 3.0, meaning that women with the all indicators endorsed in the domain were three time more likely to experience a revocation than women who had no indicators endorsed.

equally. (See Table 4.)

*Figure 22. Revocations associated with the three methods of overall need ratings for the test sample: Non-Indigenous men (N = 7,714)*

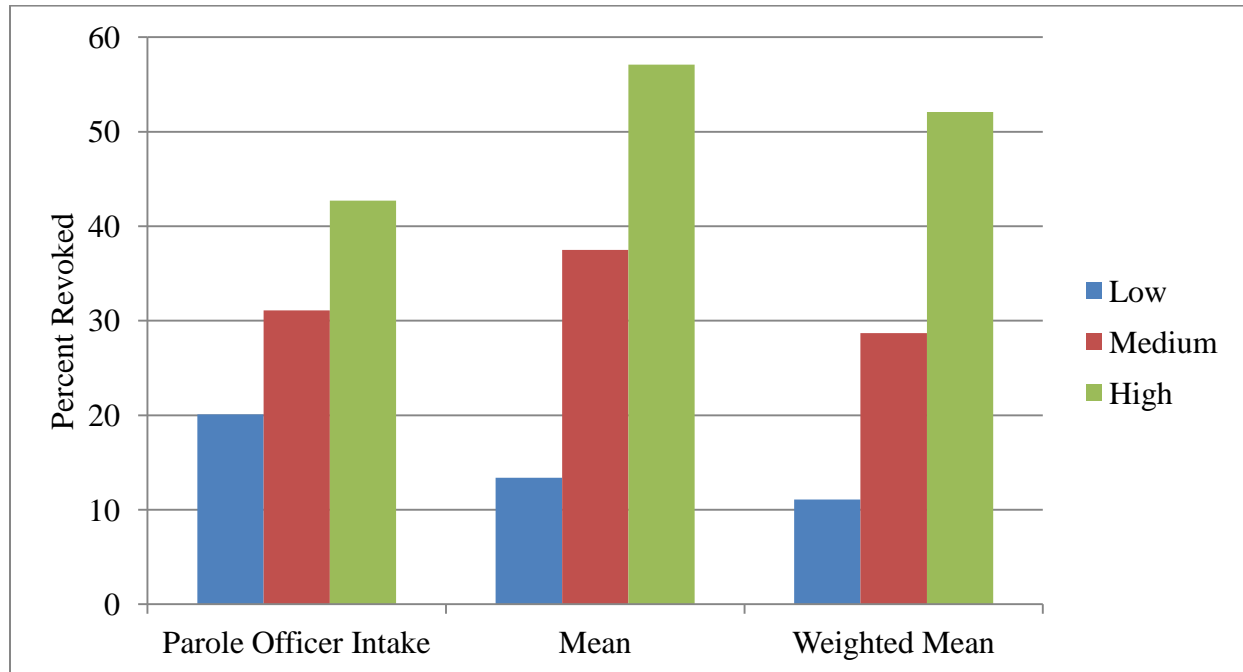


Table 4

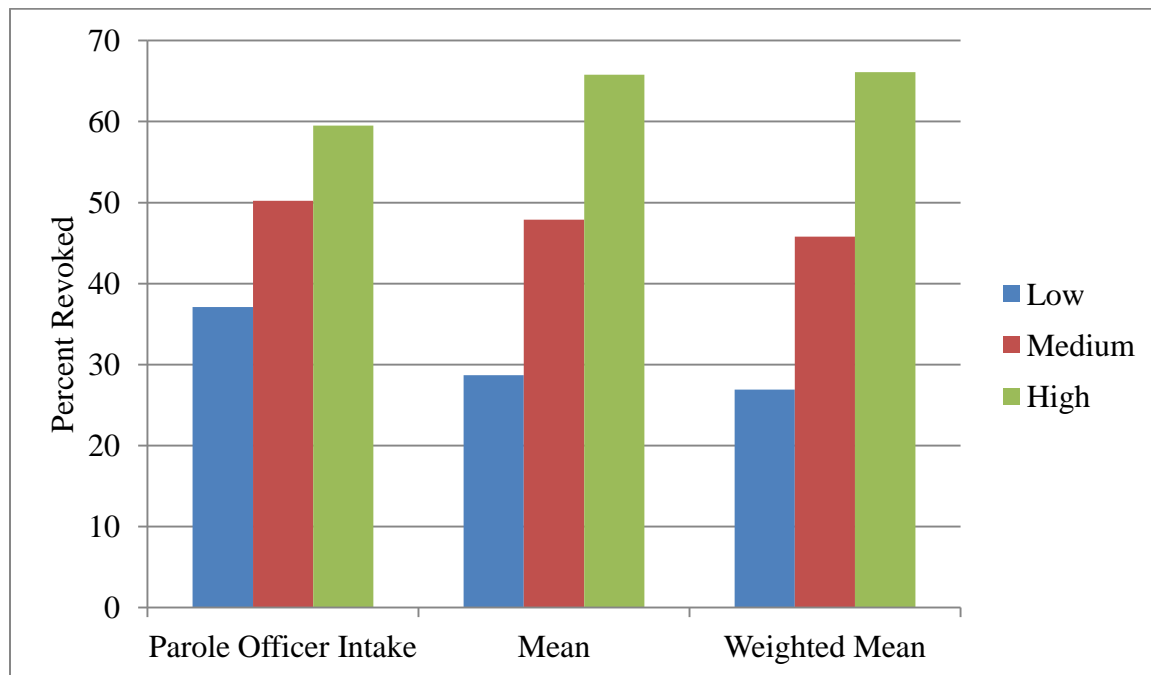
*Revocations associated within the three overall need rating methods for the test sample: Men*

	Non-Indigenous Men (N = 7,714)			Indigenous Men (N = 2,319)		
	% Revoked	N Revoked	Total N	% Revoked	N Revoked	Total N
Parole Officer Intake Rating						
Low	20.1	144	718	37.1	39	105
Medium	31.1	1,098	3,531	50.2	482	961
High	42.7	1,478	3,465	59.5	745	1,253
Calculated Mean Rating						
Low	13.4	307	2,286	28.7	68	237
Medium	37.5	1,313	3,503	47.9	462	964
High	57.1	1,100	1,925	65.8	736	1,118
Calculated Weighted Mean Rating						
Low	11.1	179	1,619	26.9	56	208
Medium	28.7	776	2,707	45.8	419	914
High	52.1	1,765	3,388	66.1	791	1,197

### Indigenous men

Figure 23 shows the percentages of revocations for the ratings on the three methods for Indigenous men. Each rating method shows the desired increase in revocations from low to medium to high ratings. However, the mean method and weighted mean method show greater differentiation in the percentages of revocations across the ratings than the parole officer intake ratings. The predictive validity of the two calculated overall need ratings were similar and showed greater predictive power than the parole officers' ratings for Indigenous men. Interestingly, the weightings developed specifically for Indigenous men did not provide an advantage over simply weighting each domain equally. (See Table 4.)

*Figure 23.* Revocations associated with the three methods of overall need ratings for the test sample: Indigenous men. (N = 2,319)



### Women

Results of the overall need ratings using the three methods for federally sentenced women are displayed in Figure 24 and 25. Figure 24 applied the cut off scores that were used for federally sentenced men. As observed for the men, each of the three methods of rating of overall need showed the expected increase in percentage of revocations as ratings increased from low to medium to high. The two calculated overall ratings performed similarly and both had greater differentiation among the ratings than the intake ratings. Therefore, the calculated ratings of

overall need had greater predictive validity for federally sentenced women than the intake rating, and applying weights specifically developed for women offenders to the domains did not have an advantage over weighting the domains equally (See Table 5).

Using the adjusted cut-off scores that reduced the proportion of women in the high risk group further improved the distribution and predictive power of both of the calculated methods over the parole officer ratings (Figure 25) with differences between the outcomes for women rated low risk and those rated high risk increasing in both cases.

*Figure 24.* Revocations associated with the three methods of overall need ratings for the test sample using the unadjusted calculated scores: Women ( $N = 575$ )

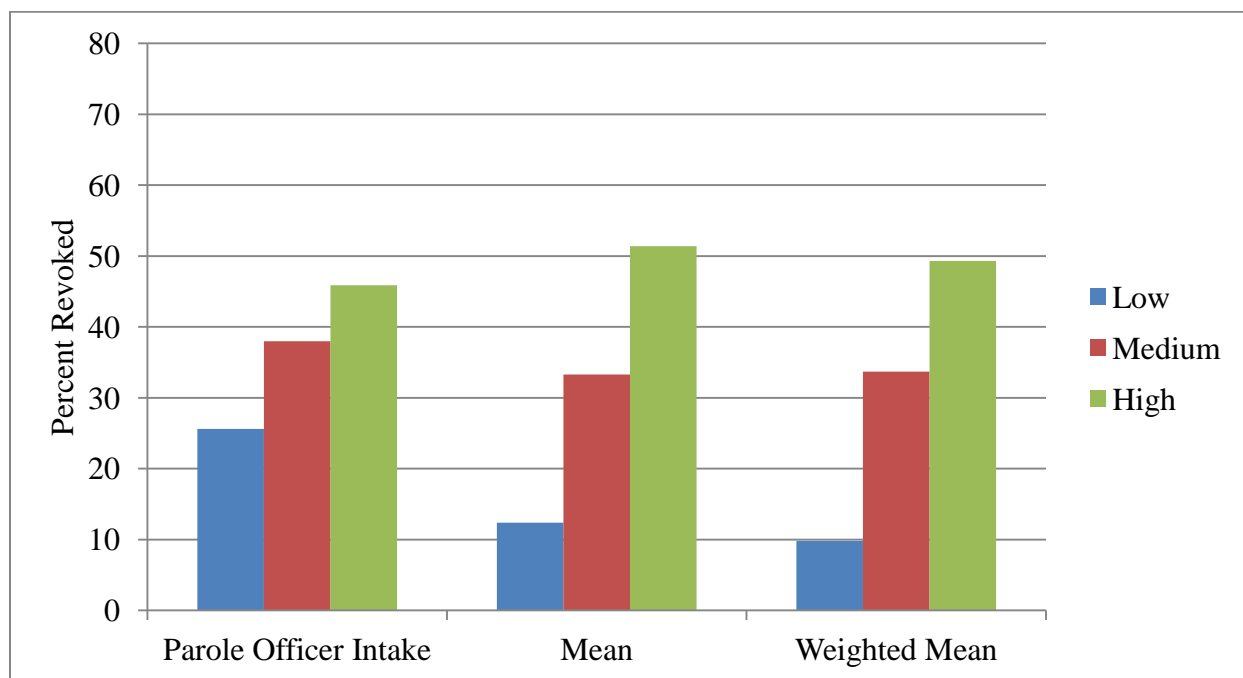




Figure 25. Revocations associated with the three methods of overall need ratings for the test sample using the adjusted calculated scores: Women ( $N = 575$ )

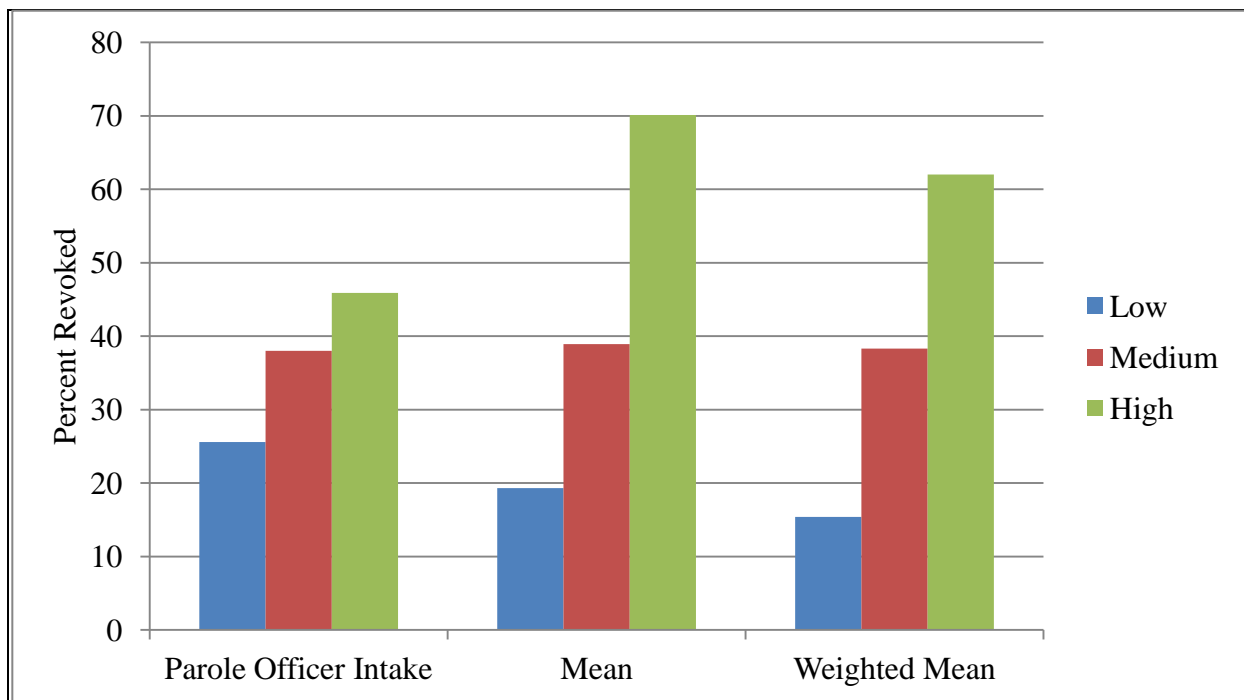


Table 5

*Revocations by overall need rating methods for the test sample: Women (N = 575)*

	% Revoked	N Revoked	Total N
Parole Officer Intake Rating			
Low	25.6	31	121
Medium	38.0	112	295
High	45.9	73	159
Calculated Mean Rating (unadjusted)			
Low	12.4	12	97
Medium	33.3	77	231
High	51.4	127	247
Calculated Weighted Mean Rating (unadjusted)			
Low	9.8	10	102
Medium	33.7	59	175
High	49.3	147	298
Calculated Mean Rating (adjusted)			
Low	19.3	31	161
Medium	38.9	131	337
High	70.1	54	77
Calculated Weighted Mean Rating (adjusted)			
Low	15.4	22	143
Medium	38.3	119	311
High	62.0	75	121

### **Split-half Reliability of Intake and Calculated Overall Need Ratings**

Split-half reliability analyses were conducted to determine whether results for the weighted mean overall need ratings apply to a new set of offenders. The comparisons were also completed for the parole officers' intake and mean overall need ratings. Smaller groups – Indigenous men, and women – were expected to show greater disparity between the development and test samples. For the weighted mean overall need ratings for all groups, the strength of the results was expected to be lower in the test sample than the development sample.

#### **Non-Indigenous men**

Results for non-Indigenous men presented in Table 6 show that the percentages of revocations at the different need rating levels were remarkably similar between the development and test samples. The percentages of revocations in the development sample were identical to the test sample across the weighted mean overall need ratings. The chi-squared statistics and hazard ratios were also similar between development and test samples for all three overall need rating

methods. Although there was an evident drop in chi-squared statistics and hazard ratios from the development sample to the test sample for the weighted mean overall need rating method, the results were consistent enough to conclude that this was a reliable method. The greatest difference between development and test sample results was observed for the parole officers' intake ratings. Both of the calculated methods appeared to be more reliable, having less variability within each rating level than the parole officers' intake ratings.

Furthermore, the Harrell's C statistics show that the mean and weighted mean methods of rating overall need had equally strong predictive power and these calculated methods had a greater predictive power than the parole officers' intake ratings.

Table 6

*Split-half reliability of associations between revocations and parole officer intake ratings and calculated overall need ratings: Non-Indigenous men*

Rating	Development Sample			Test Sample		
	Percent of Revocations	$\chi^2$	Hazard ratio	Percent of Revocations	$\chi^2$	Hazard ratio
Parole officer intake ratings of overall need						
Low <sup>a</sup>	21 (258/742)	257.9 <sup>b</sup>	.59 <sup>c</sup>	20 (144/718)	304.7 <sup>b</sup>	.59 <sup>c</sup>
Medium	32 (1,159/3,614)	59.6	1.9	31 (1,098/3,531)	57.9	2.0
High	42 (1,424/3,417)	176.3	3.1	43 (1,478/3,465)	189.2	3.3
Mean ratings of overall need						
Low <sup>a</sup>	14 (331/2,358)	1,043.4 <sup>b</sup>	.68 <sup>c</sup>	13 (307/2,286)	1,045.2 <sup>b</sup>	.68 <sup>c</sup>
Medium	38 (1,327/3,476)	466.5	3.8	37 (1,313/3,503)	426.7	3.7
High	56 (1,083/1,939)	1,012.2	7.7	57 (1,100/1,925)	991.0	7.9
Weighted mean ratings of overall need						
Low <sup>a</sup>	11 (188/1,705)	1,065.6 <sup>b</sup>	.68 <sup>c</sup>	11 (179/1,619)	995.4 <sup>b</sup>	.68 <sup>c</sup>
Medium	29 (771/2,667)	213.3	3.3	29 (776/2,707)	192.2	3.2
High	52 (1,782/3,401)	769.1	8.7	52 (1,765/3,388)	700.8	8.2

*Note.* All survival analysis chi-squared values and associated hazard ratios were significant at the  $p < .001$  level.

<sup>a</sup>The low rating was the reference category in survival analyses.

<sup>b</sup>Chi-squared values indicate results for the model as a whole.

<sup>c</sup>Harrell's C statistic.

### **Indigenous men**

As expected, the results for Indigenous men showed greater differences between the development and test samples than were observed for non-Indigenous men (Table 7). Presumably, this is simply due to the smaller groups of Indigenous men (2,321 in the development sample and 2,319 in the test sample) than the groups of non-Indigenous men (7,773 in the development sample and 7,714 in the test sample) which allows for a greater influence of chance in producing the results.

Despite this, the similar results between the development and test samples indicated good reliability for each of the overall need rating methods. The test sample displayed a decrease in strength of results for the weighted mean method of rating overall need for Indigenous men, as was the case for non-Indigenous men. However, the results of this calculated method of rating overall need were still stronger than the parole officers' intake ratings.

As was observed for the non-Indigenous men, the greatest difference between the results of the development and test samples was seen in the parole officers' intake ratings. Calculated overall need ratings appeared to be more reliable than intake ratings for Indigenous men.

Also as observed for non-Indigenous men, the Harrell's C statistics indicate stronger predictive power, and more consistent predictive power associated with the calculated overall need ratings than the parole officers' intake ratings for Indigenous men.

Table 7

*Split-half reliability of associations between revocations and parole officer intake ratings and calculated overall need ratings: Indigenous men*

Rating	Development Sample			Test Sample		
	Percent of Revocations	$\chi^2$	Hazard ratio	Percent of Revocations	$\chi^2$	Hazard ratio
Parole officer intake ratings of overall need						
Low <sup>a</sup>	44 (35/79)	26.9 <sup>b</sup>	.53 <sup>c</sup>	37 (39/105)	57.4 <sup>b</sup>	.56 <sup>c</sup>
Medium	53 (509/961)	5.8	1.5	50 (482/961)	11.7	1.8
High	59 (759/1,281)	14.3	1.9	59 (745/1,253)	31.0	2.5
Calculated mean ratings of overall need						
Low <sup>a</sup>	25 (59/239)	239.6 <sup>b</sup>	.62 <sup>c</sup>	29 (68/237)	232.0 <sup>b</sup>	.62 <sup>c</sup>
Medium	49 (484/988)	56.0	2.8	48 (462/964)	40.0	2.3
High	69 (760/1,094)	154.5	5.5	66 (736/1,118)	141.4	4.6
Calculated weighted mean ratings of overall need						
Low <sup>a</sup>	23 (49/215)	243.1 <sup>b</sup>	.62 <sup>c</sup>	27 (56/208)	234.8 <sup>b</sup>	.62 <sup>c</sup>
Medium	48 (454/943)	53.8	3.1	46 (419/914)	35.4	2.3
High	69 (800/1,163)	145.5	6.1	66 (791/1,197)	128.8	4.9

*Note.* All survival analysis chi-squared values and associated hazard ratios were significant at the  $p < .001$  level except for the intake need rating of medium in the development sample which was significant with a  $p$  of .016.

<sup>a</sup>The low rating was the reference category in survival analyses.

<sup>b</sup>Chi-squared values in the low overall need rating rows indicate results for the model as a whole.

<sup>c</sup>Harrell's C statistic.

## Women

Table 8 shows the split-half reliability results for the parole officer intake ratings and both the adjusted and unadjusted calculated ratings for federally sentenced women. As was observed for both groups of men, each method appeared reliable. Percentages of revocations in the development sample were similar to the test sample, and the results of survival analyses were also similar between the two groups. Unlike the men however, the results for the test sample were equally strong as the development sample for the weighted mean overall need rating method, and the parole officers' intake ratings did not appear less reliable than the calculated

ratings. The split-half reliability using the adjusted cut-off scores were also acceptable, showing reasonably consistent results between the development and test samples.

Furthermore, the Harrell's C statistics indicate that using either the adjusted or unadjusted cut-off scores for the calculated ratings, and both methods of calculating overall need had greater predictive power regarding revocations for women offenders. These statistics were also more consistent between the test and development samples for the calculated ratings of women than the parole officers' intake ratings.

Table 8

*Split-half reliability of associations between revocations and parole officer intake and calculated overall need ratings: Women*

Rating	Development Sample			Test Sample		
	Percent of Revocations	$\chi^2$	Hazard ratio	Percent of Revocations	$\chi^2$	Hazard ratio
Parole officer intake ratings of overall need						
Low <sup>a</sup>	21 (26/122)	27.1 <sup>b</sup>	.60 <sup>c</sup>	26 (31/121)	23.0 <sup>b</sup>	.59 <sup>c</sup>
Medium	37 (116/311)	15.2	2.3	38 (112/295)	11.9	2.0
High	46 (86/187)	26.9	3.2	46 (73/159)	22.9	2.8
Calculated mean ratings of overall need (unadjusted)						
Low <sup>a</sup>	9 (9/96)	54.9 <sup>b</sup>	.65 <sup>c</sup>	12 (12/97)	62.3 <sup>b</sup>	.66 <sup>c</sup>
Medium	35 (85/244)	23.3	5.5	33 (77/231)	18.7	3.9
High	48 (134/280)	44.2	10.1	51 (127/247)	47.1	8.3
Calculated weighted mean ratings of overall need (unadjusted)						
Low <sup>a</sup>	8 (8/96)	58.8 <sup>b</sup>	.65 <sup>c</sup>	10 (10/102)	58.5 <sup>b</sup>	.66 <sup>c</sup>
Medium	30 (55/181)	18.8	5.2	34 (59/175)	20.2	4.7
High	48 (165/343)	42.8	10.9	49 (147/298)	46.0	9.5
Calculated mean ratings of overall need (adjusted)						
Low <sup>a</sup>	17.4 (28/161)	59.0 <sup>b</sup>	.64 <sup>c</sup>	19.3 (31/161)	83.5 <sup>b</sup>	.66 <sup>c</sup>
Medium	41.7 (155/372)	37.6	3.6	38.9 (131/337)	30.4	3.1
High	51.7 (45/87)	58.3	6.5	70.1 (54/77)	81.2	8.0
Calculated weighted mean ratings of overall need (adjusted)						
Low <sup>a</sup>	12.9 (19/147)	72.7 <sup>b</sup>	.66 <sup>c</sup>	15.44 (22/143)	81.7 <sup>b</sup>	.67 <sup>c</sup>
Medium	39.9 (134/336)	39.5	4.7	38.3 (119/311)	31.6	3.8
High	54.7 (75/137)	70.7	9.0	62.0 (75/121)	75.5	8.9

*Note.* All survival analysis chi-squared values and associated hazard ratios were significant at the  $p < .001$  level.

<sup>a</sup>The low rating was the reference category in survival analyses.

<sup>b</sup>Chi-squared values in the low overall need rating rows indicate results for the model as a whole.

<sup>c</sup>Harrell's C statistic.

## Discussion

The results support the use of calculated methods for determining domain and overall ratings of need among all groups of offenders. There was greater predictive validity of the calculated domain ratings and the calculated overall ratings compared to the parole officers' ratings, and the distributions of domain and overall need ratings favoured the calculated methods. Results of the exploratory analysis comparing the two calculated methods indicated that the method of using the weighted mean did not provide better prediction over the simpler mean method. The split-half reliability was stronger for both the calculated mean and weighted mean overall need rating methods than the parole officers' overall need ratings for both groups of men, and was similar across the three methods for women.

The distributions of prevalence of need ratings of the domains of the DFIA-R are important to CSC in order to reliably identify targets for correctional interventions, assess areas that may present barriers to reintegration into the community and to profile the population of offenders. The DFIA-R domains reflect the "big four" risk factors –antisocial behaviour, personality pattern, cognitions, and associates, as well as the other factors in the "central eight" – substance abuse, marital and family, employment and education, and leisure and recreation (Andrews et al., 2006). Since the indicators and domains were selected for offender populations, they should be evenly or normally distributed within these populations.

Both calculated overall ratings and domain ratings based on the indicators achieve a more normal distribution than the parole officers' intake ratings for men and women. The parole officers' overall ratings had fewer than 10% of non-Indigenous men and fewer than 5% of Indigenous men in the low need category. The calculated ratings more than tripled the number of non-Indigenous men and more than doubled the number of Indigenous men in the low category. Nearly 30% of men in the mean method and 21% in the weighted mean method for non-Indigenous men were rated as low need, and 10% of Indigenous men in the mean method and 9% in the weighted mean method were rated as low. Since the calculated ratings were based on the indicators, we can conclude that men are often seen by parole officers as posing moderate or high dynamic risk even when few indicators are endorsed. Women were more likely to receive a rating of high overall need according to the calculated overall ratings using the formula applied to men that is based on a percentage of indicators endorsed. A higher proportion of high need



ratings for women that results from the calculated methods may reflect the very disadvantageous background of federally sentenced women and benefit women by signaling the type of correctional interventions they require. The calculated methods predict outcomes more accurately meaning that, within the women offender population, those rated high need are more likely to be revoked than women rated moderate or low need. It should be noted that this is a relative finding within the women offender population and does not mean that high needs women are more likely to be revoked than high needs men; rather, it demonstrates the strength of the difference in outcomes for women at each need rating. We found that the distributions and the predictive power of the calculated methods were further improved by making the classification for the moderate and high need ratings more stringent for the women.

The calculated domain ratings had the advantage that they were more normally distributed than the parole officers' ratings. The differences between the distributions of calculated and parole officers' intake domain ratings were often quite dramatic. For example, the proportion of non-Indigenous men with community functioning ratings of asset or no need using the parole officers' assessment at intake was 68%, 3.5 times larger than the 19.5% when the calculated domain rating method was applied. These types of results suggest that the indicators are not always informing the parole officers' domain ratings despite research supporting the correlation between indicators and domain ratings (Stewart et al., 2017). Further evidence of the face validity of the calculated method is provided by the observation that men and women receiving a no need rating actually have no indicators endorsed while using parole officers' ratings many men and women were rated as no need despite having indicators endorsed.

### **Associations of Domain Ratings with Revocations**

The utility of the DFIA-R is linked with the extent to which ratings on needs are related to outcomes in the community – its ability to identify factors that present barriers to reintegration and are relevant to criminal behavior. The calculated ratings based directly on the proportion of DFIA-R indicators endorsed were stronger predictors of revocations than parole officer ratings. This was true across all domains and all three study groups – non-Indigenous men, Indigenous men, and women. The results showed greater differentiation in revocations across the calculated domain rating levels. Calculated domain ratings eliminated counter-intuitive results such as a higher rate of revocations associated with the low and moderate intake ratings of than ratings of high or higher rates of revocations at the intake ratings of asset or no need than low need. The

distributions of the calculated domain ratings discussed above improve the predictive validity, and the low proportion of offenders with no need was appropriate.

The consistently stronger relationship between calculated ratings and revocations should not be surprising. There is a long history in applied psychology and forensic risk assessment of calculated assessments outperforming clinical judgements (Grove, 2005; Grove & Meehl, 1996, Wormith, 2017). One reason the DFIA-R results may follow this trend is related to the amount of information parole officers are required to consider in each rating. Even the domain with the fewest indicators – the community functioning domain – may be too large to allow assessors to simultaneously consider all the information. Human working memory, that is, the capacity to receive, process, and remember information at a given time is restricted (Miller, 1956). This introduces the likelihood that certain indicators may be given more importance and others may be overlooked without any empirical basis to support these decisions. The calculated method to determine domain ratings excluded a few indicators based on empirical evidence that they were not related to outcome and otherwise considered all indicators equally, thus producing more reliable ratings and stronger predictions of revocations.

### **Associations of Overall Ratings with Revocations**

Similarly, the calculated overall ratings considered all the domains, and produced greater differentiation among the rating levels regarding revocations than parole officers' intake ratings. The calculated ratings had better predictive validity than the parole officers' intake ratings. The improvement in predictive validity was in large part due to reducing misclassification of offenders between the low and medium overall need ratings. Both methods of calculated overall ratings realized this advantage and it was present for all three groups of offenders. The higher rate of revocations in the parole officers' intake ratings of low need compared to the calculated overall ratings of low suggests that many offender intake ratings of low should have been medium need. Furthermore, since the group of offenders with calculated overall ratings of low was larger than the intake ratings of low, many offenders with intake ratings of medium actually belonged to the low group.

Although the calculated method improved the tool's distribution of ratings and predictive power a potential concern was the increase in the percentage of women offenders rated as high need using the calculated method over the parole officer ratings. While this accurately reflected the greater number of indicators endorsed among women, higher overall need ratings could be

problematic in that it could affect release decisions disadvantageously for women. We therefore examined the impact on the validity of the tool when we adjusted the formula for allocating women to a need rating. Results indicated that the predictive validity of the calculated methods was further improved by making the classification for the moderate and high need ratings more stringent for the women. We would therefore recommend that this would be the preferred scoring option for women offenders. Split half reliability for women using these adjusted cut-off scores was unaffected.

### **Split-Half Reliability of Overall Ratings.**

Although the split-half reliability results suggested that all three methods of rating overall need appeared reasonably reliable for non-Indigenous men, Indigenous men, and women, the parole officers' intake ratings were less reliable than the calculated ratings for both groups of men. The weighted mean method of rating overall need was expected to have the lowest split-half reliability since it relied on hazard ratios calculated for one group and tested on another. The increased variability in scores and ratings of this method were not great enough to question the reliability of this method. The variability in parole officers' overall need ratings at intake, however, while reasonably consistent was noticeably greater than for the calculated methods. This suggests that the intake ratings were less accurate and more variable, consistent with the conclusion that there would be more frequent misclassifications associated with parole officers' ratings.

### **Limitations**

This study only examined revocations of conditional release without differentiating revocations for technical infractions or revocations with an offence. Future research should examine revocations with an offence and reoffending post-warrant expiry. In addition, we did not specifically examine 'strength-based' factors referred to as 'asset' ratings in the DFIA-R. Very few offenders were given this rating on the domains but future research should determine if calculations should be adjusted to reflect the number of domains that have these ratings. It should also be noted that the results were based on the initial assessment of need which for many offenders could have been conducted be at least a couple of years prior to their release. Recent research suggests that dynamic assessments conducted later in the sentence, closer to the time of release, provide a better prediction of outcome on release (Mulvey, Schubert, Pitzer, Hawes,

Piquero, & Cardwell, 2016). With respect to the results for women offenders, given lower base rates of revocation and smaller numbers we were unable to disaggregate by Indigenous and non-Indigenous background. We recognise, based on the profile information derived from the DFIA-R, that these two groups do differ with respect to the extent of their need (Stewart et al., 2017). However, the results for non-Indigenous and Indigenous men in the current study still support the stronger predictive validity and reliability of the calculated overall and domain ratings over the parole officers' intake ratings, and the strength of results for women would suggest that this would be true for both non-Indigenous and Indigenous women. Since the weighted mean method of calculating the overall ratings, which was based on different formulae for non-Indigenous men, Indigenous men and women, demonstrated similar predictive ability as the mean method, the differences between non-Indigenous men, Indigenous men, and women – and by extension between non-Indigenous and Indigenous women – are not substantial enough to support different methods of calculating ratings.

In assessing the impact of the indicators on release we did not control for intervening influences that might have attenuated the impact of some indicators. For example, participation in correctional programs and education and employment programs could reduce the extent to which indicators assessed as present at intake were associated with outcome on release and implementation of a well structured release plan could affect the extent to which indicators on the community function domain affect outcome. This approach could therefore underestimate the importance of some indicators; however, where the significant association exists it is likely to be one that represents an actual effect.

Finally, we found that adjusting the calculated formulas used to allocate women offenders to the need levels for the overall rating improved the predictability of the tool and was therefore a more accurate reflection of the impact of overall need on outcomes for women. A more thorough investigation of the impact of various cut-off values should be explored in future research. Further improvements in predictive power for women and similar improvements for non-Indigenous and Indigenous men could be realized.

## **Conclusions**

Although the parole officers' ratings are valid and reliable as established by Stewart and colleagues (Stewart et al., 2017), calculated ratings improve both the predictive validity and reliability of the DFIA-R. Using calculated scoring methods, parole and program officers who work directly with offenders will have more accurate information on which need domains require focused intervention for men, women, and Indigenous offenders. Reducing misclassification of offenders will improve the relevance and effectiveness of the offenders' correctional plans.

It should be noted that the calculated method does not take the human element out of the assessment procedure. Both the calculated methods studied here still rely on the input of parole officers in the assessment of the indicators; only the calculation of domain and overall ratings would be automated. Indeed, the interview process that results in the population of the indicators provides the individual contact required to determine the offenders' consideration of their histories and the factors that contributed to their criminality, as well as the degree to which they are ready to be engaged in addressing the factors. The calculated methods, however, provide a simple low-cost alternative to the parole officers' ratings and have better measurement properties. Importantly, these methods accommodate the differences between men and women and Indigenous offenders in their calculation and therefore can respond to the criticism that generic tools are not adapted to various offender groups. If a electronically calculated rating is adopted by CSC consideration could be given to allowing for an assessor over-ride but we maintain that this over-ride should require a justification and the calculated score should be presented in addition so that future research can determine whether over-rides are of value.

## References

- Andrews, D.A., & Bonta, J. (2006). *The psychology of criminal conduct* (4<sup>th</sup> ed.). Newark, NJ: LexisNexis/Matthew Bender.
- Andrews, D.A., Bonta, J., & Wormith, J.S. (2006). The recent past and near future of risk and/or need assessment. *Crime Delinquency*, 52(7), 7-27. DOI: 10.1177/0011128705281756.
- Brown, S.L., & Motiuk, L.L. (2005). *The Dynamic Factors Identification and Analysis (DFIA) Component of the Offender Intake Assessment (OIA) Process: A meta-analytic, psychometric and consultative review* (Research Report R-164). Ottawa, ON: Correctional Service of Canada.
- Grove, W.M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, 61(10), 1233-1243. DOI: 10.1002/jclp.20179.
- Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293-323. <http://dx.doi.org/10.1037/1076-8971.2.2.293>
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97 <http://dx.doi.org/10.1037/h0043158>
- Motiuk, L.L., & Porporino, F. (1989). *Field test of the Community Risk/Needs Management Scale: A study of offenders on caseload*. (Research Report R-06). Ottawa, ON: Correctional Service of Canada.
- Mulvey, E. P., Schubert, C. A., Pitzer, L., Hawes, S., Piquero, A., & Cardwell, S. (2016). An examination of change in dynamic risk of offending over time among serious juvenile offenders. *Journal of Criminal Justice*, 45, 48-53. doi: 10.1016/j.jcrimjus.2016.02.008
- Stewart, L.A., Wardrop, K., Wilton, G., Thompson, J., Derkzen, D., & Motiuk, L.L. (2017). *Reliability and validity of the Dynamic Factors Identification and Analysis – Revised*. (Research Report R-395). Ottawa, ON: Correctional Service of Canada.
- Wormith, J.S. (2017). Automated offender risk assessment: The next generation or a black hole? *Criminology & Public Policy*, 16(1), 281-303. DOI:10.1111/1745-9133.12277

## Appendix A: Hazard Ratio Weights Applied to Calculate Overall Need Ratings

Table A1

*Hazard ratio weights applied to calculate overall need ratings for non-Indigenous men, Indigenous men, and women*

Domain	Hazard Ratio Weight
<b>Non-Indigenous Men</b>	
Employment/Education	1.957
Marital/Family	1.013
Associates	2.277
Substance Abuse	2.281
Community Functioning	1.953
Personal/Emotional	0.815
Attitude	2.479
<b>Non-Indigenous Men</b>	
Employment/Education	2.202
Marital/Family	0.990
Associates	2.143
Substance Abuse	1.759
Community Functioning	1.268
Personal/Emotional	1.077
Attitude	1.428
<b>Women</b>	
Employment/Education	2.529
Marital/Family	1.570
Associates	1.869
Substance Abuse	3.023
Community Functioning	1.983
Personal/Emotional	0.486
Attitude	1.132

*Note* These hazard ratios are the result of the multiple regressions that included all the domains and should not be interpreted to represent the individual predictability of the domains. The Personal Emotional domain for example predicts outcomes on its own for women but it shares variance with the other domains in the regression.

Below are the formulae that determined the weighted overall need ratings.

Non-Indigenous men

(Mean of employment/education indicators \* 1.957  
 + mean of marital/family indicators \* 1.013  
 + mean of associates indicators \* 2.277  
 + mean of substance abuse indicators \* 2.281

- + mean of community functioning indicators \* 1.953
- + mean of personal/emotional indicators \* 0.815
- + mean of attitude indicators \* 2.479)/7

Low rating: a value less than or equal to 0.375 (0.25 times the mean weight of 1.499)

Moderate rating: a value greater than 0.375 but less than or equal to 0.750 (0.5 times 1.499).

High rating: a value greater than 0.750.

#### Indigenous men

(Mean of employment/education indicators \* 2.202

- + mean of marital/family indicators \* 0.990
- + mean of associates indicators \* 2.143
- + mean of substance abuse indicators \* 1.759
- + mean of community functioning indicators \* 1.268
- + mean of personal/emotional indicators \* 1.077
- + mean of attitude indicators \* 1.428)/7

Low rating: a value less than or equal to 0.388 (0.25 times the mean weight of 1.552)

Moderate rating: a value greater than 0.388 but less than or equal to 0.776 (0.5 times 1.552).

High rating: a value greater than 0.776.

#### Women (unadjusted)

(Mean of employment/education indicators \* 2.529

- + mean of marital/family indicators \* 1.570
- + mean of associates indicators \* 1.869
- + mean of substance abuse indicators \* 3.023
- + mean of community functioning indicators \* 1.983
- + mean of personal/emotional indicators \* 0.486
- + mean of attitude indicators \* 1.132)/7

Low rating: a value less than or equal to 0.450 (0.25 times the mean weight of 1.799)

Moderate rating: a value greater than 0.450 but less than or equal to 0.900 (0.5 times 1.799).

High rating: a value greater than 0.900.



Women (adjusted)

The weightings were unchanged; only the cut-offs for the ratings were adjusted as follows.

Low rating: a value less than or equal to 0.594 (0.33 times the mean weight of 1.799)

Moderate rating: a value greater than 0.594 but less than or equal to 1.187 (0.66 times 1.799)

High rating: a value greater than 1.187.

**Appendix B: Survival analyses comparing parole officer intake and calculated domain ratings on revocations**

Table B1.

*Survival analysis results for parole officer intake and calculated domain ratings predicting revocations for non-Indigenous men.*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	<i>p</i>	$\chi^2$	Hazard Ratio	<i>p</i>
<b>Employment/education</b>						
Asset or no need <sup>a</sup>						
Low Need	46.4	1.36	< .001	71.6	1.90	< .001
Moderate Need	431.7	2.31	< .001	258.5	3.31	< .001
High Need	489.2	3.47	< .001	492.6	5.39	< .001
<i>Model Fit</i>						
n	15,487			15,400		
Wald $\chi^2$	756.6 <sup>b</sup>	.61 <sup>c</sup>	< .001	1,064.3 <sup>d</sup>	.63 <sup>c</sup>	< .001
<b>Marital/family</b>						
Asset or no need <sup>a</sup>						
Low Need	18.1	1.19	< .001	142.5	1.88	< .001
Moderate Need	71.7	1.34	< .001	456.2	3.19	< .001
High Need	15.4	1.21	< .001	291.7	3.86	< .001
<i>Model Fit</i>						
n	15,487			15,479		
Wald $\chi^2$	81.1 <sup>e</sup>	.53 <sup>c</sup>	< .001	671.3 <sup>e</sup>	.59 <sup>c</sup>	< .001
<b>Associates</b>						
Asset or no need <sup>a</sup>						
Low Need	5.5	1.14	.019	86.0	2.35	< .001
Moderate Need	226.2	1.77	< .001	414.6	6.10	< .001
High Need	434.9	2.34	< .001	525.5	8.77	< .001
<i>Model Fit</i>						
n	15,487			15,483		
Wald $\chi^2$	514.2 <sup>f</sup>	.58 <sup>c</sup>	< .001	1,328.1 <sup>e</sup>	.64 <sup>c</sup>	< .001
<b>Substance Abuse</b>						
No need <sup>a</sup>						
Low Need	59.4	1.49	< .001	75.8	1.64	< .001
Moderate Need	355.2	2.27	< .001	519.4	3.19	< .001
High Need	1,064.0	3.48	< .001	935.5	4.73	< .001
<i>Model Fit</i>						
n	15,487			15,467		
Wald $\chi^2$	1,207.6 <sup>e</sup>	.63 <sup>c</sup>	< .001	1,317.7 <sup>e</sup>	.64 <sup>c</sup>	< .001

Table B1 *Continued*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	$p$	$\chi^2$	Hazard Ratio	$p$
<b>Community functioning</b>						
Asset or no need <sup>a</sup>						
Low Need	86.3	1.49	< .001	128.4	1.84	< .001
Moderate Need	516.7	2.17	< .001	574.9	3.42	< .001
High Need	448.0	3.00	< .001	1,059.9	5.21	< .001
<i>Model Fit</i>						
n	15,487			15,478		
Wald $\chi^2$	818.0 <sup>c</sup>	.59 <sup>c</sup>	< .001	1480.6 <sup>g</sup>	.65 <sup>c</sup>	< .001
<b>Personal/emotional</b>						
No need <sup>a</sup>						
Low Need	4.0	0.88	.044	48.4	2.53	< .001
Moderate Need	40.8	1.30	< .001	149.0	5.08	< .001
High Need	256.8	1.89	< .001	242.1	8.19	< .001
<i>Model Fit</i>						
n	15,487		< .001	15,483		< .001
Wald $\chi^2$	393.8 <sup>h</sup>	.58 <sup>c</sup>	< .001	1,107.9 <sup>i</sup>	.63 <sup>c</sup>	< .001
<b>Attitude</b>						
Asset or no need <sup>a</sup>						
Low Need	0.6	0.95	0.437	95.5	1.95	< .001
Moderate Need	90.9	1.52	< .001	481.0	4.36	< .001
High Need	507.6	2.61	< .001	818.9	7.15	< .001
<i>Model Fit</i>						
n	15,487			15,486		< .001
Wald $\chi^2$	771.8 <sup>j</sup>	.61 <sup>c</sup>	< .001	1,707.4 <sup>k</sup>	.66 <sup>c</sup>	

<sup>a</sup>The asset or no need category served as the reference category.

<sup>b</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,487$ ) = 6.7,  $p = .009$ ). The hazard ratios associated with the intake employment/education domain change at different points of follow-up time.

<sup>c</sup>Harrell's C statistic.

<sup>d</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,400$ ) = 14.7,  $p < .001$ ). The hazard ratios associated with the calculated employment/education domain change at different points of follow-up time.

<sup>e</sup>The assumption of proportional hazards was not violated.

<sup>f</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,487$ ) = 8.7,  $p = .003$ ). The hazard ratios associated with the intake associates domain change at different points of follow-up time.

<sup>g</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,478$ ) = 10.3,  $p = .001$ ). The hazard ratios associated with the calculated community functioning domain change at different points of follow-up time.

<sup>h</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,487$ ) = 12.2,  $p < .001$ ). The hazard ratios associated with the intake personal/emotional domain change at different points of follow-up time.

<sup>i</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,483$ ) = 11.0,  $p < .001$ ). The hazard ratios associated with the calculated personal/emotional domain change at different points of follow-up time.

<sup>j</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,487$ ) = 5.7,  $p = .017$ ). The hazard ratios associated with the intake attitude domain change at different points of follow-up time.

<sup>k</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 15,486$ ) = 4.8,  $p = .029$ ). The hazard ratios associated with the intake attitude domain change at different points of follow-up time.

Table B2

*Survival analysis results for parole officer intake ratings and calculated domain ratings predicting revocations for Indigenous men*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	<i>p</i>	$\chi^2$	Hazard Ratio	<i>p</i>
<b>Employment/education</b>						
Asset or no need <sup>a</sup>						
Low Need	7.3	1.30	.007	9.8	1.69	.002
Moderate Need	95.8	2.29	< .001	39.6	2.83	< .001
High Need	153.2	3.23	< .001	75.8	4.23	< .001
<i>Model Fit</i>						
n	4,640			4,629		
Wald $\chi^2$	254.1 <sup>b</sup>	.59 <sup>c</sup>	< .001	328.7 <sup>d</sup>	.60 <sup>c</sup>	< .001
<b>Marital/family</b>						
Asset or no need <sup>a</sup>						
Low Need	3.0	1.11	.086	19.0	1.70	< .001
Moderate Need	24.3	1.27	< .001	48.7	2.31	< .001
High Need	4.0	1.13	.045	65.7	2.86	< .001
<i>Model Fit</i>						
n	4,640			4,640		
Wald $\chi^2$	24.4 <sup>c</sup>	.52 <sup>c</sup>	< .001	119.7 <sup>e</sup>	.56 <sup>c</sup>	< .001
<b>Associates</b>						
Asset or no need <sup>a</sup>						
Low Need	2.4	0.88	.125	13.6	2.11	< .001
Moderate Need	33.8	1.39	< .001	48.1	3.95	< .001
High Need	81.0	1.70	< .001	83.3	6.19	< .001
<i>Model Fit</i>						
n	4,640			4,640		
Wald $\chi^2$	120.6 <sup>e</sup>	.56 <sup>c</sup>	< .001	346.7 <sup>f</sup>	.60 <sup>c</sup>	< .001
<b>Substance Abuse</b>						
No need <sup>a</sup>						
Low Need	15.3	1.56	< .001	11.1	2.06	< .001
Moderate Need	46.4	1.91	< .001	33.3	3.31	< .001
High Need	98.5	2.42	< .001	58.3	4.82	< .001
<i>Model Fit</i>						
n	4,640			4,637		
Wald $\chi^2$	128.3 <sup>e</sup>	.55 <sup>c</sup>	< .001	206.6 <sup>e</sup>	.57 <sup>c</sup>	< .001
<b>Community functioning</b>						
Asset or no need <sup>a</sup>						
Low Need	9.1	1.20	.003	28.4	1.75	< .001
Moderate Need	58.3	1.44	< .001	84.9	2.57	< .001
High Need	79.7	1.85	< .001	154.8	3.50	< .001
<i>Model Fit</i>						
n	4,640			4,639		
Wald $\chi^2$	111.2 <sup>e</sup>	.55 <sup>c</sup>	< .001	276.6 <sup>g</sup>	.60 <sup>c</sup>	< .001

Table B2 *Continued*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	$p$	$\chi^2$	Hazard Ratio	$p$
Personal/emotional						
No need <sup>a</sup>						
Low Need	0.2	1.06	.632	3.9	1.82	< .001
Moderate Need	12.0	1.32	< .001	15.9	3.34	< .001
High Need	45.6	1.67	< .001	22.7	4.25	< .001
<i>Model Fit</i>						
N	4,640			4,639		
Wald $\chi^2$	72.4 <sup>e</sup>	.54 <sup>c</sup>	< .001	248.5 <sup>e</sup>	.59 <sup>c</sup>	< .001
Attitude						
Asset or no need <sup>a</sup>						
Low Need	0.1	0.97	.706	23.2	1.65	< .001
Moderate Need	30.1	1.38	< .001	100.0	2.76	< .001
High Need	71.3	1.64	< .001	163.0	3.72	< .001
<i>Model Fit</i>						
n	4,640			4,639		
Wald $\chi^2$	93.3 <sup>e</sup>	.55 <sup>c</sup>	< .001	332.0 <sup>e</sup>	.60 <sup>c</sup>	< .001

<sup>a</sup>The asset or no need category served as the reference category.

<sup>b</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N=4,640$ ) = 4.1,  $p = .042$ ). The hazard ratios associated with the intake employment/education domain change at different points of follow-up time.

<sup>c</sup>Harrell's C statistic.

<sup>d</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 4,629$ ) = 6.5,  $p = .011$ ). The hazard ratios associated with the calculated employment/education domain change at different points of follow-up time.

<sup>e</sup>The assumption of proportional hazards was not violated.

<sup>f</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 4,640$ ) = 8.69,  $p = .003$ ). The hazard ratios associated with the calculated associates domain change at different points of follow-up time.

<sup>g</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 4,639$ ) = 5.8,  $p = .016$ ). The hazard ratios associated with the calculated community functioning domain change at different points of follow-up time.

Table B3

*Survival analysis results for parole officer intake ratings and calculated domain ratings predicting revocations for women.*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	$p$	$\chi^2$	Hazard Ratio	$p$
<b>Employment/education</b>						
Asset or no need <sup>a</sup>						
Low Need	11.9	2.15	< .001	4.8	2.27	< .001
Moderate Need	52.2	4.14	< .001	16.7	4.41	< .001
High Need	52.1	4.81	< .001	31.8	7.67	< .001
<i>Model Fit</i>						
n	1,195			1,193		
Wald $\chi^2$	76.6 <sup>b</sup>	.61 <sup>c</sup>	< .001	101.1 <sup>b</sup>	.64 <sup>c</sup>	< .001
<b>Marital/family</b>						
Asset or no need <sup>a</sup>						
Low Need	0.3	1.09	.599	5.6	2.72	< .001
Moderate Need	21.1	1.79	< .001	16.9	5.48	< .001
High Need	4.5	1.38	.033	25.0	8.13	< .001
<i>Model Fit</i>						
n	1,195			1,192		
Wald $\chi^2$	26.0 <sup>b</sup>	.57 <sup>c</sup>	< .001	75.4 <sup>d</sup>	.63 <sup>c</sup>	< .001
<b>Associates</b>						
Asset or no need <sup>a</sup>						
Low Need	1.1	0.77	.302	3.2	2.53	.074
Moderate Need	19.6	2.10	< .001	13.9	6.55	< .001
High Need	45.1	3.04	< .001	22.1	10.88	< .001
<i>Model Fit</i>						
n	1,195			1,192		
Wald $\chi^2$	75.9 <sup>b</sup>	.61 <sup>c</sup>	< .001	97.3 <sup>b</sup>	.63 <sup>c</sup>	< .001
<b>Substance Abuse</b>						
No need <sup>a</sup>						
Low Need	1.6	1.53	.211	11.3	2.71	< .001
Moderate Need	18.4	2.77	< .001	38.1	4.79	< .001
High Need	79.7	5.40	< .001	70.3	7.81	< .001
<i>Model Fit</i>						
n	1,195			1,193		
Wald $\chi^2$	103.3 <sup>b</sup>	.63 <sup>c</sup>	< .001	99.4 <sup>b</sup>	.64 <sup>c</sup>	< .001
<b>Community functioning</b>						
Asset or no need <sup>a</sup>						
Low Need	0.6	0.90	.444	3.7	2.00	.055
Moderate Need	13.5	1.51	< .001	16.9	4.13	< .001
High Need	13.4	1.89	< .001	28.1	6.07	< .001
<i>Model Fit</i>						
n	1,195			1,189		
Wald $\chi^2$	28.6 <sup>b</sup>	.58 <sup>c</sup>	< .001	77.1 <sup>b</sup>	.62 <sup>c</sup>	< .001

Table B3 *Continued*

Domain Rating	Parole Officer Intake Ratings			Calculated Ratings		
	$\chi^2$	Hazard Ratio	<i>p</i>	$\chi^2$	Hazard Ratio	<i>p</i>
<b>Personal/emotional</b>						
No need <sup>a</sup>						
Low Need	2.5	0.60	.113	0.7	1.55	.387
Moderate Need	0.4	1.17	.511	3.5	2.56	.063
High Need	7.2	1.86	.007	7.8	4.18	.005
<i>Model Fit</i>						
n	1,195			1,195		
Wald $\chi^2$	39.4 <sup>e</sup>	.59 <sup>c</sup>	< .001	54.1 <sup>f</sup>	.60 <sup>c</sup>	< .001
<b>Attitude</b>						
Asset or no need <sup>a</sup>						
Low Need	0.3	0.92	.572	11.6	1.59	< .001
Moderate Need	5.9	1.32	.015	54.3	2.83	< .001
High Need	9.8	1.54	.002	54.5	4.30	< .001
<i>Model Fit</i>						
n	1,195			1,190		
Wald $\chi^2$	15.3 <sup>b</sup>	.56 <sup>c</sup>	.002	84.6 <sup>g</sup>	.62 <sup>c</sup>	< .001

<sup>a</sup>The asset or no need category served as the reference category.

<sup>b</sup>The assumption of proportional hazards was not violated.

<sup>c</sup>Harrell's C statistic.

<sup>d</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1, *N* = 1,192) = 6.8, *p* = .009). The hazard ratios associated with the calculated marital/family domain change at different points of follow-up time.

<sup>e</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1, *N* = 1,195) = 8.0, *p* = .005). The hazard ratios associated with the intake personal/emotional domain change at different points of follow-up time.

<sup>f</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1, *N* = 1,195) = 6.2, *p* = .013). The hazard ratios associated with the calculated personal/emotional domain change at different points of follow-up time.

<sup>g</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1, *N* = 1,190) = 6.3, *p* = .012). The hazard ratios associated with the calculated attitude domain change at different points of follow-up time.

## Appendix C: Survival Analyses Comparing Parole Officer Ratings and Calculated Overall Ratings on Revocations

Table C1

*Survival analysis results for parole officer intake, mean, and weighted mean overall ratings predicting revocations for the test sample of men offenders.*

Rating	Parole Officer Rating			Calculated Mean			Calculated Weighted Mean		
	$\chi^2$	Hazard Ratio	$p$	$\chi^2$	Hazard Ratio	$p$	$\chi^2$	Hazard Ratio	$p$
Non-Indigenous Men									
Low <sup>a</sup>									
Medium	57.9	2.0	< .001	426.7	3.7	< .001	192.2	3.2	< .001
High	189.2	3.3	< .001	991.0	7.9	< .001	700.8	8.2	< .001
<i>Model Fit</i>									
n	7,714			7,714			7,714		
Wald $\chi^2$	304.7 <sup>b</sup>	.59 <sup>c</sup>	< .001	1,045.2 <sup>d</sup>	.68 <sup>c</sup>	< .001	995.4 <sup>e</sup>	.68 <sup>c</sup>	< .001
Indigenous Men									
Low <sup>a</sup>									
Medium	11.7	1.8	< .001	40.0	2.3	< .001	35.4	2.3	< .001
High	31.0	2.5	< .001	141.4	4.6	< .001	128.8	4.9	< .001
<i>Model Fit</i>									
n	2,319			2,319			2,319		
Wald $\chi^2$	57.4 <sup>f</sup>	.56 <sup>c</sup>	< .001	232.0 <sup>g</sup>	.62 <sup>c</sup>	< .001	234.8 <sup>h</sup>	.62 <sup>c</sup>	< .001

<sup>a</sup>The low need category was the reference group.

<sup>b</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 7,714$ ) = 6.2,  $p = .013$ ). The hazard ratios associated with the intake overall need ratings change at different points of follow-up time for non-indigenous men.

<sup>c</sup>Harrell's C statistic.

<sup>d</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 7,714$ ) = 7.8,  $p = .005$ ). The hazard ratios associated with the overall need ratings calculated with the mean method change at different points of follow-up time for non-indigenous men.

<sup>e</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 7,714$ ) = 8.7,  $p = .003$ ). The hazard ratios associated with the overall need ratings calculated with the weighted mean method change at different points of follow-up time for non-indigenous men.

<sup>f</sup>The assumption of proportional hazards was not violated.

<sup>g</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 2,319$ ) = 12.0,  $p < .001$ ). The hazard ratios associated with the overall need ratings calculated with the mean method change at different points of follow-up time for indigenous men.

<sup>h</sup>The assumption of proportional hazards was violated (Wald  $\chi^2$  (1,  $N = 2,319$ ) = 5.9,  $p = .015$ ). The hazard ratios associated with the overall need ratings calculated with the weighted mean method change at different points of follow-up time for indigenous men.



Table C2

*Survival analysis results for parole officer intake rating and calculated mean and weighted mean overall ratings predicting revocations for the test sample of women offenders.*

Rating	Parole Officer Intake Rating			Calculated Mean			Calculated Weighted Mean		
	$\chi^2$	Hazard Ratio	<i>p</i>	$\chi^2$	Hazard Ratio	<i>p</i>	$\chi^2$	Hazard Ratio	<i>p</i>
Women (unadjusted)									
Low <sup>a</sup>									
Medium	11.9	2.0	< .001	18.8	3.9	< .001	20.2	4.7	< .001
High	22.9	2.8	< .001	47.1	8.3	< .001	46.0	9.5	< .001
<i>Model Fit</i>									
N	575			575			575		
Wald $\chi^2$	23.0 <sup>b</sup>	.59 <sup>c</sup>	< .001	62.3 <sup>b</sup>	.66 <sup>c</sup>	< .001	58.5 <sup>b</sup>	.66 <sup>c</sup>	< .001
Women (adjusted)									
Low <sup>a</sup>									
Medium	11.9	2.0	< .001	30.4	3.1	< .001	31.6	3.8	< .001
High	22.9	2.8	< .001	81.2	8.0	< .001	75.5	8.9	< .001
<i>Model Fit</i>									
n	575			575			575		
Wald $\chi^2$	23.0 <sup>b</sup>	.59 <sup>c</sup>	< .001	83.5 <sup>b</sup>	.66 <sup>c</sup>	< .001	81.7 <sup>b</sup>	.67 <sup>c</sup>	< .001

<sup>a</sup> The low need category was the reference group.

<sup>b</sup> The assumption of proportional hazards was not violated.

<sup>c</sup> Harrell's C statistic.