



**EVIDENT  
CHANGE**  
Inform Systems. Transform Lives.

**NEW  
HAMPSHIRE  
CHILD WELFARE  
SDM<sup>®</sup> RISK  
VALIDATION  
STUDY REPORT**

## NEW HAMPSHIRE DIVISION FOR CHILDREN, YOUTH AND FAMILIES

October 2022

Structured Decision Making® and SDM® are registered in the US Patent and Trademark Office.



### ABOUT EVIDENT CHANGE

Evident Change promotes just and equitable social systems for individuals, families, and communities through research, public policy, and practice. For more information, call (800) 306-6223 or visit us online at [EvidentChange.org](https://EvidentChange.org) and [@Evident\\_Change](https://twitter.com/Evident_Change) on Twitter.

© 2022 Evident Change

# CONTENTS

- Background.....1
  - Study Goals.....1
  - Project Approach .....1
  - Risk Validation Steering Committee .....2
  
- Validation Kickoff .....2
  
- Examining Current Risk Assessment Performance .....6
  - Utility.....7
  - Accuracy ..... 8
  - Equity .....9
  - Identifying Areas for Improvement..... 11
  
- Exploring Risk Assessment Revisions ..... 13
  - Comparing Utility ..... 13
  - Comparing Accuracy ..... 14
  - Comparing Equity ..... 16
  - Selecting a Model.....20
  
- Next Steps: Implementing the Revised Risk Assessment..... 22

APPENDIX: Model B Risk Assessment Item Analysis

# BACKGROUND

In 2019, the New Hampshire Division for Children, Youth and Families (DCYF) partnered with Evident Change (formerly the NCCD Children’s Research Center or CRC) to update their current suite of Structured Decision Making® (SDM) assessments. The intake assessment, safety assessment, risk reassessment, and reunification assessment are being updated collaboratively by a workgroup of local stakeholders, DCYF partners, and Evident Change staff.

SDM® risk assessments are actuarial tools completed for the household at the end of an assessment to classify families by their likelihood of future child welfare system involvement. By identifying families at the highest risk of future system involvement, an agency can target post-assessment services. Because the risk assessment is actuarial, updates to the assessment items, scoring, and risk-level thresholds are informed by an analytics process. The updates are then vetted by a stakeholder group to ensure the revised tool supports and upholds the agency’s values, works accurately and equitably across families by subpopulations such as race and ethnicity, and provides utility to workers in the field.

The following sections describe the risk validation partnership and process used by Evident Change and DCYF, results from study phases, and steps to implement the revised risk assessment.

## STUDY GOALS

The goals of the New Hampshire risk validation study were:

- Examine how the current risk assessment is performing across three SDM principles: utility (how useful the assessment is), accuracy (how accurately it classifies families by their likelihood of subsequent involvement), and equity (how comparatively well it works for different subgroups in the population).
- If current risk assessment does *not* meet stakeholders’ and Evident Change’s expectation for performance across these three principles, model revisions using available data. Then examine revisions to the risk assessment to explore how well they improve utility, accuracy, and/or equity.

## PROJECT APPROACH

Risk assessments are actuarial tools designed and revised with an emphasis on data analytics and statistical relationships. However, because child welfare staff complete these tools in the field and use them to guide their decisions about services for children and families, a critical step in the validation process is to solicit feedback from agency and community stakeholders and guide the validation with that feedback. Reviewing the data and ultimately making validation choices therefore requires a human-centered design approach. The

study was broken into three stages: validation kickoff, current risk assessment findings, and revised model selection. Each stage began with data analytics that were followed by presentations, discussions, and workgroups with a risk validation steering committee.

## RISK VALIDATION STEERING COMMITTEE

DCYF convened a steering committee composed of agency leadership, agency supervisors, agency line staff, and staff from community partner agencies. The steering committee's role was to ask questions, provide feedback, and make decisions about the risk assessment's performance throughout the validation process. This group was invited to participate in all three stages of the study and worked in partnership with DCYF administrators and Evident Change. Risk assessment validations involve numerous ways to organize data and explore the risk classification. This results in risk models that can be compared to balance usability, equity, and predictive accuracy. The steering committee was responsible for reviewing the validation analytics, vetting the analytics in relation to local practice, and ensuring the resultant risk assessment was the most appropriate for use by DCYF staff with New Hampshire families.

## VALIDATION KICKOFF

Conducting a successful risk validation study largely depends on the extent to which the previous risk assessment was implemented and used as intended. To examine current use and to determine the feasibility of a full risk validation study, Evident Change conducted a preliminary implementation analysis to explore SDM risk assessment completion rates, risk assessment findings, and alignment between case actions and risk-based case recommendations.

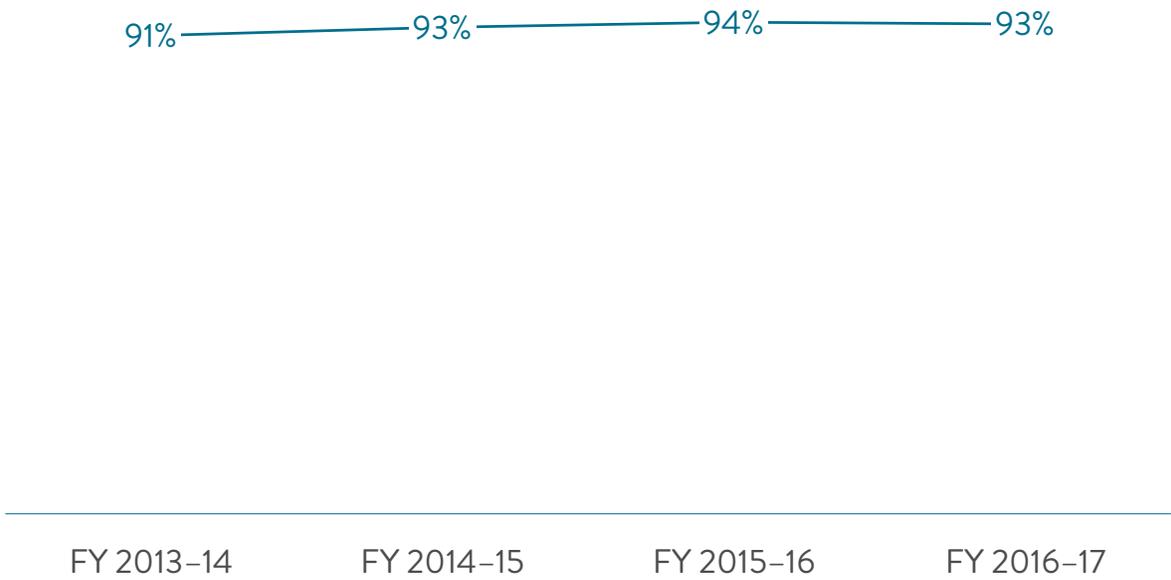
Evident Change used two cohorts of reports accepted for assessment (one including reports received between July 1, 2013, and June 30, 2017; and another including reports received between July 1, 2018, and June 30, 2019) to understand risk assessment completion on primary households<sup>1</sup> over time.<sup>2</sup> Risk completion rates were high across the five fiscal years. The completion rate increased from 91% in fiscal year 2013–14 to 95% in fiscal year 2018–19 (figures 1 and 2).

---

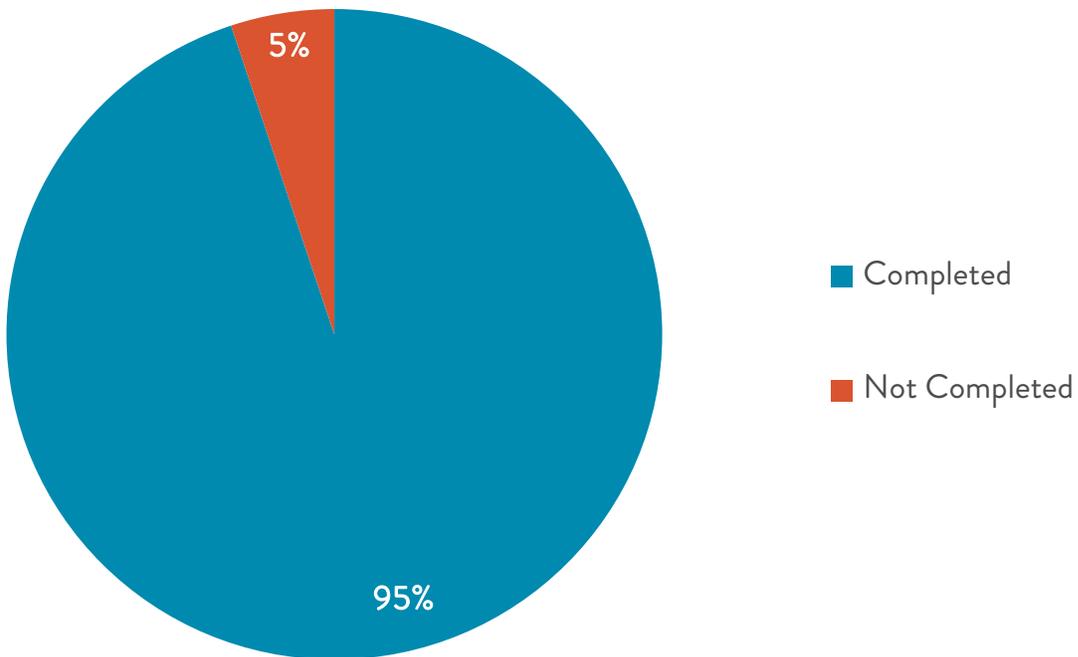
<sup>1</sup>“Primary household” is an indicator in the iam\_household table in Bridges.

<sup>2</sup> Risk assessment completion rates for fiscal year 2013–14 through fiscal year 2016–17 came from a New Hampshire Integrated Assessment SDM management report created by Evident Change in March 2018. Data on reports accepted for assessment that were received in fiscal year 2018–19 were from a separate data extract used for the risk validation analytics. Methods varied slightly between the two analyses.

**Figure 1**  
**Risk Tool Completion Rates**



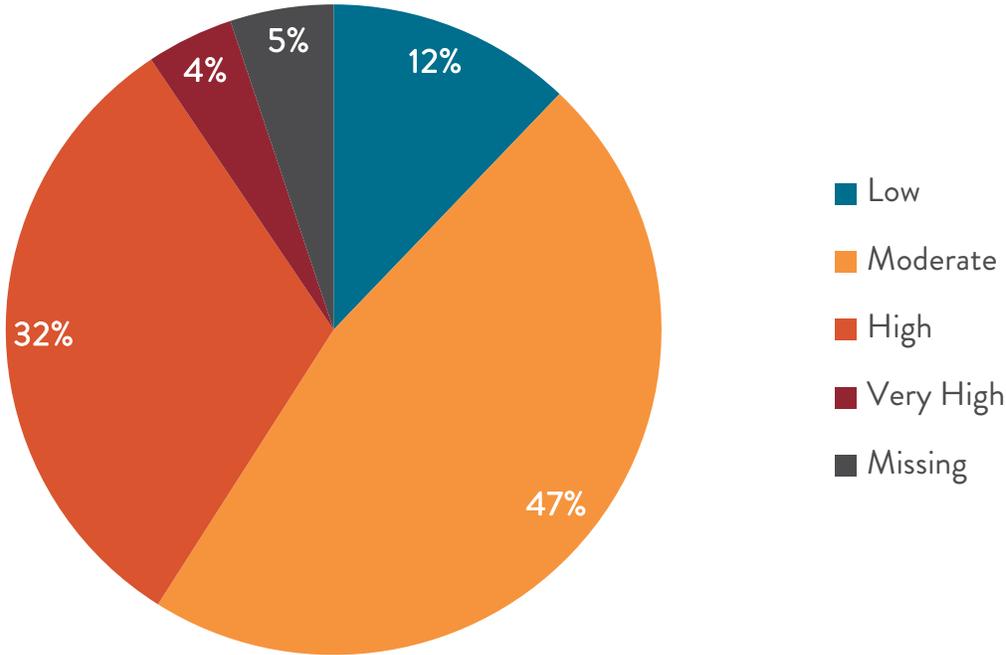
**Figure 2**  
**Risk Tool Completion Rate: Fiscal Year 2018-19**  
**N = 9,663**



SDM risk assessments are used to guide worker decisions regarding whether to open a case for ongoing services following the assessment. Most often, services are recommended for families at the high- and very high-risk levels and for families with dangers still present at the close of the assessment, regardless of allegation findings. In New Hampshire, at the time of this analysis, new family service cases were recommended for all families with founded allegations who were classified as high or very high risk and some families with founded allegations who were classified as moderate risk. Families with unfounded allegations who were classified as moderate, high, or very high risk were recommended for community services. The policy in place at that time limited the use of risk assessment findings when making ongoing family service decisions.<sup>3</sup>

Evident Change examined how often ongoing family services cases were opened by risk level for fiscal year 2018–19. Almost half (47%) of assessments resulted in a moderate-risk classification while only 4% were very high risk. Although 36% of assessments resulted in high or very high-risk classifications (Figure 3), only 12% of those high or very high-risk assessments resulted in new or continued family service cases (not shown).

**Figure 3**  
**Risk Distribution: Fiscal Year 2018–19**



<sup>3</sup> DCYF developed and will be implementing a matrix that considers SDM safety and risk findings as well as allegation findings in guiding decisions about provision of community-based voluntary, DCYF-managed voluntary, or DCYF court-mandated services.

To better understand the relationship between risk level and ongoing case decisions, Evident Change examined family service case decisions by finding and risk level. For investigations with founded allegations, the pattern of case-opening decisions by risk level aligned with SDM recommendations. Specifically, a larger proportion of high and very high-risk investigations with founded allegations resulted in ongoing family services than did low- and moderate-risk founded investigations. As expected based on current policy, only a small percentage of unfounded assessments resulted in new or continued open family service cases.

<b>FAMILY SERVICES CASE OPENED/CONTINUED BY RISK LEVEL AND FINDING<sup>4</sup></b>					
<b>RISK LEVEL</b>	<b>FAMILY SERVICES OPENED/CONTINUED</b>				<b>N</b>
	<b>NO*</b>		<b>YES</b>		
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>	
<b>Founded Investigation</b>					
Low	43	91.5%	4	8.5%	47
Moderate	181	84.2%	34	15.8%	215
High	154	45.7%	183	54.3%	337
Very high	40	19.8%	162	80.2%	202
Missing	12	92.3%	1	7.7%	13
<b>Total Founded</b>	<b>430</b>	<b>52.8%</b>	<b>384</b>	<b>47.2%</b>	<b>814</b>
<b>Unfounded Investigation</b>					
Low	1,111	99.9%	1	0.1%	1,112
Moderate	4,271	99.4%	27	0.6%	4,298
High	2,665	98.6%	38	1.4%	2,703
Very high	192	93.2%	14	6.8%	206
Missing	31	100.0%	0	0.0%	31
<b>Total Unfounded</b>	<b>8,270</b>	<b>99.0%</b>	<b>80</b>	<b>1.0%</b>	<b>8,350</b>

\*Includes a small number of assessments missing family service decisions.

Evident Change presented implementation analysis results to the steering committee at the project kickoff in January 2020. Results favored continuing with the full risk validation study, and the analysis served as a starting point for the remaining risk validation work.

<sup>4</sup> There were 499 assessments missing an overall disposition or with a disposition of “CHINS – Services”; these are excluded from the table.

# EXAMINING CURRENT RISK ASSESSMENT PERFORMANCE

In the next stage of the study, Evident Change researchers looked at the relationship between the current risk classification of a sample of families and child protective services (CPS) outcomes to determine how well the risk classification related to future DCYF involvement. Evident Change conducted this research using information available from Bridges, New Hampshire's child welfare case management system.

The sample consisted of families<sup>5</sup> who were part of an assessment that began between January 1 and December 31, 2017 (the sample period). If a family had more than one assessment during that timeframe, the first such assessment (the "sample assessment") was selected. Evident Change observed, via recorded data, subsequent CPS outcomes for each family during a standardized 18-month follow-up period after the sample assessment.<sup>6</sup> Using a standardized follow-up period ensures that the timeframe for subsequent DCYF involvement for each family in the sample is the same. The outcome measures included new assessments for abuse or neglect, new founded assessments, and new assessments resulting in child placements.

Among the 7,455 sample families remaining after exclusions, families were stratified by race/ethnicity of the youngest alleged child victim with race/ethnicity information recorded on the sample assessment. Hispanic/Latinx and White families and those with unknown or missing races/ethnicities were randomly divided into either a construction sample or a validation sample. Due to small sample sizes, the remaining families were all in the construction sample to ensure ample representation of the different races/ethnicities.

Evident Change used the construction sample (n=3,971) to examine the utility, accuracy, and equity of the current risk assessment to determine its performance and construct revised risk assessments. The validation sample (n=3,484) was used to verify the revised risk assessments' accuracy. Validating the instrument on a separate population provides a more robust estimation of how a risk assessment will perform when used on a population it was not created on.

---

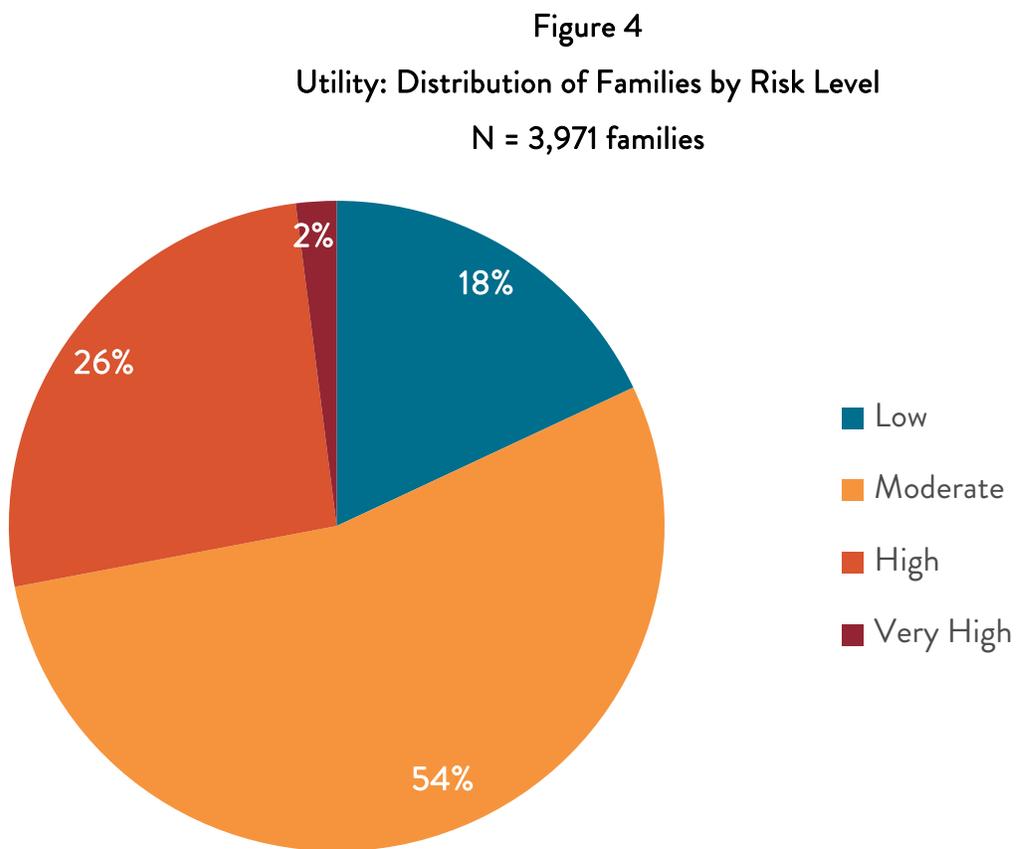
<sup>5</sup> For analysis purposes, Evident Change created families by identifying allegedly maltreated children from each assessment that began between January 1 and December 31, 2017, and combining groups of allegedly maltreated children with common members. For example, if Child A and Child B were allegedly maltreated children identified in one assessment and Child B and Child C were allegedly maltreated children identified in another assessment, all three were combined into one family.

<sup>6</sup> Assessments lasting longer than 60 days were capped at the 60-day mark for the purpose of this study.

## UTILITY

To determine whether the current risk assessment was performing at optimal levels, Evident Change first examined its utility. The risk assessment is meant to identify families classified into higher risk levels so that resources can be targeted toward them in the hope of reducing subsequent CPS involvement. Risk assessments should result in risk classifications that have a meaningful proportion of families at each risk level. This effort, however, depends on agency workload and demand for services. The risk assessment distribution (i.e., how many families are classified as high or very high risk and therefore recommended for ongoing resources and engagement) should reflect this.

More than half (54%) of families in the construction sample were classified as moderate risk on the current risk assessment while 18% were classified as low risk, 26% as high risk, and 2% as very high risk (Figure 4).<sup>7</sup>



<sup>7</sup> The risk validation construction sample distribution differs from the distribution shown in Figure 3. The time periods differ, and the distribution in Figure 4 includes only the first risk assessment for any family during the sample period while the distribution in Figure 3 includes all risk assessments completed during the period.

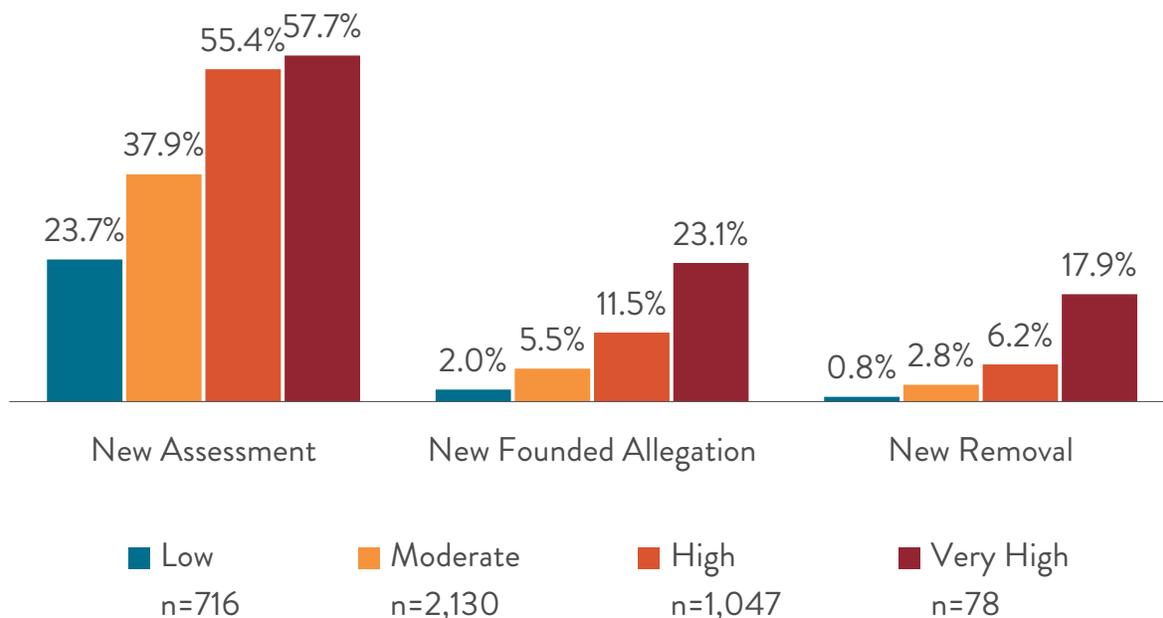
## ACCURACY

Another way to assess the current risk assessment's performance is to examine how well the assessment distinguishes the rate of subsequent CPS involvement during the outcome period. Risk assessments are considered accurate when the proportion of families with an outcome substantially increases with each increase in risk level. Additionally, accurate risk assessments will classify families so that those classified as lower risk have a below-average chance of subsequent involvement while those classified as higher risk have an above-average chance. Moderate-risk families should have outcome rates similar to the outcome average or base rate (i.e., the outcome rate of the entire group regardless of risk level).

Among the sampled families, 40% were involved in a new assessment at least once during the follow-up period, 7% experienced a new assessment that was founded for abuse or neglect, and 4% had at least one child removed as a result of a new assessment (not shown).

The current overall risk assessment classification (based on the higher of the neglect and abuse risk levels) performed accurately at all risk levels. In other words, there was an increase in the outcome rates from low to moderate, moderate to high, and high to very high risk (Figure 5). Note, however, that the new assessment outcome rate among very high-risk families was only slightly higher than that for high-risk families. This may be due to the smaller number of very high-risk families but warrants attention when considering revisions. Additionally, the outcome rates for moderate-risk families were slightly below the base rate for all three outcomes.

**Figure 5**  
**Accuracy: Outcomes by Current Risk Level**  
N = 3,971 families

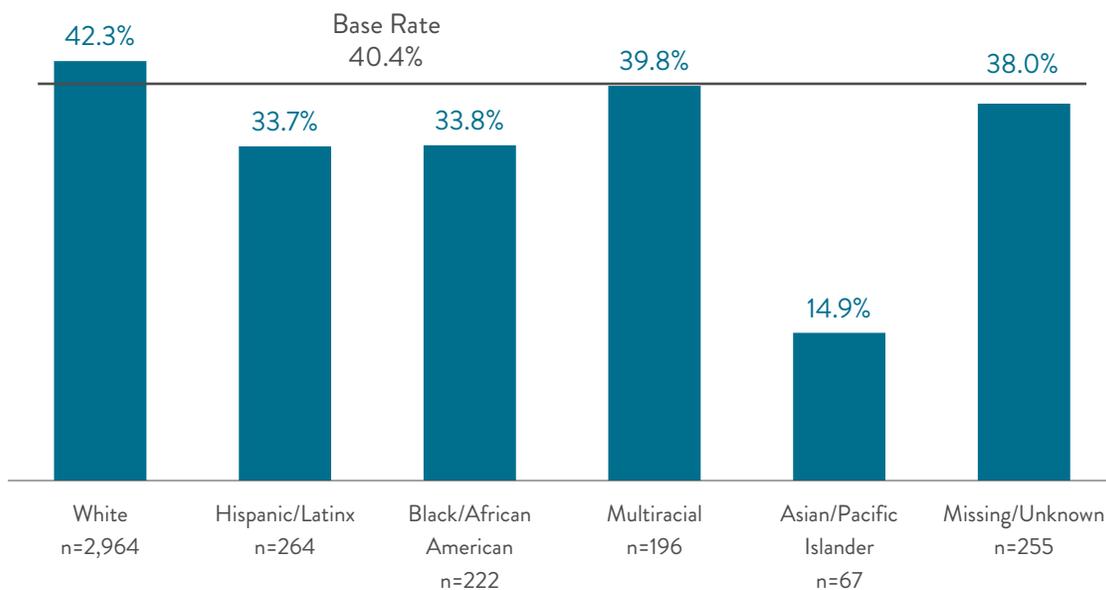


## EQUITY

Evident Change also examined the current risk assessment's equity. The risk assessment should work equitably across families in various subgroups (e.g., region, race/ethnicity). In other words, the risk level should mean the same thing regardless of which groups a family belongs to. For instance, the likelihood of future CPS involvement for families classified as high risk in one group should be similar to the likelihood for families classified as high risk in other groups. The same should be true for all risk levels. Note that equity can be difficult to achieve if base rates differ significantly by subgroup.

Evident Change examined CPS outcome rates and risk assessment classification findings by the race/ethnicity of the youngest child alleged to be maltreated using race/ethnicity information from the sample assessment.<sup>8</sup> The proportion of families with subsequent DCYF involvement varied by subgroup. Assessment outcome rates for Hispanic/Latinx and Black/African American families were similar (around 34%). The new assessment outcome rate for families with unknown or missing race/ethnicity information, multiracial families, and White families was higher, at 38.0%, 39.8%, and 42.3%, respectively; and the rate was much lower for Asian/Pacific Islander families (14.9%). The new assessment outcome rate for White families was above the average for the entire group (Figure 6). Note that these analyses by race/ethnicity exclude American Indian/Alaska Native families, as the sample size does not support drawing conclusions.

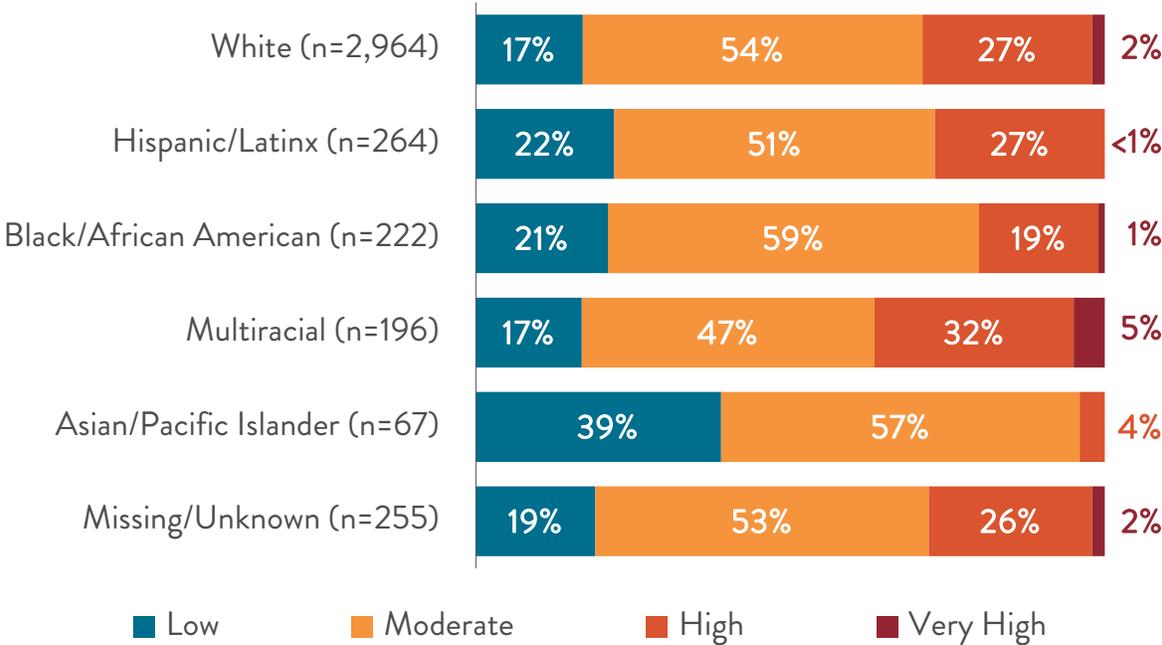
**Figure 6**  
**Equity: New Assessment Outcome by Race/Ethnicity**  
**N = 3,971 families**



<sup>8</sup>Evident Change also examined new assessment outcome rates by finding (founded, incomplete, unfounded) but focused more on the results by race/ethnicity to assess the current assessment's equity.

The distribution of families into the high- and very high-risk groups was similar for White families (29%), Hispanic/Latinx families (28%), and families with unknown or missing race/ethnicity information (28%). The proportion of multiracial families (36%) in the high- and very high-risk groups was higher than all other race/ethnicity groups while the proportion of Black/African American (21%) and Asian/Pacific Islander (5%) families was lower than the other race/ethnicity groups (Figure 7).

**Figure 7**  
**Equity: Current Risk Level Distribution by Race/Ethnicity**  
**N = 3,971 families**

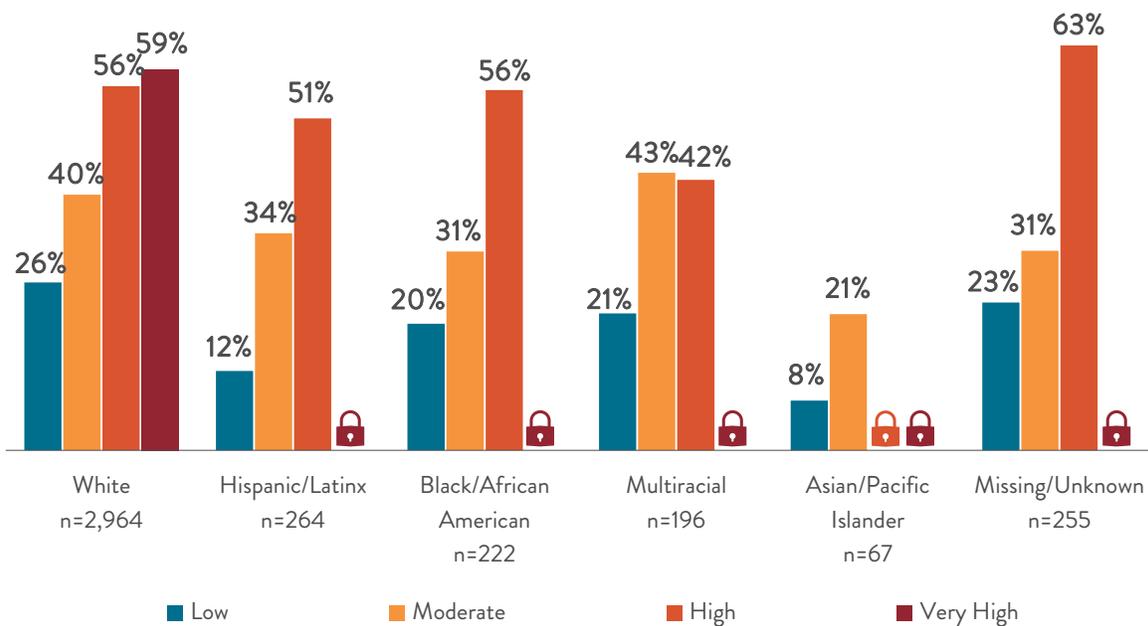


With one exception among the race/ethnicity groups and risk levels shown in Figure 8, an increase in the risk classification corresponded to an increase in the new assessment outcome rates across race/ethnicity subgroups. Of those shown, the exception included the high-risk outcome rate for multiracial families, which was lower than the outcome rate for moderate-risk families in the same group (Figure 8). For the Black/African American, Hispanic/Latinx, missing/unknown, and multiracial subgroups, the number of very high-risk families was small. In addition, fewer than 25 Asian/Pacific Islander families were classified as high risk.

Outcome rates within a risk classification varied across race/ethnicity groups, and some crossed over other risk classification levels. The new assessment outcome rate for moderate-risk Asian/Pacific Islander families was lower than the new assessment outcome rate for low-risk White families and families with missing or unknown race/ethnicity information and was just slightly lower than the new assessment outcome rate for low-risk multiracial families (prior to rounding, the outcome rates were 21.1% for moderate-risk Asian/Pacific Islander families and 21.2% for low risk multiracial families). These findings could be due to the significantly

lower new assessment outcome rates for Asian/Pacific Islander families. These observations are known as “crossover,” meaning that the outcome rates for a risk level for a given subgroup are more similar to the outcome rates for a different risk level for other subgroups. An assessment working equitably for different subgroups should have little or no crossover.

**Figure 8**  
**Equity: New Assessment by Current Risk Level and Race/Ethnicity**  
**N = 3,971 families**



Note: Any subgroup with a sample size less than 25 was too small to report and is indicated by the lock symbol.

## IDENTIFYING AREAS FOR IMPROVEMENT

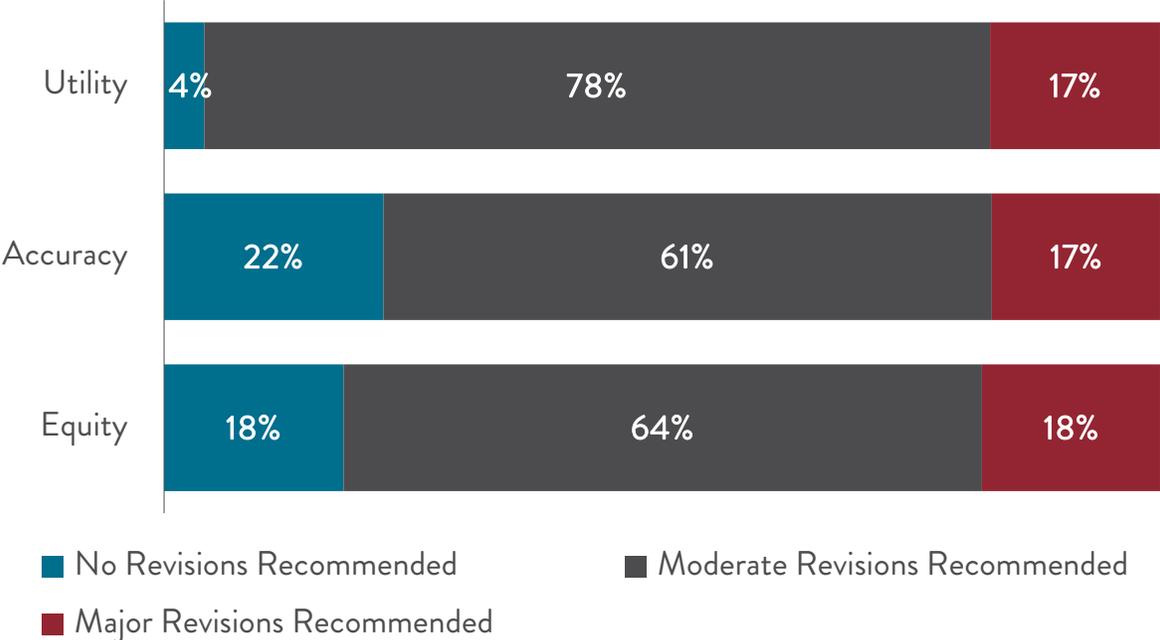
While the current risk assessment classified families with reasonable accuracy, the results suggest several areas for potential improvement. For example, the extremely low proportion of families classified as very high risk paired with the small distinction in new assessment outcome rates between high- and very high-risk families calls into question how meaningful this classification group is for families served by DCYF. Also, the current assessment did not accurately classify multiracial families with respect to the new assessment outcome observed. Low-risk families of each race/ethnicity group should have outcome rates similar to one another, and the same should be true for all other risk classifications. However, a stepwise pattern of increasing outcome rates by risk classification was not evident for multiracial families.

Evident Change presented the data analytics described above to the steering committee during a workgroup session and received feedback as well as votes on the level of revisions needed to improve each performance area (distribution, accuracy, and equity). More than three quarters (78%) of voting steering committee members recommended moderate revisions to the risk assessment tool to improve utility, while 17% recommended major revisions. There was consensus (not shown) on exploring whether the percentage of high- and very high-risk families could be reduced to align more closely with the agency resources available to provide ongoing family services to high- or very high-risk families while maintaining the tool’s accuracy and equity. Additionally, there was a recommendation to examine how a three-level tool (low, moderate, and high risk) would perform, due to the small number of families in the very high-risk group.

Of voters on accuracy issues, 61% recommended moderate revisions to the current risk assessment, 17% recommended major revisions, and 22% recommended no revisions. Voters felt that the new assessment outcome rate for families classified as very high risk was not as high as they would like and was not a large distinction from the rate of high-risk families who experienced a new assessment within 18 months of their sample assessment.

About two thirds (64%) of voters on the question recommended moderate revisions to the current risk assessment to address equity issues, 18% recommended no revisions, and 18% recommended major revisions (Figure 9). Note that steering committee polls were limited in sample size, and should be interpreted in the context of small group sizes.

**Figure 9**  
**Does Current Risk Assessment Need Any Revisions?**



# EXPLORING RISK ASSESSMENT REVISIONS

Using results from the current risk examination and feedback from the steering committee, Evident Change began exploring how revisions to the current tool could improve performance on one or more of the criteria described above. This process began with examining the bivariate relationships (i.e., correlations between individual items and outcomes and cross-tabulations of items with outcome measures) between the family assessment characteristics that workers observed and recorded in Bridges at the time of the sample assessment and the family's subsequent DCYF involvement. This includes SDM risk assessment items recorded by workers, allegations of abuse or neglect made at the time of the sample incident, and families' DCYF involvement prior to the sample incident. Individual items were selected for inclusion in the revised assessments based on their statistical association with outcomes. Items that demonstrated a significant statistical association with any CPS outcome were included in regression analyses. Stepwise regression analyses were used to identify items with the strongest relationships when similar items were being considered.

Evident Change researchers used these bivariate and multivariate relationships to construct hundreds of revised risk assessment models based on different item combinations, item scores, and cut points for each index (neglect and abuse). Evident Change then tested the models against construction sample outcomes to evaluate their classification capabilities. Evident Change also examined distribution and equity for the construction sample. Finally, risk assessment revisions were applied to the validation sample to examine classification findings with a different sample to ensure that the assessments were not overfitted to the construction sample on which they were developed.<sup>9</sup>

After examining the different models, Evident Change identified two potential revised risk assessments, described in the following as Model A and Model B. Both assessments improved tool performance in one or more of the areas described above and incorporated feedback provided by the steering committee. These revised risk assessment models were presented to the risk steering committee. The presentation compared the revised models to each other as well as to the current risk assessment.

## COMPARING UTILITY

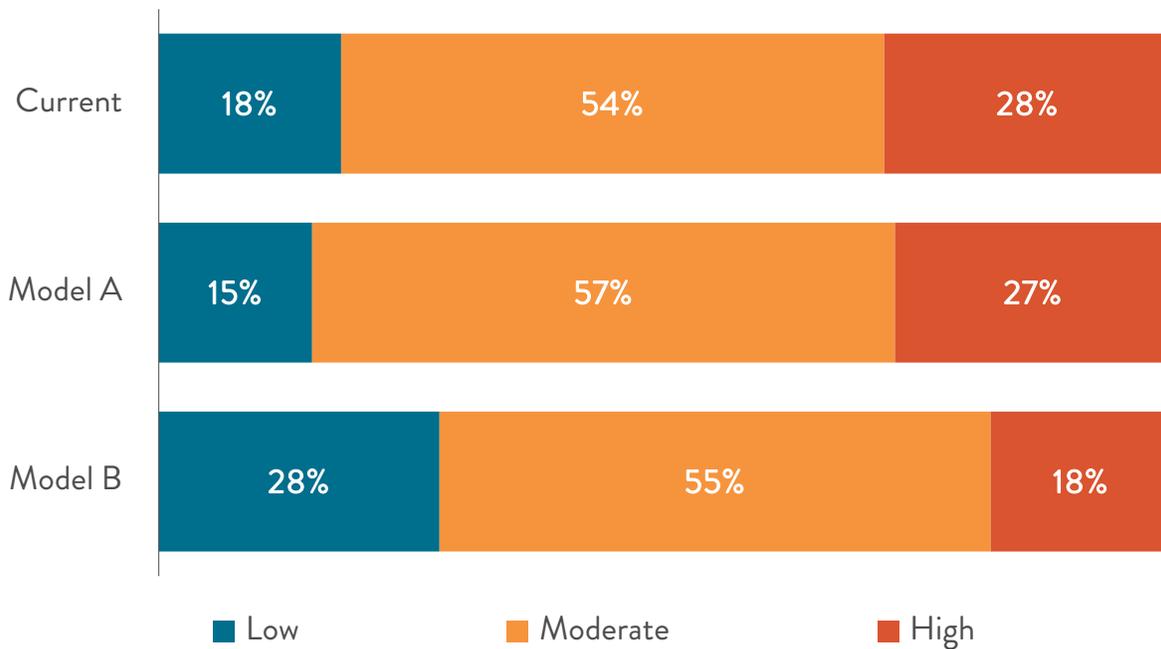
The data analytics described above supported changing to a three-level risk assessment. For comparison purposes and because the recommendation to provide ongoing services is the same for both, high and very high risk groups from the current risk assessment were combined. Model A and B are also three-level risk assessments. The scored risk level distributions of Model A and the current risk assessment were similar,

---

<sup>9</sup> All three models performed well when tested on the validation sample, suggesting that each one, including the final revised model selected by the group, will work well on another group of assessments.

although Model A classified about 3% fewer families into the low-risk and 3% more into the moderate-risk group. Model B, on the other hand, classified fewer families into the high-risk group and more families into the low-risk group than Model A and the current risk assessment (Figure 10).

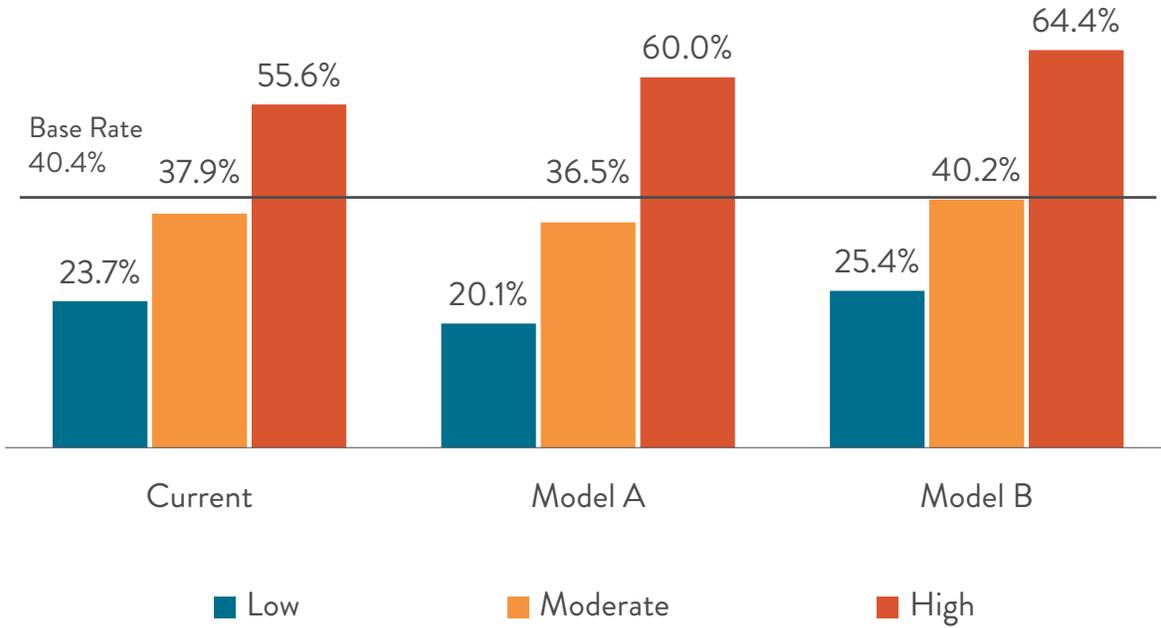
**Figure 10**  
**Utility: Risk Level Distribution by Version**  
N = 3,971 families



## COMPARING ACCURACY

Even though Model A's scored risk level distribution was similar to that of the current risk assessment, the low-risk outcome rates for Model A were lower and the high-risk outcome rates were higher than those for the current risk assessment, providing greater distinction between outcomes for families in the low- and high-risk groups. Families classified as low risk in Model B experienced slightly higher outcome rates than families classified as low risk in the other two models; but in Model B, families classified as high risk also experienced the highest outcome rates. Thus, Model B still had a greater distinction between outcomes for families in the low- and high-risk groups than the current risk assessment. Outcome rates for moderate-risk families were similar to the base rate for Model B and were lower than the base rate for both the current risk assessment and Model A (figures 11, 12, and 13).

**Figure 11**  
**Accuracy: Assessment Outcome by Version**  
**N = 3,971 families**



**Figure 12**  
**Accuracy: Founded Allegation Outcome by Version**  
**N = 3,971 families**

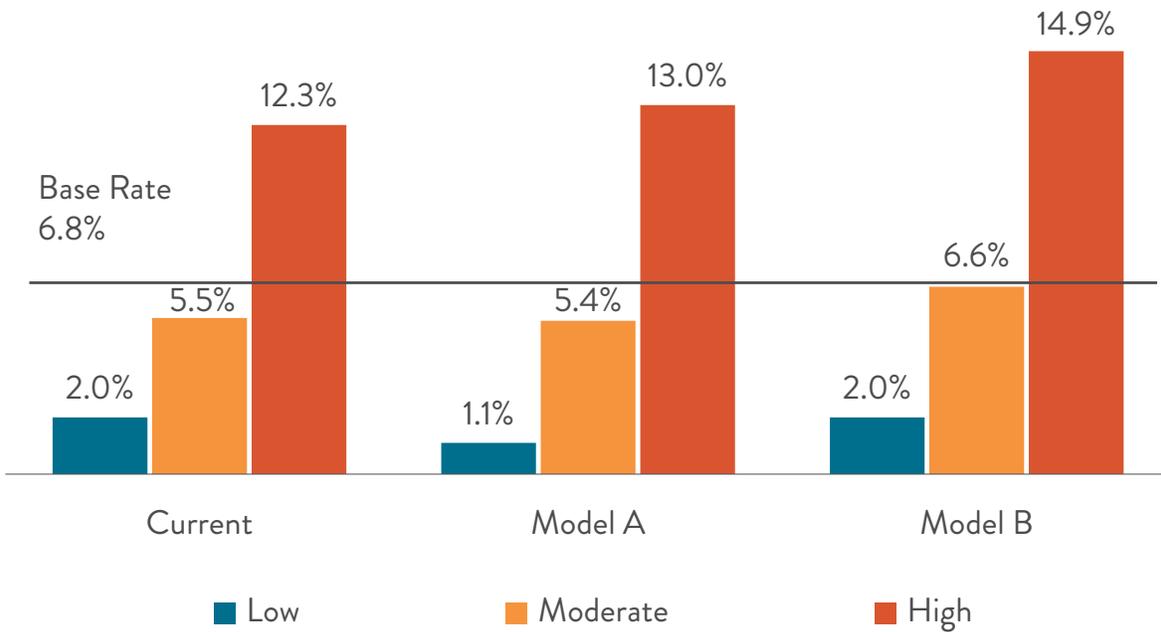
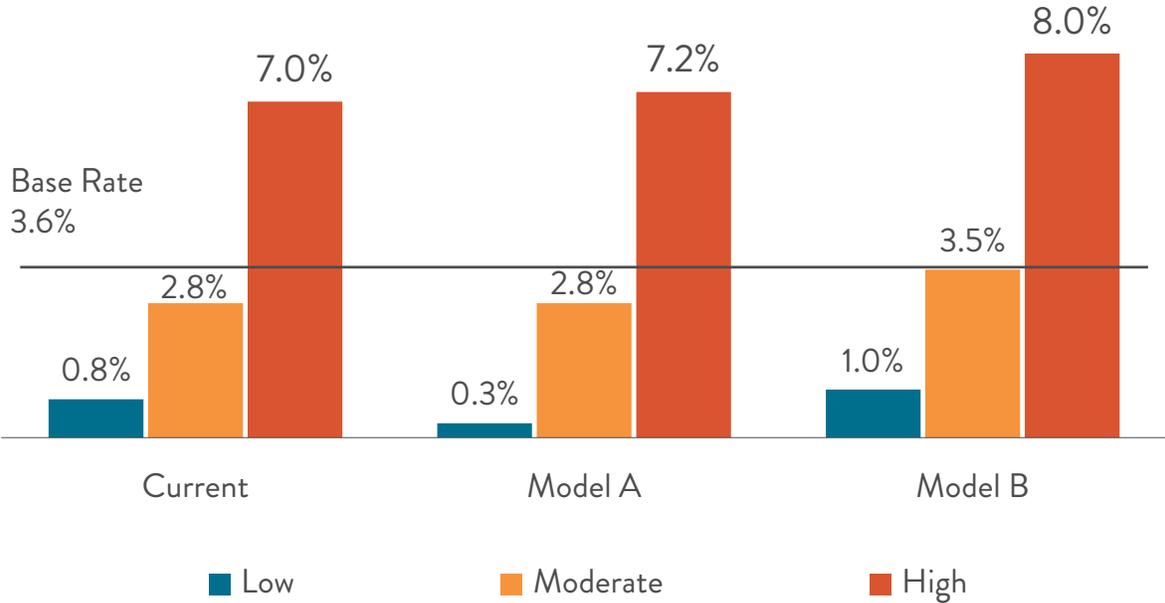


Figure 13  
 Accuracy: Removal Outcome by Version  
 N = 3,971 families



**COMPARING EQUITY**

Both Model A and Model B performed better overall in general than the current risk assessment. They also produced classifications with more similar proportions and equitable performance across race/ethnicity subgroups than the current assessment. For example, the revised assessments classified more similar proportions into the high-risk level across subgroups than the current assessment did, particularly among multiracial families (figures 14, 15, and 16).

Figure 14

Equity Current Model: Risk Level Distribution by Race/Ethnicity

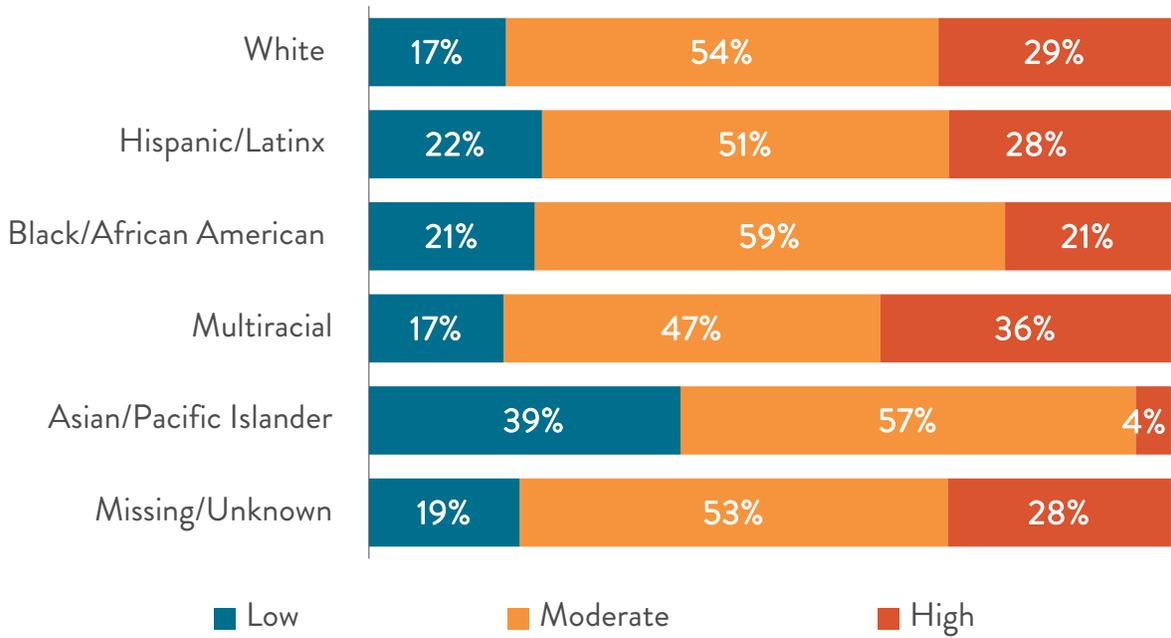


Figure 15

Equity Model A: Risk Level Distribution by Race/Ethnicity

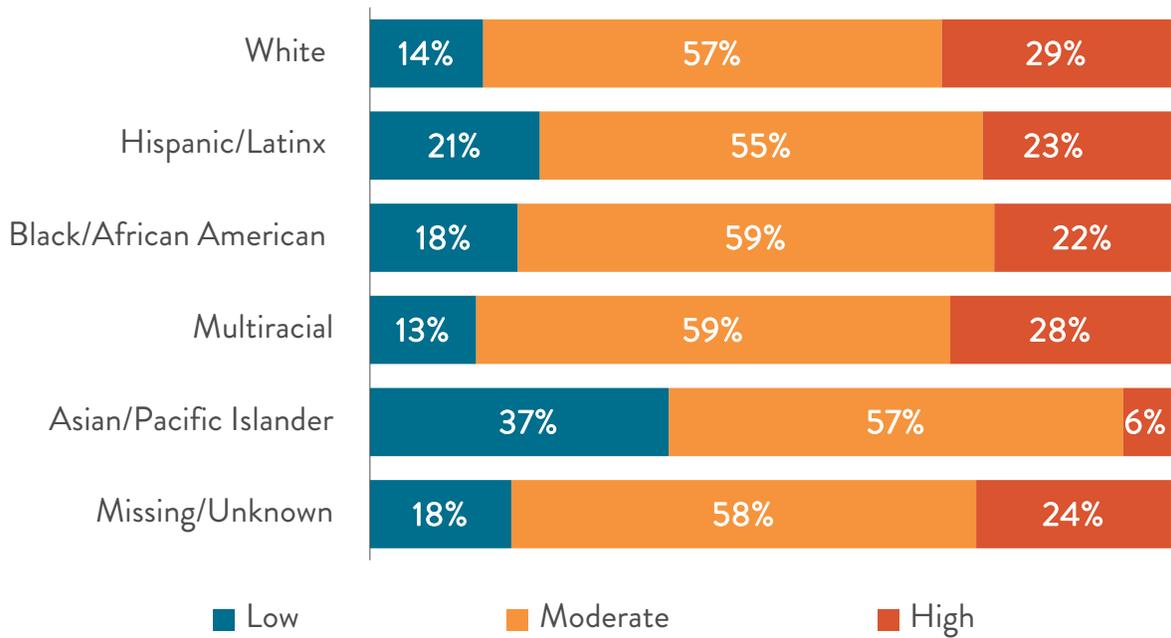
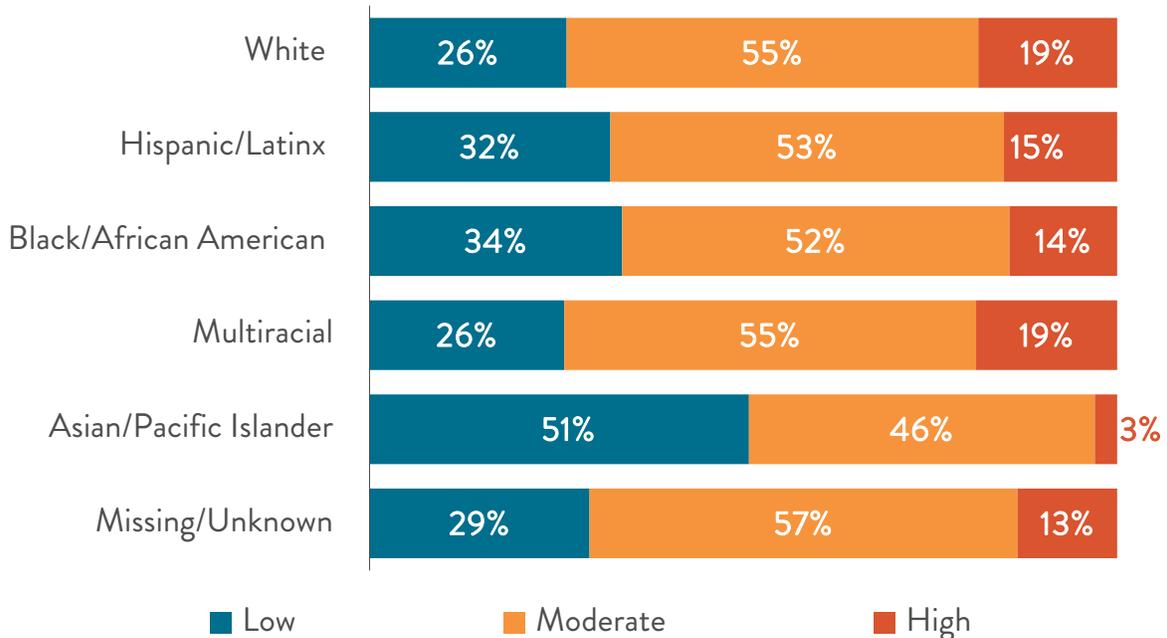


Figure 16

Equity Model B: Risk Level Distribution by Race/Ethnicity



After changing the current risk assessment into a three-level assessment, an increase in risk classification corresponded to an increase in new assessment outcome rates for all three models for all but one race/ethnicity group. Additionally, when the current model was examined as a three-level assessment, the outcome rates for moderate- and high-risk multiracial families was nearly the same.

The distinction in the new assessment outcome rates between multiracial families classified as moderate and high risk was greater for Model A and B than the current risk assessment. In addition, both models A and B had greater distinction between outcomes for families in the low- and high-risk groups across all race/ethnicity groups for which results are shown than the current risk assessment. While moderate-risk Black/African American families had higher outcome rates than low-risk Black/African American families across all models, the difference was smaller for Model B than Model A and the current risk assessment. Although there was some crossover between low- and moderate-risk groups, no models had crossover between moderate- and high-risk groups (figures 17, 18, and 19). Crossover between moderate- and high-risk groups is more concerning because the decision to provide ongoing services is guided by a risk classification of “high.”

Figure 17

Equity Current Model: New Assessment by Risk Level and Race/Ethnicity

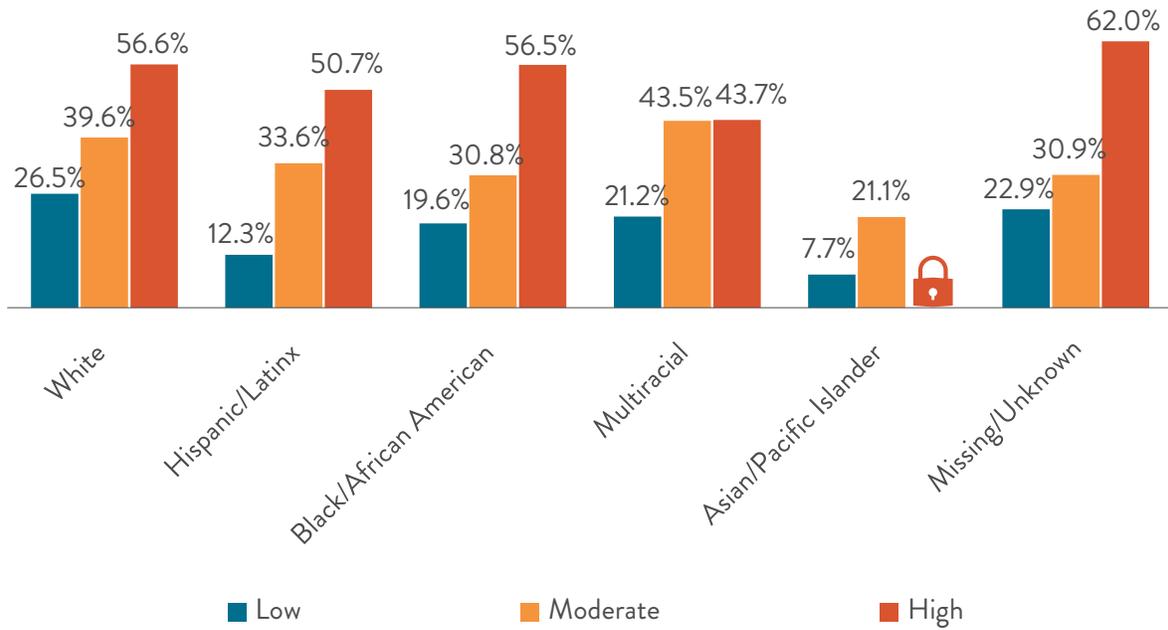


Figure 18

Equity Model A: New Assessment by Risk Level and Race/Ethnicity

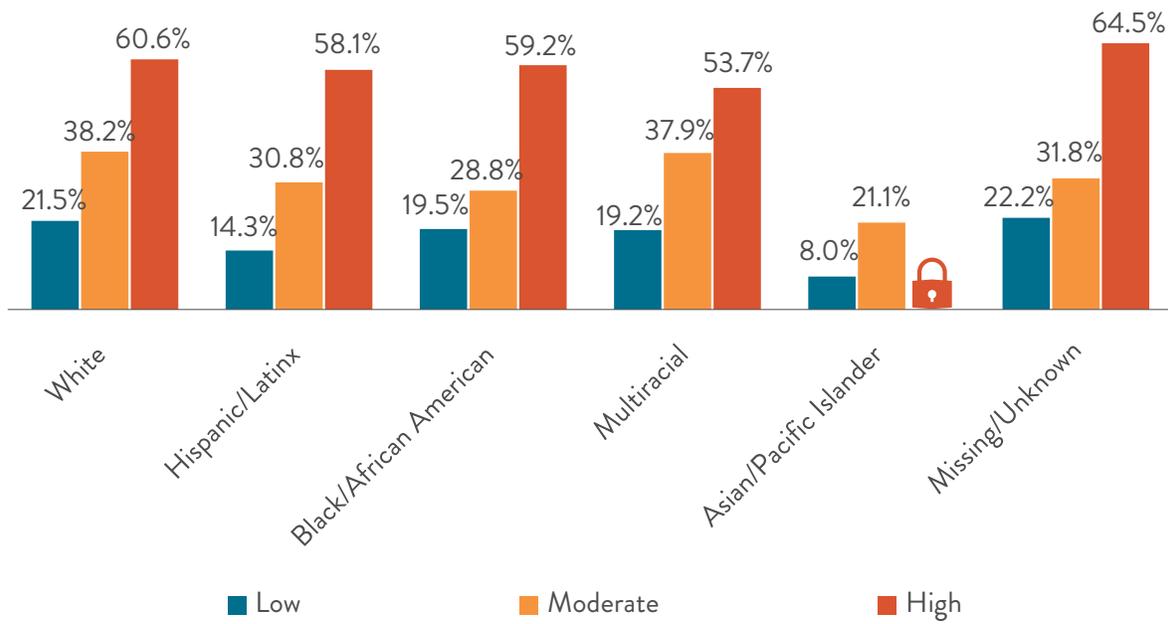
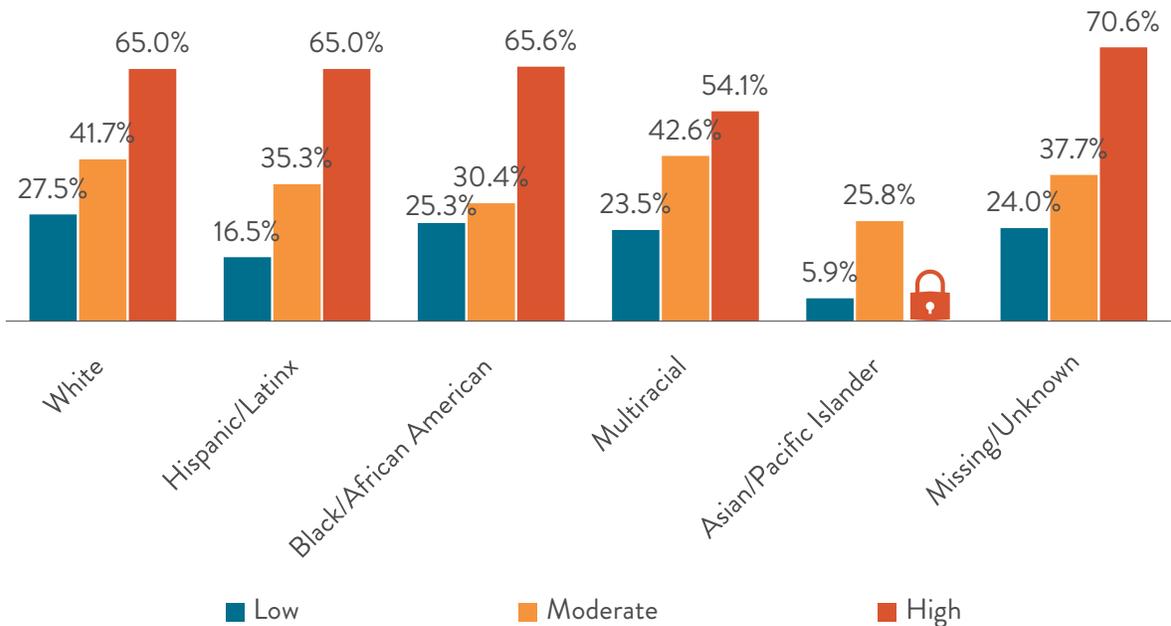


Figure 19

Equity Model B: New Assessment by Risk Level and Race/Ethnicity

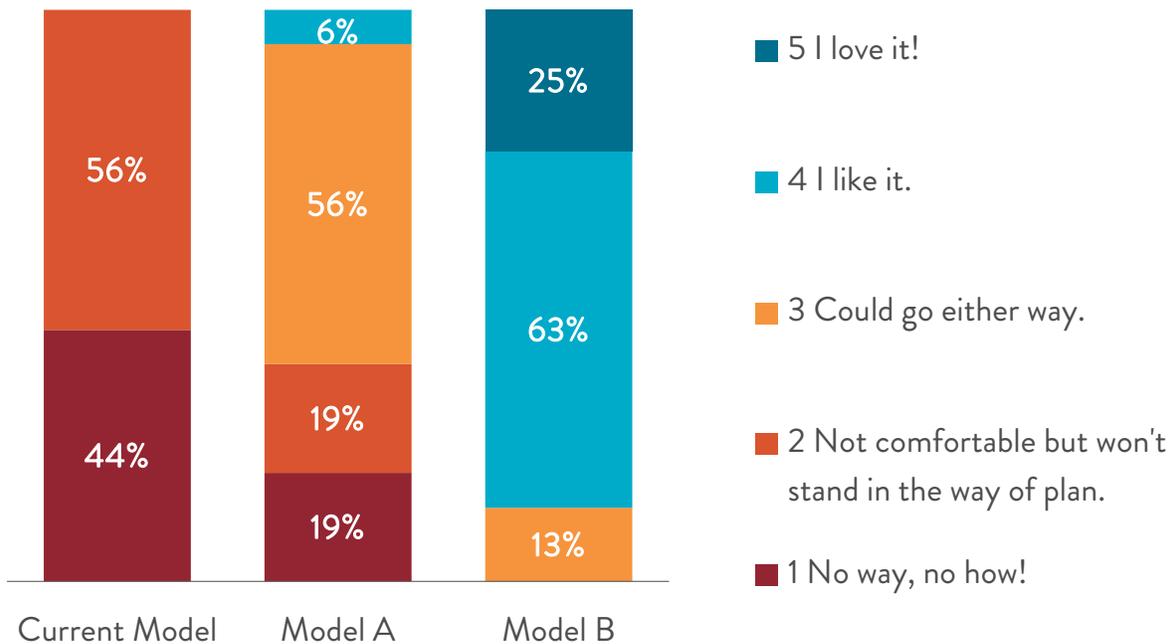


## SELECTING A MODEL

Evident Change presented the data described above to the steering committee at the last workgroup session. After comparing the models and discussing in smaller groups and as a whole, members were asked to rate each model using a five-point “gradients of agreement” scale ranging from “No way, no how” to “I love it” (Figure 20). The current model received entirely “No way, no how” and “Not comfortable” votes; many voters chose the middle options for Model A; and Model B received a high percentage of “I like it” and “I love it” votes. Based on these votes and discussion with DCYF administration, DCYF decided to move forward with Model B for their new SDM risk assessment. Note that steering committee polls were limited in sample size and should be interpreted in the context of small group sizes.

Figure 20

What Is Your Opinion on Implementing Each Model?



In summary, although the current DCYF risk assessment performed adequately when classifying families by the risk of future CPS involvement, Model B demonstrated improved accuracy and equity. Model B also offered better alignment between the risk distribution of families and the agency’s workload and service availability. Finally, there were higher outcome rates for high-risk families in Model B and better distinction between new assessment outcome rates for moderate- and high-risk multiracial families than on the current risk assessment. Although the low-risk outcome rates for Model B were higher than the low-risk outcome rates for the other models, the distinction between outcomes for families in the low- and high-risk groups was greater than that for the current risk assessment and were similar to that for Model A. Outcome rates for Model B were also closest to the base rate. The smaller distinction between the new assessment outcome rate for low- and moderate-risk Black/African American families should be monitored as DCYF moves forward with Model B.

See the appendix for the revised Model B risk assessment item analysis.

# NEXT STEPS: IMPLEMENTING THE REVISED RISK ASSESSMENT

Following the steering committee decision, a workgroup of DCYF and Evident Change staff reviewed and finalized the revised assessment. Next steps following the validation include inter-rater reliability testing to ensure that the assessment produces consistent results across staff and circumstances and field testing to solicit feedback from line staff about utility and ease of using the new assessment in the field. Any recommended changes based on inter-rater reliability testing or field testing will be made to the final assessment tool; once finalized, the risk assessment will be built into the agency's new Granite Families system.

To support implementation of and fidelity to the SDM tools, Evident Change strongly recommends including risk assessment analytics in any continuous quality improvement processes in place or in development. Specifically, it will be important to monitor distribution, accuracy, and equity over time. Changes in tool functioning may indicate the need for revisions through a new risk calibration or validation study. Additionally, using case reading or other methods to examine and provide feedback and coaching on worker use and application of the risk assessment in practice can strengthen implementation.

# APPENDIX: MODEL B RISK ASSESSMENT ITEM ANALYSIS

TABLE A1

MODEL B NEGLECT INDEX ITEM ANALYSIS

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME								
	n	%	NEW ASSESSMENT FOR NEGLECT				NEW FOUNDED NEGLECT ALLEGATION				
			n	%	r	pVALUE	n	%	r	pVALUE	
<b>TOTAL</b>	3,971	100.0%	1,388	35.0%				247	6.2%		
<b>Current assessment is for neglect</b>					.102	<.001				.074	<.001
No	973	24.5%	257	26.4%				30	3.1%		
Yes	2,998	75.5%	1131	37.7%				217	7.2%		
<b>Prior assessments for neglect</b>					.282	<.001				.150	<.001
None	2,289	57.6%	579	25.3%				87	3.8%		
One	811	20.4%	307	37.9%				51	6.3%		
Two or three	625	15.7%	329	52.6%				68	10.9%		
Four or more	246	6.2%	173	70.3%				41	16.7%		
<b>Prior assessments for abuse</b>					.140	<.001				.057	<.001
None	2,927	73.7%	906	31.0%				158	5.4%		
One or more	1,044	26.3%	482	46.2%				89	8.5%		
<b>Household has previously received family services</b>					.153	<.001				.119	<.001
No	3,631	91.4%	1,188	32.7%				194	5.3%		
Yes	340	8.6%	200	58.8%				53	15.6%		
<b>Number of children involved in the child abuse or neglect incident</b>					.140	<.001				.042	.009
One	2,670	67.2%	819	30.7%				148	5.5%		
Two or three	1,168	29.4%	492	42.1%				87	7.4%		
Four or more	133	3.3%	77	57.9%				12	9.0%		

TABLE A1

## MODEL B NEGLECT INDEX ITEM ANALYSIS

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME							
			NEW ASSESSMENT FOR NEGLECT				NEW FOUNDED NEGLECT ALLEGATION			
	n	%	n	%	r	p VALUE	n	%	r	p VALUE
<b>TOTAL</b>	3,971	100.0%	1,388	35.0%			247	6.2%		
<b>Current housing</b>					.058	<.001			.035	.029
Neither of the below applies	3,841	96.7%	1,323	34.4%			233	6.1%		
Current housing is physically unsafe and/or family is homeless at time of alleged incident	130	3.3%	65	50.0%			14	10.8%		
<b>Primary caregiver mental health</b>					.120	<.001			.075	<.001
No problem	3,167	79.8%	1,016	32.1%			168	5.3%		
Current or past problem	804	20.2%	372	46.3%			79	9.8%		
<b>Either caregiver substance abuse</b>					.141	<.001			.126	<.001
No problem	2,654	66.8%	802	30.2%			108	4.1%		
Current or past substance abuse	1,317	33.2%	586	44.5%			139	10.6%		
<b>Age of youngest child in the home</b>					.063	<.001			.073	<.001
Age 2 or older	3,131	78.8%	1,046	33.4%			166	5.3%		
Under age 2	840	21.2%	342	40.7%			81	9.6%		
<b>Characteristics of children in household</b>										
<i>Medically fragile or failure to thrive</i>					.035	.027			.005	.745
No	3,916	98.6%	1,361	34.8%			243	6.2%		
Yes	55	1.4%	27	49.1%			4	7.3%		
<i>Mental health or behavioral problem</i>					.028	.082			-.008	.615
No	3,364	84.7%	1,157	34.4%			212	6.3%		
Yes	607	15.3%	231	38.1%			35	5.8%		

**TABLE A1**

**MODEL B NEGLECT INDEX ITEM ANALYSIS**

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME							
			NEW ASSESSMENT FOR NEGLECT				NEW FOUNDED NEGLECT ALLEGATION			
	n	%	n	%	r	pVALUE	n	%	r	pVALUE
<b>TOTAL</b>	3,971	100.0%	1,388	35.0%			247	6.2%		
<b>Adult household relationships in the past year</b>					.039	.013			.031	.054
No violence	3,383	85.2%	1,156	34.2%			200	5.9%		
Violence	588	14.8%	232	39.5%			47	8.0%		
<b>Primary caregiver has a history or abuse or neglect as a child</b>					.071	<.001			.082	<.001
No	3,707	93.4%	1,262	34.0%			211	5.7%		
Yes	264	6.6%	126	47.7%			36	13.6%		

**TABLE A2**

**MODEL B ABUSE INDEX ITEM ANALYSIS**

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME							
			NEW ASSESSMENT FOR ABUSE				NEW FOUNDED ABUSE ALLEGATION			
	n	%	n	%	r	pVALUE	n	%	r	pVALUE
<b>TOTAL</b>	3,971	100.0%	731	18.4%			41	1.0%		
<b>Current assessment is for abuse</b>					.077	<.001			.028	.082
No	2,459	61.9%	395	16.1%			20	0.8%		
Yes	1,512	38.1%	336	22.2%			21	1.4%		
<b>Prior assessments for abuse</b>					.182	<.001			.063	<.001
None	2,927	73.7%	419	14.3%			22	0.8%		
One	675	17.0%	183	27.1%			7	1.0%		
Two or more	369	9.3%	129	35.0%			12	3.3%		

TABLE A2

## MODEL B ABUSE INDEX ITEM ANALYSIS

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME									
	n	%	NEW ASSESSMENT FOR ABUSE				NEW FOUNDED ABUSE ALLEGATION					
			n	%	r	pVALUE	n	%	r	pVALUE		
<b>TOTAL</b>	3,971	100.0%	731	18.4%					41	1.0%		
<b>Household has previously received family services</b>					.072	<.001				.026	.101	
No	3,729	93.9%	660	17.7%					36	1.0%		
Yes	242	6.1%	71	29.3%					5	2.1%		
<b>Prior injury to a child resulting from abuse or neglect</b>					.037	.019				.029	.066	
No	3,864	97.3%	702	18.2%					38	1.0%		
Yes	107	2.7%	29	27.1%					3	2.8%		
<b>Primary caregiver's assessment of incident</b>					.058	<.001				.051	.001	
Neither blames child for or justifies maltreatment	3,827	96.4%	692	18.1%					37	1.0%		
Blames child for or justifies maltreatment	130	3.3%	30	23.1%					2	1.5%		
Both	14	0.4%	Not shown due to small sample size									
<b>Primary caregiver's assessment of incident</b>												
<i>Blames the child for maltreatment</i>					.057	<.001				.023	.139	
No	3,899	98.2%	706	18.1%					39	1.0%		
Yes	72	1.8%	25	34.7%					2	2.8%		
<i>Justifies maltreatment</i>					.032	.044				.053	.001	
No	3,885	97.8%	708	18.2%					37	1.0%		
Yes	86	2.2%	23	26.7%					4	4.7%		
<b>Adult household relationships in the past year</b>					.029	.069				.007	.681	
No violence	3,383	85.2%	607	17.9%					34	1.0%		
Violence	588	14.8%	124	21.1%					7	1.2%		

TABLE A2

MODEL B ABUSE INDEX ITEM ANALYSIS

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME							
	n	%	NEW ASSESSMENT FOR ABUSE				NEW FOUNDED ABUSE ALLEGATION			
			n	%	r	pVALUE	n	%	r	pVALUE
TOTAL	3,971	100.0%	731	18.4%			41	1.0%		
<b>Primary caregiver characteristics</b>										
<i>Provides insufficient emotional or psychological support</i>					.027	.087			.034	.031
No	3,880	97.7%	708	18.2%			38	1.0%		
Yes	91	2.3%	23	25.3%			3	3.3%		
<i>Employs excessive or inappropriate discipline</i>					.032	.045			.015	.342
No	3,932	99.0%	719	18.3%			40	1.0%		
Yes	39	1.0%	12	30.8%			1	2.6%		
<i>Domineering</i>					.059	<.001			.017	.284
No	3,936	99.1%	716	18.2%			40	1.0%		
Yes	35	0.9%	15	42.9%			1	2.9%		
<b>Primary caregiver has a history of abuse or neglect as a child</b>					.040	.011			.023	.152
No	3,707	93.4%	667	18.0%			36	1.0%		
Yes	264	6.6%	64	24.2%			5	1.9%		
<b>Either caregiver substance abuse</b>					-.009	.576			.029	.072
No problem	2,654	66.8%	495	18.7%			22	0.8%		
Current or past substance abuse	1,317	33.2%	236	17.9%			19	1.4%		
<b>Characteristics of children in household</b>										
<i>Delinquency or CHINS</i>					.036	.023			-.010	.529
No	3,794	95.5%	687	18.1%			40	1.1%		
Yes	177	4.5%	44	24.9%			1	0.6%		
<i>Developmental or learning disability</i>					.051	.001			.041	.010
No	3,576	90.1%	635	17.8%			32	0.9%		
Yes	395	9.9%	96	24.3%			9	2.3%		

**TABLE A2**

**MODEL B ABUSE INDEX ITEM ANALYSIS**

ITEM/RESPONSE	ITEM DISTRIBUTION		18-MONTH OUTCOME									
	n	%	NEW ASSESSMENT FOR ABUSE				NEW FOUNDED ABUSE ALLEGATION					
			n	%	r	pVALUE	n	%	r	pVALUE		
<b>TOTAL</b>	3,971	100.0%	731	18.4%					41	1.0%		
<i>Mental health or behavioral problem</i>					.071	<.001					.033	.039
No	3,364	84.7%	580	17.2%					30	0.9%		
Yes	607	15.3%	151	24.9%					11	1.8%		
<b>Number of children involved in the child abuse or neglect incident</b>					.054	.001					-.014	.377
One, two, or three	3,749	94.4%	671	17.9%					40	1.1%		
Four or more	222	5.6%	60	27.0%					1	0.5%		