CASE STUDY
PARTICIPATORY SDM®
RISK VALIDATION IN NEW HAMPSHIRE
OCTOBER 2022
AUTHORS

Chris Scharenbroch  
Sierra Fischer*  
Cayley Farrell  
Andrea Bogie  
Erin Wicke Dankert

*Texas Department of Family and Protective Services

ACKNOWLEDGMENTS

Evident Change gratefully acknowledges the generous contributions of the risk steering committee. We thank the New Hampshire Division for Children, Youth, and Families, in particular Kathleen Grondine and Joseph Ribsam, for their partnership. We also recognize the following Evident Change staff, who committed their time and expertise to conducting the risk validation study and creating associated materials.

New Hampshire Division for Children, Youth, and Families  
Joseph E. Ribsam, Director  
Kathleen Grondine, QI Administrator II  
Kristin Booth, Senior Planner  
Jennifer Ross-Ferguson, Field Administrator  
Beverly Dubiel, CPS Supervisor  

Evident Change  
Sarah Beach  
Sue Breckenridge  
Erin Espinosa, PhD  
Raelene Freitag, PhD  
Debra Greene  
Erin Hanusa  
Lucy Hodgman  
Emerson Ives  
Dayana Kupisk, PhD  
Dave Lee  
Mike Malloy  
Emily Ramasamy

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 and @Evident_Change on Twitter.

© 2022 Evident Change
PROJECT APPROACH AND GOALS

In 2020, Evident Change collaborated with the New Hampshire Division for Children, Youth, and Families (DCYF) to validate its actuarial risk assessment, which had been used since 2011. (The agency used a different version of the Structured Decision Making® [SDM] risk assessment beginning in 2001). Risk validation studies rely on extensive data analyses to identify risk factors to best classify families by their likelihood of future child welfare system involvement. How well a risk assessment maximizes this likelihood for families classified at higher risk levels is a measure of accuracy.

Growing knowledge about the limitations and potentially negative impact of using data-driven applications such as risk assessments has led research teams to create a more inclusive process for understanding the trade-offs inherent in risk assessment implementation and validation. This type of process seeks to balance a risk assessment’s statistical accuracy with an imperative to offer an equitable and useful application in practice. Fundamental to this process is establishing a shared value proposition for how the risk assessment should work. To do this, a broad group of stakeholders with diverse perspectives needs to help design, validate, and implement the risk assessment.

Evident Change and DCYF used a collaborative, stakeholder-informed approach to conduct a participatory risk validation of the SDM® risk assessment in early 2020. The approach borrows principles of human-centered design (Freckmann et al., 2020) and action research (Streubert & Carpenter, 2011) to center attention on those who are most impacted, to seek action to improve practice, and to study the effect of a practice improvement. Including and trusting impacted individuals and communities has been shown to increase empathy, generate new ideas, and be necessary to promote equity in implementation science (Metz et al., 2021).

Human-centered design puts the onus of design and development on stakeholders. This movement forces reflection on how assessments, either standardized or locally customized, work toward agency and community goals and alleviate or perpetuate systematic biases. This approach often involves mixed methods—supplementing data analytics with qualitative information (from surveys, focus groups, observation)—and consensus-building activities, such as workgroups, to refine assessments, definitions, and policy (Saxena et al., 2020). Human-centered design calls for transparency, internal and independent ethical reviews, and human-led design choices. Human-centered design also supports mechanisms to ensure that design choices are working as intended, such as continuous quality improvement.

Risk validation can similarly benefit from these human-centered design features. The goal of using a participatory risk validation approach such as the one used in New Hampshire is to inform empirical analysis with feedback and direction from those who use the tool in practice, who develop its policy, and/or who are impacted by its findings.
Goals for the New Hampshire risk validation study included the following.

- Engage a diverse stakeholder group in every step of the risk validation process to ensure that the evaluation of the state’s current risk assessment and any changes to the assessment meet shared values around equity, accuracy, and utility.
- Understand how the current risk assessment is performing across three of the SDM system 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 did not meet stakeholders’ and Evident Change’s expectation for performance across these three principles, model revisions using available data and explore with the stakeholder group how revisions to the risk assessment improve utility, accuracy, and/or equity.
- Collaboratively recommend a revised risk assessment to DCYF leadership.

The validation process conducted by Evident Change and DCYF was broken into three stages: shared learning, evaluating the current assessment, and evaluating revised models.

**RISK VALIDATION STEERING COMMITTEE**

DCYF convened a steering committee composed of staff at all levels, DCYF licensed alcohol and drug counselors, domestic violence partners, a representative from the Office of Child Advocate, and parent partners with previous DCYF involvement. 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.

The steering committee engaged with Evident Change once for each stage of the project. The first meeting occurred onsite in New Hampshire as part of the SDM update project kickoff. The second and third meetings of the steering committee were held virtually. Prior to the second two meetings Evident Change developed and distributed an agenda outlining the topic and objectives and a worksheet that provided data findings and questions to guide small- and large-group discussions. Copies of the agendas and worksheets can be found in the appendix of this report.

Each steering committee meeting involved a presentation of data followed by small- and large-group discussions. Meetings involved multiple opportunities for steering committee members to engage in the information and offer their expertise. Throughout each session, Evident Change used polls to gather
consensus on multiple decisions and identify areas where more discussion was required before making a final decision.

**RISK FINDINGS BY PROJECT STAGE**

**STAGE 1: SHARED LEARNING**

To support the participatory risk validation, Evident Change and DCYF analyzed the policy, practice, and administrative data to learn how the risk assessment was currently being used. Researchers from Evident Change conducted a comprehensive empirical analysis adhering to a rigorous and established methodology previously used to validate the SDM risk assessment. The empirical analysis set a foundation for shared understanding throughout the entire validation process. It examined 7,455 families who came to DCYF and had a risk assessment completed between January 1 and December 31, 2017. All available system information was aggregated and analyzed to describe families with as much detail as possible. This included demographic and trend profiles and an examination of the statistical relationship between the risk assessment items and outcomes of subsequent DCYF involvement. (See the appendix for sample and model analytics.) The empirical analysis was complemented by a policy and practice analysis of when, how, and for which families a risk assessment should be completed.

Data analysis and findings from the empirical and policy analysis were presented and discussed with the steering committee throughout the validation. Data visualizations and concepts were designed to facilitate conversations among a diverse group of stakeholders. Evident Change intentionally used the same visualizations for utility, accuracy, and equity for every engagement throughout the process. (See examples provided in Appendix B.)

For example, accuracy is often quantified by statistical measures, such as the area under the curve (AUC), to measure sensitivity and specificity of the risk scale. These measures, while available for researchers, were not the focus of the data visualizations. As a basic measure, if a risk classification is accurate, low-risk families should experience lower rates of subsequent system involvement compared with high-risk families.

---


2 In addition to these requirements, Evident Change evaluates for reliability. For DCYF, these tests were done after model selection. This work is briefly discussed in the Next Steps section.
The data can be shown with simple data visualization using bar charts. For example, Figure 1 presents families from the validation study disaggregated by race/ethnicity and compares the rate of new investigation within 18 months by risk level. When a risk assessment is accurate, the bars display a stepwise increase with each increase in risk level from low to very high. (See the pattern for White families in Figure 1.) If this pattern is not present, or if the increases in the outcome rate are smaller than expected, it is easy for the steering committee to see this. (See the pattern for multiracial families in Figure 1.) Data visualizations such as Figure 1 were used to engage with the steering committee while Evident Change researchers were simultaneously analyzing and holding risk models to standard levels of statistical rigor.

Figure 1

Is the Risk Assessment Accurate?

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

This strategy led to a familiarity with the information and a shared language for the group to evaluate the assessments. Because the steering committee included a mix of child welfare staff who routinely use the risk assessment and community stakeholders who may know less about the assessment, a series of learning conversations, considerations, and activities were held to introduce and orientate the group. For example, smaller groups were asked to compare the distribution and accuracy of risk models and rank the models according to race equity. To do this, members actively considered what race/ethnicity groups have a larger or smaller percentage of families and examined how a model differentiates race/ethnicity groups by their likelihood of experiencing subsequent system involvement.

Simultaneous to the concepts of risk validation, Evident Change facilitated a discussion on how the SDM risk assessment aligned with the agency’s service provision policies. Evident Change typically recommends
providing services to all high- and very high-risk families, regardless of whether allegations were substantiated. At the time of this case study, DCYF’s legislative-driven policy allowed the agency to provide ongoing services to families at high and very high risk only if the allegations were substantiated. This limited the number of higher-risk families without allegation substantiations who could receive DCYF services.

Data were shared to examine how families experience system involvement by both risk level and allegation finding (Figure 2). Of families identified as low risk, 24% experienced at least one more new investigation in the 18 months after their risk assessment. This rate was substantially lower than high-risk families (58%). These rates also show that regardless of the allegation finding (substantiated or unsubstantiated), families classified as very high risk return to the system more often. One goal of post-investigation services is preventing families from experiencing subsequent system involvement. These data support why risk is a valuable consideration for offering ongoing services to higher-risk families.

**Figure 2**

*New Investigation in 18 Months by Risk Level and Investigation Finding*

![Bar chart showing new investigations by risk level and investigation finding](image)

3 Evident Change examined a population of all families who experienced an assessment and had a risk assessment completed between January 1 and December 31, 2017. These 7,455 families were followed for 18 months after their initial assessment, with subsequent DCYF involvement tallied. These data are from the risk construction sample (53%, n=3,971) of all families.
STAGE 2: EVALUATING THE CURRENT RISK ASSESSMENT

The next stage of the process aimed to review and determine what areas of the risk assessment, if any, need improvement.

UTILITY

As a first step, the steering committee was invited to consider utility by exploring whether the risk distribution is meaningful and supportive to policy and practice. Data on the current risk assessment showed that 18% of families were classified as low risk, 54% as moderate, 28% as high, and 2% as very high risk (not shown). Given these data, the steering committee reflected on resource implications for DCYF and the community. More specifically, they asked if the distribution reflected the number of families that DCYF and/or the community could work with.

ACCURACY

Next, groups considered accuracy. Overall, 40% of all families were involved in a new investigation during the 18-month outcome period. Of families assessed as low risk, 24% had a new investigation, meaning about one in four low-risk families experienced at least one new investigation, and 76% did not. This outcome rate was compared with 38% of moderate-risk, 55% of high-risk, and 58% of very high-risk families who had this outcome (Figure 3).

The steering committee spent time considering whether the model identifies families that DCYF and the community should continue to work with. For example, is the difference between the outcome rates of moderate- and high-risk families sufficient to justify a different response from the agency or community?
After considering utility and accuracy, the committee reviewed the data for an equity analysis. Race equity was centered by ensuring the risk level meant the same thing for everyone, regardless of whether families identified as Asian/Pacific Islander, Black/African American, Hispanic or Latino, White, or multiracial. Utility and accuracy measures offer a comparative understanding. The steering committee was asked, “To what extent do these measurements look similar across subgroups?” Distributions can help reveal whether some race/ethnicity groups are more or less frequently classified as high risk (see Figure 4). Accuracy measures examine if families at each risk level return to the system at a similar rate, regardless of race/ethnicity (see Figure 5).

---

4 Race/ethnicity groups reflect the way data are captured in the agency’s administrative database.
Figure 4
Equity: Risk Distribution by Family Race/Ethnicity
N = 3,971

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>White (n=2,964)</td>
<td>17%</td>
<td>54%</td>
<td>27%</td>
<td>2%</td>
</tr>
<tr>
<td>Hispanic/Latinx (n=264)</td>
<td>22%</td>
<td>51%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>Black/African American (n=222)</td>
<td>21%</td>
<td>59%</td>
<td>19%</td>
<td>1%</td>
</tr>
<tr>
<td>Multiracial (n=196)</td>
<td>17%</td>
<td>47%</td>
<td>32%</td>
<td>5%</td>
</tr>
<tr>
<td>Asian/Pacific Islander (n=67)</td>
<td>39%</td>
<td>57%</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>Missing/Unknown (n=255)</td>
<td>19%</td>
<td>53%</td>
<td>26%</td>
<td>2%</td>
</tr>
</tbody>
</table>

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

Figure 5
Equity: New Assessment by Risk Level by Family Race/Ethnicity
N = 3,971

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>White n=2,964</td>
<td>26%</td>
<td>56%</td>
<td>59%</td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latinx n=264</td>
<td>40%</td>
<td>51%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black/African American n=222</td>
<td>20%</td>
<td>21%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiracial n=196</td>
<td>21%</td>
<td>43%</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific Islander n=67</td>
<td>8%</td>
<td>21%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing/Unknown n=255</td>
<td>23%</td>
<td>31%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Any subgroup with a sample size less than 25 was too small to report and is indicated by the lock symbol.
During this stage of the validation study, Evident Change researchers leveraged data from the empirical analysis to create a worksheet with data on distribution, accuracy, and equity. The worksheet included activities for understanding and questions for consideration. The information was shared in advance, and materials were presented in large-group orientations. Afterward, members of the steering committee participated in virtual breakout rooms to discuss each measure. Each small group began with a short activity to develop an understanding of each concept and then posed questions for consideration.

Polling was used to gather group feedback about whether the areas of utility (distribution), accuracy, or equity required revisions. The polls helped set the direction for validation efforts. Steering committee members were polled on the three concepts using a “traffic light” scale: A red light means major revisions are recommended; a yellow light indicates moderate revisions are recommended, and a green light means no revisions are recommended. Any category with a red light was considered a priority for the revision process. The polling results suggested the steering committee wanted to see improvements in all three areas.

**KEY FINDINGS FROM CURRENT RISK ASSESSMENT REVIEW**

After reviewing the existing risk assessment performance and polling results, the committee recommended exploring revised assessment models that better align with DCYF’s capacity to provide services. This meant the risk models should explore distributions that organize fewer families into the high-risk group. In addition, the policy review suggested moving to a three-level assessment (low, moderate, and high) from the current four levels (low, moderate, high, and very high). The equity analysis demonstrated a call for focused attention on multiracial families to improve both the accuracy and the distribution of the tool. The steering committee noted that distribution for multiracial families had a greater percentage of high- and very high-risk families and that accuracy was not performing well by risk level. These findings were given to the Evident Change research team, along with an imperative that any revised models maintain or improve the accuracy and racial equity for all families.

**STAGE 3: EVALUATING REVISED MODELS**

Evident Change used the risk validation dataset to explore revisions that could improve performance. Hundreds of models were developed, each with a unique set of items and risk level thresholds. Bivariate and multivariate analyses were used to develop, analyze, and evaluate combinations of items and scoring algorithms (i.e., the logic and scoring for the assessment). Evident Change distilled the choices by ranking the models according to feedback from the steering committee while maintaining valid distribution, accuracy, and equity. Two models—model A and model B for this case study—were presented to the committee.

Evident Change created a worksheet to compare analytics from each model and to facilitate the final workgroup. Data were presented on service capacity to inform the utility measure. Distribution and accuracy
measures were presented and examined for each model and disaggregated by family race/ethnicity. (See Appendix B: Analytic Model Comparisons.)

The steering committee was convened to weigh the following options: select a new revised risk model, opt to stay with the current model, or recommend more research. Models A and B were introduced with a basic analysis of how individual risk items on each are statistically related to the three outcomes. The composition of each model, including the items and scoring, was discussed. Each model’s distribution, accuracy, and equity were shared, and the group discussed benefits and trade-offs.

Data were presented and discussed in a format similar to previous stages: large-group presentations about each risk concept, followed by small-group discussions. A key to this engagement was focusing on the familiar data points and visualizations used throughout this process. This helped ensure workgroup members were comfortable with the information and equipped to evaluate the models. In this process, the workgroup set the value standard by reviewing results from different models and building consensus on which model is best. Choosing a model in this way helps create an equity value standard. Predictive models use fairness adjustments that often become part of the algorithm itself (Corbett–Davies et al., 2017).

All members were asked to rate each potential risk model (current, A, or B) on a scale of 1 to 5 (Figure 6). Models A and B were more agreeable to the steering committee than the current assessment. The rating was fairly directive: Participants overwhelmingly favored model B, which had no ratings of 1 ("No way, no how!"). Steering committee rankings were used to inform final model recommendations, which were vetted and approved by Evident Change and agency leadership.

**Figure 6**

Poll Results: What Is Your Opinion of Each Risk Model?

- **Current Model**
  - 5 I love it: 44%
  - 4 I like it: 56%
  - 3 Could go either way: 19%
  - 2 Not comfortable but won’t stand in the way of plan: 19%
  - 1 No way, no how: 13%

- **Model A**
  - 5 I love it: 56%
  - 4 I like it: 19%
  - 3 Could go either way: 19%
  - 2 Not comfortable but won’t stand in the way of plan: 63%

- **Model B**
  - 5 I love it: 25%
  - 4 I like it: 6%
  - 3 Could go either way: 13%
  - 2 Not comfortable but won’t stand in the way of plan: 63%
  - 1 No way, no how: 19%
FINAL CONSIDERATIONS

The workgroup reached consensus that model B could outperform the current risk model. Structural changes to the assessment included revising and rescoring a number of questions on the assessment and using different thresholds for risk levels. The new assessment moves from a four-level classification to a three-level classification. The current risk assessment identified a small percentage of families as very high risk, and there was no clear policy distinction for these families. Collapsing to a three-level assessment was a design choice made to align with policy.

The estimates from Model B suggest a risk distribution of low at 15%, moderate at 57%, and high at 27%. In this model, 64.4% of all high-risk families experienced a new investigation within 18 months (up nearly 10 percentage points from the current model’s 55.6%). These data suggest that classification accuracy could be improved, and fewer families would be classified as high risk, which would be better aligned with service availability.

The revised model also demonstrated improved distinction in outcome rates for moderate- and high-risk multiracial families while maintaining race/equity values for all families. Outcome rates were also closest to the base rate in model B, meaning moderate-risk families will experience outcome rates very similar to the average for all families.

With these considerations in mind, the steering committee reached consensus on supporting model B, and this decision was subsequently approved by DCYF leadership.

NEXT STEPS

Prior to implementation, the revised risk assessment went through a workgroup process to write clear item definitions and policy about risk assessment completion and use of the risk assessment results. Some members of the workgroup were also steering committee members. The final draft underwent inter-rater reliability (IRR) testing to examine the consistency of the revised assessment and field testing to gather feedback from staff when the assessment was applied in field conditions. (Revised assessment findings were not used to guide case decisions during field testing.) IRR and field testing provide opportunities for a broader group of staff to try the assessment and provide feedback on it. Findings of the IRR and field tests were used to make small modifications to item structures and definitions prior to implementation.

The validation of the risk assessment and the workgroup process led directly to practice changes in New Hampshire. New policy will allow for risk level, allegation finding, and safety assessment to be considered in offering services. This will help ensure that high-risk families can receive more intensive services (DCYF-managed cases—voluntary or court-ordered—and community-based voluntary cases), moderate-risk families are referred to community services, and low-risk families do not receive services.
SUMMARY

If a child welfare agency is looking to support post-investigation decision making, exploring risk assessment is a place to start. But agencies and researchers must proceed with caution: When leveraging data tainted by structural racism and inequity, it is imperative to do so without furthering system disparity.

Risk solutions require constant dedication to values. Developing risk assessments that impact lives is a great responsibility. Studies and risk validations have shown that risk assessments accurately classify families who are likely to return to the child welfare system as higher risk than those who are less likely to return. Actuarial risk assessments are intentionally transparent so that everyone, including workers and families, can review the risk factors and algorithms that contribute to the family’s risk level. These data offer a critical view to examine a risk assessment’s utility, accuracy, and equity. Given this knowledge and transparency, how do we use the information?

The New Hampshire case study demonstrates how an inclusive group of stakeholders can play an essential role in the development, selection, and eventual promotion and support of the risk assessment. This group came from varied backgrounds. Risk validation, of course, requires statistical modeling and principles, but these rarely offer a singular solution, and choices emerge that require trade-offs among utility, accuracy, and equity. Finding the right balance requires inviting varied expertise into the development process. This study describes a process to engage with stakeholders, present data, and build consensus to support a risk-based decision-support solution.

Participatory risk validation research encourages everyone—families, staff, community, administrators, researchers, and subject-matter experts—to collectively evaluate an assessment’s value. These groups must use data and analytics to examine disproportionality and inequity in the system and the risk data to balance an application that will support children and families and meet the agency’s needs. We know that not all system interventions are equally beneficial to all families. To understand what constitutes harm means leading with the impact on children and families while holding perspectives across differences and culture, service opportunities, civil liberties, and agency accountability. If this standard can be met, a risk assessment can offer necessary decision-making support to child welfare agencies.
REFERENCES


APPENDICES

Participant guides were developed and shared with stakeholders. The handouts were distributed to increase data literacy on the risk assessment validation, inform design choices, and generate feedback.

Copies of theses handouts are presented in this appendix.

A: INTRODUCTION TO ACCURACY AND EQUITY
B: MODEL SELECTION GUIDE
C: RISK ASSESSMENT GOALS AND VALUES: REFLECTION QUESTIONS
APPENDIX A: INTRODUCTION TO ACCURACY AND EQUITY

STAGE II STEERING COMMITTEE EXERCISE

INTRODUCTION

While developing risk assessment models can involve complicated math, the process of selecting the right model should go beyond numbers. It should also incorporate an agency’s values about distribution, accuracy, and equity. This worksheet is intended to help you become familiar with the functioning of the current New Hampshire Division for Children, Youth and Families (DCYF) risk assessment model and begin to reflect on values to incorporate into an updated risk assessment.

POLICY DECISION: ONGOING SERVICE PROVISION

The purpose of a Structured Decision Making® (SDM) risk assessment is to support workers in making decisions about providing post-assessment services to families. SDM® policy recommends providing services to families most likely to have a subsequent assessment. The SDM risk assessment classifies families as being at low, moderate, high, or very high risk of subsequent DCYF involvement (in this example, subsequent reassessment). Families classified as high or very high risk are typically recommended for a family service case.

SAMPLE

DCYF implemented the SDM system in 2001, and it is now time for a validation of the risk assessment. Evident Change examined families who experienced an assessment and had a risk assessment completed between January 1 and December 31, 2017. These 7,455 families were followed for 18 months after their initial assessment, with subsequent DCYF involvement tallied. The SDM risk assessment data and the rates of subsequent DCYF involvement will be used in this exercise to evaluate the current performance of the risk assessment. This exercise shows data from about half (53%, n=3,971) of the sample—this is called the construction sample. This will allow for validating the new risk assessment on the other half of the sample—called the validation sample.

DIRECTIONS

Review the functioning of DCYF’s current risk assessment model using the data below. The exercise is meant to help you become familiar with different metrics. There is not a “correct” answer. These will be covered in more detail during the risk assessment validation steering committee meeting.
ACCURACY

Accuracy: How the model differentiates groups by their likelihood of experiencing a new assessment in the future.

This figure shows what percentage of families come back to DCYF after a risk assessment. Overall, 40% of families were reassessed in 18 months, shown by the base rate line. For example, 24% of low-risk families experienced at least one new assessment, and 76% did not.

Activity

1. Circle the bar that represents the group of families that had the largest percentage come back to DCYF.
2. Put a box around the bar(s) that represents which families would have been recommended for family services.

Questions to Consider About Accuracy

- Did families in each risk level category (low, moderate, high, very high) come back to DCYF at different rates?
- Did the rates of families coming back to DCYF increase with each increase in the risk level?
- Do you believe that the model identifies families that DCYF and the community should work with?
**EQUITY**

Equity: Risk level means the same thing for everyone, regardless of differences in race/ethnicity. Reassessment rates by risk level should be similar across subgroups.

This exercise involves examining two racial subgroups.

![Risk Distribution by Race](image)

![New Investigation in 18 Months by Risk Level and Race](image)

**Activity:** Circle the percentage of families that have been recommended for family services for each group.

**Questions to Consider About Equity**

- Which group—white or multiracial—has a greater percentage of high- and very high-risk families?
- What would this risk-level distribution mean for white and multiracial families?
- Are there other meaningful subpopulations that we need to consider?

**Activity:** Highlight the base rate line.
Questions to Consider About Equity

Which group—white or multiracial families—comes back to DCYF more frequently?

**Activity:** Circle the percentage of high-risk and very high-risk families that had a subsequent assessment in each group.

Questions to Consider About Equity

- Are these percentages similar or different for the two groups?
- What does this mean for families?
APPENDIX B: MODEL SELECTION GUIDE

STAGE III MODEL REVIEW GUIDE

INTRODUCTION
While developing risk assessment models can involve complicated math, the process of selecting the right model should go beyond numbers. It should also incorporate an agency’s values about distribution, accuracy, and equity. This worksheet is intended to help you become familiar with the functioning of the current New Hampshire Division for Children, Youth and Families (DCYF) Structured Decision Making® (SDM) risk assessment model and begin to reflect on values to consider and incorporate into an updated risk assessment.

POLICY DECISION: ONGOING SERVICE PROVISION
The purpose of a Structured Decision Making® (SDM) risk assessment is to support workers in making decisions about providing post-assessment services to families. SDM® policy recommends providing services to families most likely to have a subsequent assessment. The SDM risk assessment classifies families as being at low, moderate, high, or very high risk of subsequent DCYF involvement (in this example, subsequent reassessment). Families classified as high or very high risk are typically recommended for a family service case.

SAMPLE
DCYF implemented the SDM system in 2001, and it is now time for a validation of the risk assessment. Evident Change examined families who experienced an assessment and had a risk assessment completed between January 1 and December 31, 2017. These 7,455 families were followed for 18 months after their initial assessment, with subsequent DCYF child protection involvement tallied. This exercise shows data from about half (53%, n=3,971) of the sample—this is called the construction sample and was used to create different risk assessment models. The new risk assessment model will be validated using the other half of the sample—called the validation sample.

DIRECTIONS
Use this exercise to compare three risk assessment models based on distribution, accuracy, and equity. Your review of the models will help to facilitate a group conversation about the pros and cons for the current model and two possible new versions. Rank each model across all validation criteria and, if possible, select a model that is best according to yours and DCYF’s value standards.
**DISTRIBUTION**

Distribution: The percentage of families at each risk level. Distribution is helpful when considering the usefulness of the classification provided by the model with respect to how service decisions will be made.

Distribution Model Comparison

![Risk Level Distribution by Version](image)

Questions to Consider About Distribution

- Given that SDM policy recommends providing family services to high- and very high-risk families, what are the resource implications for DCYF and the community for each model?
- How do the distributions compare across the three models? What are the implications of having a larger or a smaller high-/very high-risk group?

**Activity:** Compare and rank the different distributions produced by three risk models. Rank the models by the most preferable distribution.

**Rank:** This model’s distribution is reasonable as it relates to practice. It will appropriately inform service allocation for DCYF and the community. If you rank any model as 1, please document your rationale.

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Model A</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Model B</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
**ACCURACY**

Accuracy: Measures how well a model differentiates groups by their likelihood of experiencing subsequent system involvement.

The following figures show what percentage of families within each risk level classification came back to DCYF within 18 months of their sample investigation. For example, for the current model, 23.7% of low-risk families experienced at least one new investigation, and 76.3% did not. Overall, 40.4% of families were reinvestigated in 18 months, shown by the base rate line.

**Investigation Outcome* by Risk Level and Risk Version**

N = 3,971 families

*New investigation within 18 months.

**Substantiation Outcome* by Risk Level and Risk Version**

N = 3,971 families

*New substantiation within 18 months.

**Removal Outcome* by Risk Level and Risk Version**

N = 3,971 families

*New investigation resulting in removal within 18 months.
Questions to Consider About Accuracy

- Do the rates of families coming back to DCYF increase with each increase in the risk level?
- Do you think the models identify families that DCYF and the community should work with based on how often families return to the system?
- Are the outcomes between risk levels different enough to justify a distinction between risk levels?
- How do the outcome rates compare to the base rate?

**Activity:** Compare how well the risk classification identifies families by subsequent system involvement. Rank the models by this measure of accuracy. If you rank any model as 1, please document your rationale.

**Rank Value:** This model accurately differentiates families by their likelihood of experiencing subsequent system involvement.

<table>
<thead>
<tr>
<th>Model</th>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Agree</th>
<th>4 Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Model A</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Model B</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>
EQUITY

Equity: Ensures that risk level means the same thing for everyone, regardless of differences in race/ethnicity. The equity standard measures how distribution and rates of new assessment by risk level are similar across subgroups. For the validation study, Evident Change classified families using race/ethnicity data stored in the DCYF data system BRIDGES.

Distribution Model Comparison by Race/Ethnicity: What percentage of families within each race/ethnicity group classify as low, moderate, or high risk under each model?

<table>
<thead>
<tr>
<th>Current Model: Risk Level Distribution by Race/Ethnicity</th>
<th>White</th>
<th>Hispanic/Latinx</th>
<th>Black/African American</th>
<th>Multiracial</th>
<th>Asian/Pacific Islander</th>
<th>Missing/Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17%</td>
<td>22%</td>
<td>21%</td>
<td>17%</td>
<td>39%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>54%</td>
<td>51%</td>
<td>59%</td>
<td>47%</td>
<td>57%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td>29%</td>
<td>28%</td>
<td>21%</td>
<td>36%</td>
<td>4%</td>
<td>28%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model A: Risk Level Distribution by Race/Ethnicity</th>
<th>White</th>
<th>Hispanic/Latinx</th>
<th>Black/African American</th>
<th>Multiracial</th>
<th>Asian/Pacific Islander</th>
<th>Missing/Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14%</td>
<td>21%</td>
<td>18%</td>
<td>13%</td>
<td>37%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>57%</td>
<td>55%</td>
<td>59%</td>
<td>59%</td>
<td>57%</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td>29%</td>
<td>23%</td>
<td>22%</td>
<td>28%</td>
<td>6%</td>
<td>24%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model B: Risk Level Distribution by Race/Ethnicity</th>
<th>White</th>
<th>Hispanic/Latinx</th>
<th>Black/African American</th>
<th>Multiracial</th>
<th>Asian/Pacific Islander</th>
<th>Missing/Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26%</td>
<td>32%</td>
<td>34%</td>
<td>26%</td>
<td>51%</td>
<td>29%</td>
</tr>
<tr>
<td></td>
<td>55%</td>
<td>53%</td>
<td>52%</td>
<td>55%</td>
<td>46%</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td>19%</td>
<td>15%</td>
<td>14%</td>
<td>19%</td>
<td>3%</td>
<td>13%</td>
</tr>
</tbody>
</table>
Accuracy Model Comparison by Race/Ethnicity

What percentage of families within each race/ethnicity group and risk classification experienced a new investigation within 18 months?

Current Model: New Investigation by Risk Level and Race/Ethnicity

- White: Low 26%, Moderate 40%, High 57%
- Hispanic/Latinx: Low 12%, Moderate 34%, High 51%
- Black/African American: Low 20%, Moderate 31%, High 57%
- Multiracial: Low 21%, Moderate 43%, High 44%
- Asian/Pacific Islander: Low 8%, Moderate 21%, High 23%
- Missing/Unknown: Low 23%, Moderate 31%, High 62%

Model A: New Investigation by Risk Level and Race/Ethnicity

- White: Low 21%, Moderate 38%, High 61%
- Hispanic/Latinx: Low 14%, Moderate 31%, High 58%
- Black/African American: Low 20%, Moderate 29%, High 59%
- Multiracial: Low 19%, Moderate 38%, High 54%
- Asian/Pacific Islander: Low 8%, Moderate 21%, High 22%
- Missing/Unknown: Low 22%, Moderate 32%, High 65%

Model B: New Investigation by Risk Level and Race/Ethnicity

- White: Low 28%, Moderate 42%, High 65%
- Hispanic/Latinx: Low 16%, Moderate 35%, High 65%
- Black/African American: Low 25%, Moderate 30%, High 66%
- Multiracial: Low 24%, Moderate 43%, High 54%
- Asian/Pacific Islander: Low 6%, Moderate 26%, High 24%
- Missing/Unknown: Low 24%, Moderate 38%, High 71%
Questions to Consider About Equity

- Do any race/ethnicity groups have a larger or smaller percentage of high-/very high-risk families? If so, which groups?
- How do the base rates of subsequent DCYF involvement compare across different race/ethnicity groups?
- Did the rates of families coming back to DCYF increase with each increase in the risk level for all groups by race/ethnicity?
- Do you believe the models identify families that DCYF and the community should work with based on how often families return to the system?

Activity: Compare the distribution and accuracy of the three models by race and ethnicity. Rank the models according to equity. If you rank any model as 1, please document your rationale.

Rank Value: This model fairly classifies families by race and ethnicity.

<table>
<thead>
<tr>
<th>Model</th>
<th>Rank Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>Current</td>
<td>○</td>
</tr>
<tr>
<td>Model A</td>
<td>○</td>
</tr>
<tr>
<td>Model B</td>
<td>○</td>
</tr>
</tbody>
</table>

Selecting a Risk Assessment: Three risk assessment tools are under consideration: Current, Model A, and Model B. Help determine which tool, if any, should be implemented by DCYF to inform the decision to provide post-assessment services for families.

Activity: Fill out the rank score for each model on utility, accuracy, and equity. You may adjust your score based on new perspectives from group conversation.

Model Comparison by Validation Criteria

<table>
<thead>
<tr>
<th>Model</th>
<th>Utility</th>
<th>Accuracy</th>
<th>Equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Considerations for Model Selection

- Are the validation criteria of equal importance, or are some criteria more/less important?
- What justifications support your choice?
- Are there circumstances or conditions under which you might choose a different model?
- Are you strongly opposed to any of the models up for consideration? Why?

Report Back Polling: Use Gradients of Agreement

Be prepared to score each model using the gradients of agreement scale.

APPENDIX C: RISK ASSESSMENT GOALS AND VALUES: REFLECTION QUESTIONS

Choosing a risk assessment is not just a numbers game. It should also raise questions about values and policy.

Questions to consider before adopting a new risk assessment include the following.

- Does our expert decision-making team include varied perspectives (e.g., clients, community, frontline staff, research)?

- What is the question we are asking or the problem we are trying to solve?

- Is the outcome of interest one that disproportionately affects some groups in our community? How can we be mindful of this?

- What information would we like to know to help us solve the problem?

- Based on what you know, is risk assessment appropriate to answer the question or problem? Would it improve our current decision-making process?

- What data can we use to help define this question or problem?

- What are the limitations or drawbacks of those data?

- What policy or practice decisions will this assessment help us with?

- What policy or practice decisions will this assessment not help us with?

- What structures need to be in place to implement this risk assessment, such as policy guidance and workflow?

- How do we communicate internally and externally about the purpose and use of risk assessment?

- How can we continue to monitor and evaluate this assessment once it is implemented?