

A woman with blonde hair is sitting on a lawn in a park, holding a baby. She is wearing a light-colored, long-sleeved top. The background shows trees and a grassy area. The entire image is overlaid with a semi-transparent dark green filter.

Using Predictive Modeling to Improve Outcomes

For Children in Allegheny County

Key Partners

Research Team

- Rhema Vaithianathan, Auckland University of Technology
- Emily Putnam-Hornstein, USC
- Irene de Haan, University of Auckland
- Marianne Bitler, UC Irvine
- Tim Maloney, Auckland University of Technology
- Nan Jiang, Auckland University of Technology

Ethics

- Tim Dare, University of Auckland
- Eileen Gambrill, UC Berkeley

Evaluators

Process

Hornby-Zellar Associates

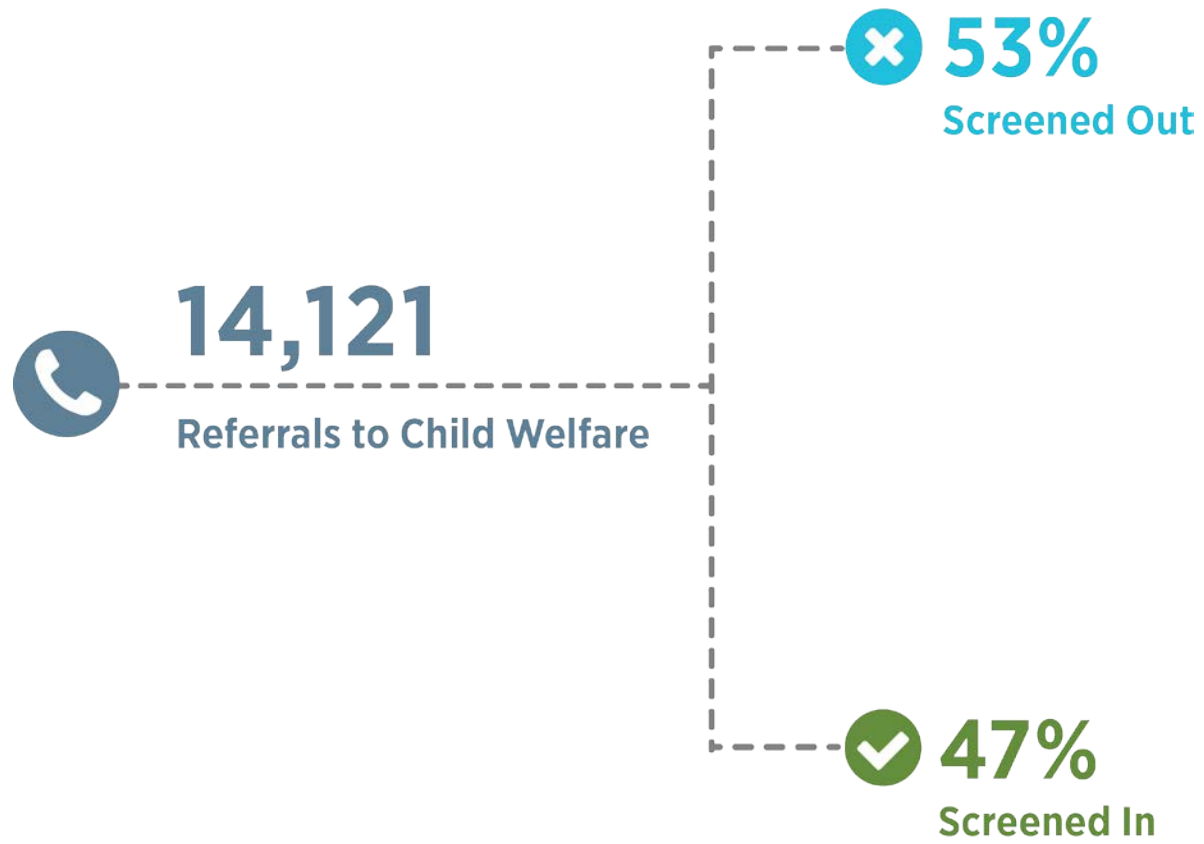
Impact

Stanford University

Technology

Implementation

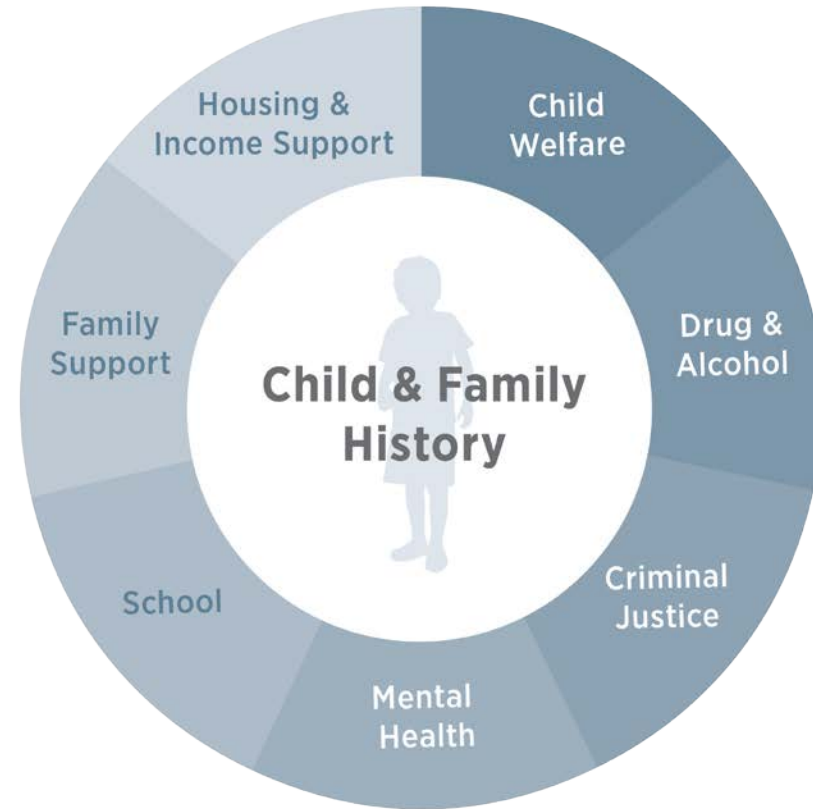
Deloitte



Today: Using Integrated Data to Inform Decision-Making

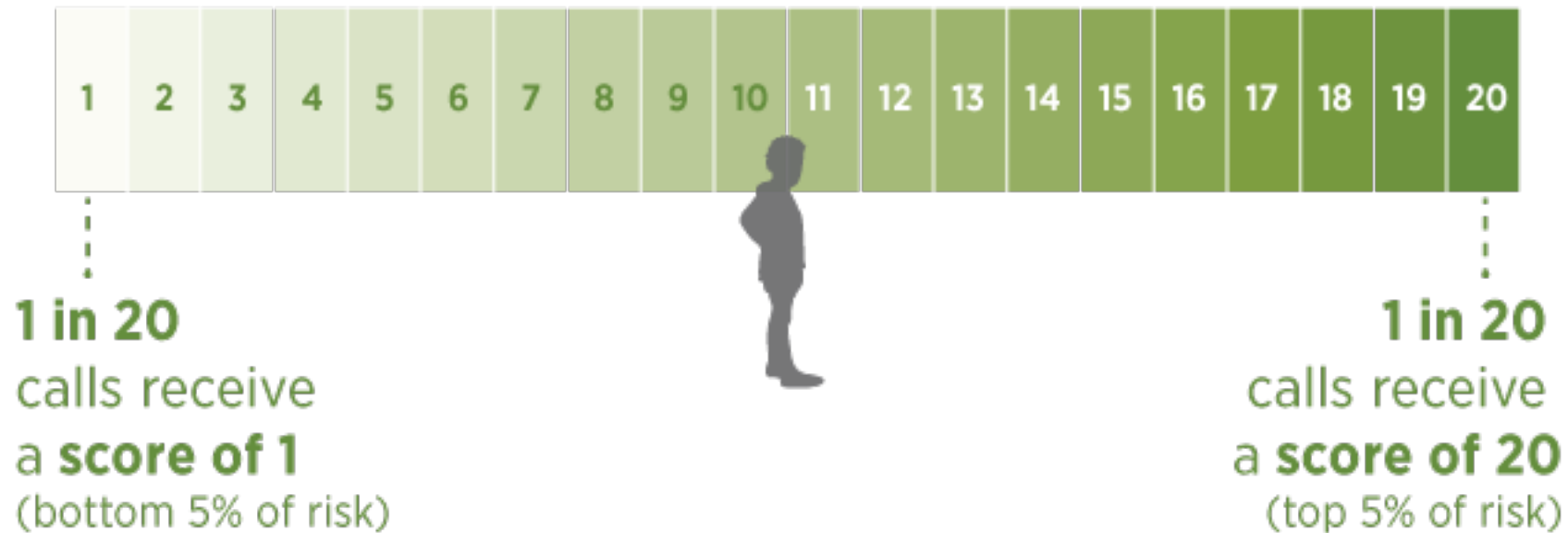
In Allegheny County, rich data are available to case workers to help inform initial maltreatment screening decisions at the child protection hotline, but

- No standardized protocols for using these data to make referral screening decisions
- No method for systematically weighting this information in an equitable manner across all referrals
- No understanding of what information is correlated / predicts future adverse outcomes for children



Developing a Screening Score

- The **screening score** is from 1 to 20
- The **higher the score, the higher the chance of the future event** (e.g., abuse, placement, re-referral) according to the data



Researchers built a screening model based on information that we already collect

They identified more than 100 factors that predict future referral or placement

To test if the model might improve the accuracy of screening decisions, we scored thousands of historical maltreatment calls and then followed the children in subsequent referrals to see how often the model was correct...



The Results: Re-Referrals



1 in 10 children

with a score of 1 were re-referred
within two years of the call.



9 in 10 children

with a score of 20 were re-referred
within two years of the call.

The Results: Out-of-Home Placements



1 in 100 children

who received a **score of 1** were placed **out-of-home** within 2 years of the call

The Results: Out-of-Home Placements



Under current practice:

**27% of highest risk cases
were screened out —**

of these, **1 in 3 are re-referred and placed** within
2 years of the initial screened out call

**48% of lowest risk cases
were screened in —**

and yet only 1.4% of those
are placed within 2 years.

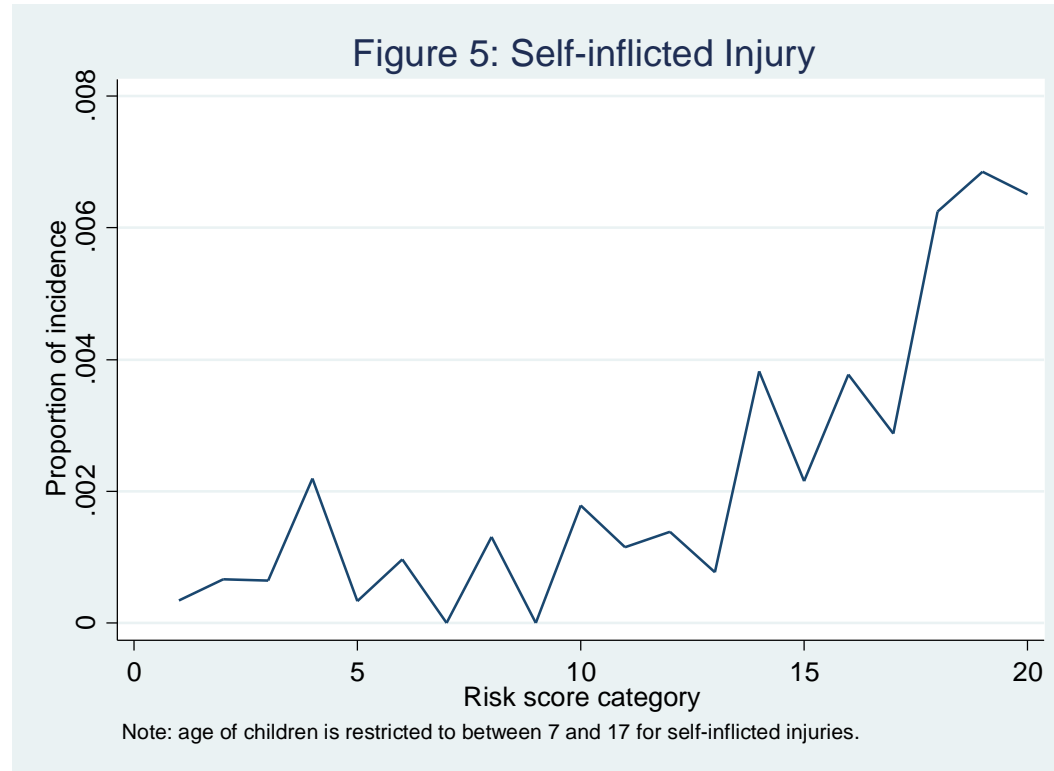


Children's Hospital Validation

- Allegheny County entered into a research agreement with Children's Hospital of Pittsburgh of UPMC in order to study relationships between the child welfare risk modeling and injury data.
- Child welfare referrals were matched with hospital event data (including emergency department visits and in-patient admissions) from February 3, 2002 to December 31, 2015.

Children's Hospital Validation

- Over a broad range of injury types *there is a positive correlation* between the scores¹ at call referral and the rate of hospital events.



¹ maximum placement risk score ever received for each child in the referral data

A woman with dark hair is sitting on a bench in a park, holding a baby. She is wearing a light-colored top. The background shows trees and a path. The image is dimmed and serves as a background for the text.

Preparing for Implementation

Ethics Assessment

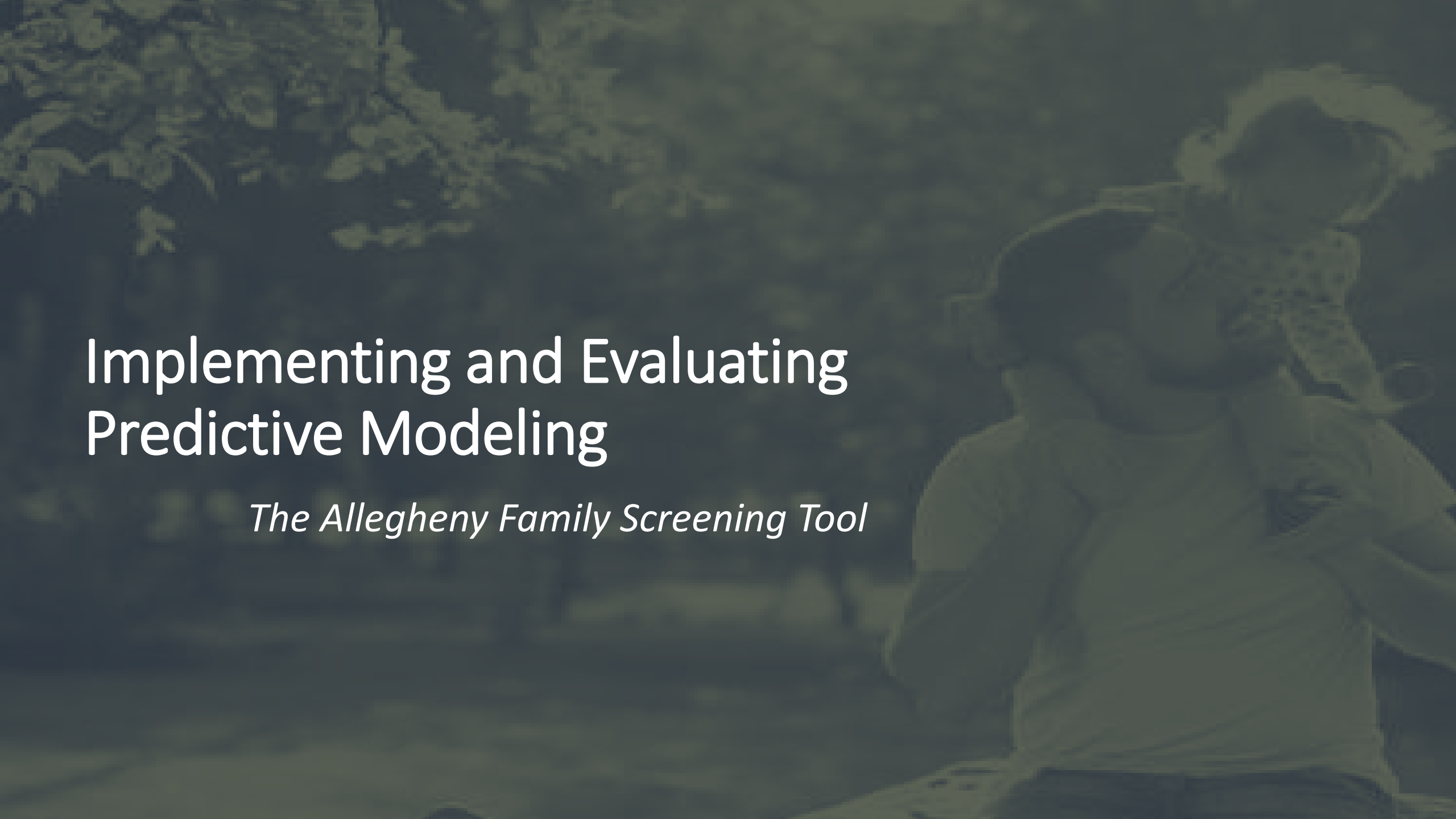
- Tool independently reviewed by ethicists from University of Auckland and UC Berkeley
- Concluded that there would be “significant ethical issues in **not** using the most accurate risk prediction measure.”
- Among key opinions:
 - The tool does not access any data that workers were not already able to utilize in decision-making
 - It is likely more *accurate* and more *transparent* than existing decision-making processes
 - The tool may reduce burdens of stigmatization by allowing for more effective targeting of services

(cont.)

Ethics Assessment

Among key opinions (*cont.*):

- Racial disparities are already present in data at many decision points, and continued vigilance will be required to avoid reinforcement of past biases. However, the writers note that:
 - The *predicted* designation of risk is designed to prompt further in-depth investigation into the family's *actual* risk status; and
 - The resulting potential interventions are designed to assist families.
- Training and ongoing monitoring will be key to ensuring and maintaining effectiveness
- While identifying at-risk families more effectively, it is further ethically required that the eventual services offered are effective

A woman with blonde hair is sitting on a lawn in a park, holding a baby. She is wearing a light-colored top. The background shows trees and a clear sky. The image is dimmed to serve as a background for the text.

Implementing and Evaluating Predictive Modeling

The Allegheny Family Screening Tool

Family Screening Tool Appearance

Screening Score
Historical Screening Scores

Family Screening Score

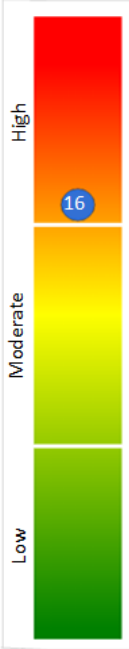
Recalculate Screening Score

The purpose of the **Family Screening Score** is to use information collected by DHS and other partners to inform screening decisions. The **Family Screening Score** is calculated by integrating and analyzing hundreds of data elements on each person related to the referral to generate an overall **Family Screening Score**. The score predicts the long-term likelihood of re-referral, if the referral is screened out without an investigation, or home removal, if the referral is screened in for investigation.

If the **Family Screening Score** meets the threshold for “mandatory screen-in,” the call must be investigated. In all other circumstances, the **Family Screening Score** provides additional information to assist the Call Screening Unit in making a call screening decision and should not replace clinical judgement.

The **Family Screening Score** is only intended to inform call screening decisions and is not intended to be used in making investigative or other child welfare decisions.

High



Last Run By:
Jane McBeth

Last Run Date:
4/7/2016, 10:32 AM

Algorithm Versions Used
Re-referral v43
Placement v22

Screening Score
Historical Screening Scores

Family Screening Score

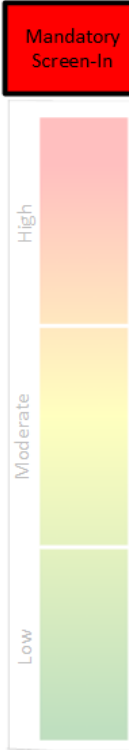
Recalculate Screening Score

The purpose of the **Family Screening Score** is to use information collected by DHS and other partners to inform screening decisions. The **Family Screening Score** is calculated by integrating and analyzing hundreds of data elements on each person related to the referral to generate an overall **Family Screening Score**. The score predicts the long-term likelihood of re-referral, if the referral is screened out without an investigation, or home removal, if the referral is screened in for investigation.

If the **Family Screening Score** meets the threshold for “mandatory screen-in,” the call must be investigated. In all other circumstances, the **Family Screening Score** provides additional information to assist the Call Screening Unit in making a call screening decision and should not replace clinical judgement.

The **Family Screening Score** is only intended to inform call screening decisions and is not intended to be used in making investigative or other child welfare decisions.

High



Last Run By:
Jane McBeth

Last Run Date:
4/7/2016, 10:32 AM

Algorithm Versions Used
Re-referral v43
Placement v22

Monitoring Performance

- 7 months of data through end of February
- Frequent internal monitoring and support activities:
 - Bi-monthly leadership meetings with updated data analyses
 - Tool modifications, functionality fixes as needed
 - Auto-generated weekly support reports regarding “high scores” screened-out
 - Informal interviews with screeners, supervisors
 - Ongoing support activities for contracted process and impact evaluations

Early Scores Differed from Expectations

As data accrued and trends materialized, the first months of the tool yielded:

- More “No Scores” than expected, including a disproportionate impact on referrals involving newborns or other very young children
- Fewer “High” scores than expected

In response to this, made an alteration of the tool to:

- Relax the tool’s requirement for a *child* to have a prior MCI (instead allowing for a score if any individual is known)
- Implemented client-matching functionality to gather data from duplicate IDs

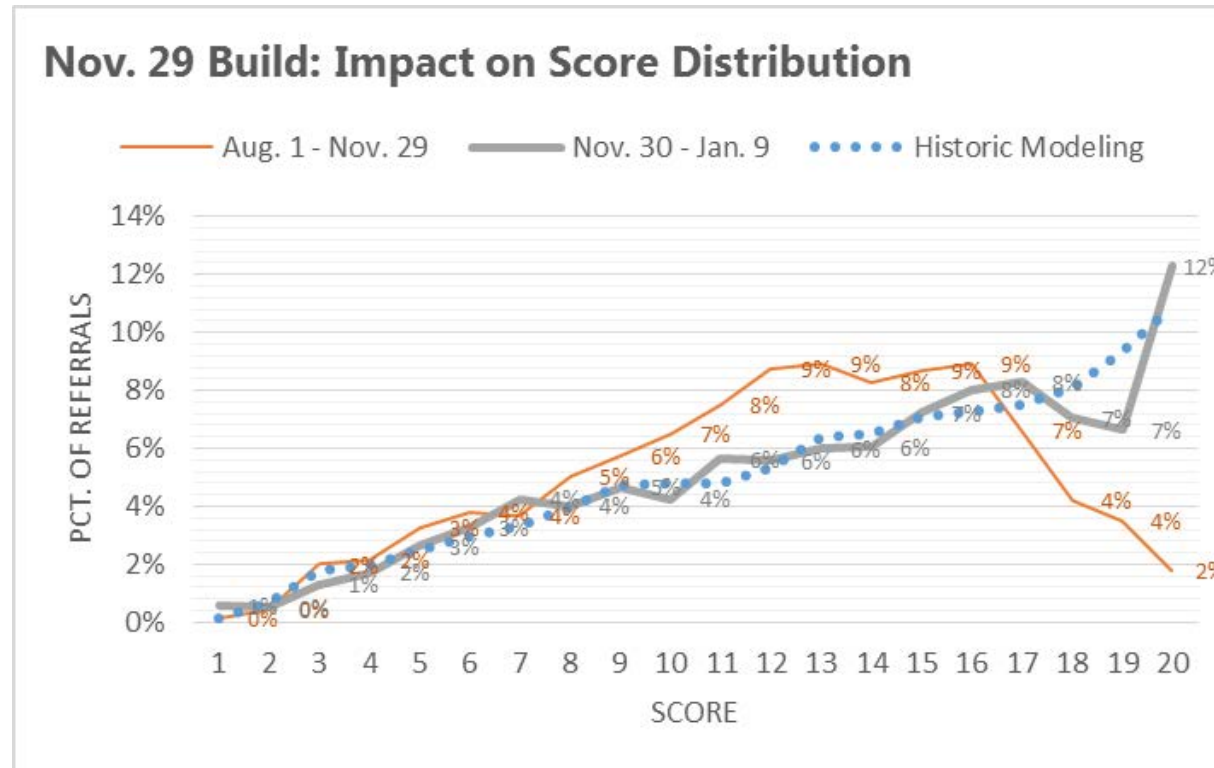
November 29th Improvements

This tool modification went live on November 29th, and changed the relative prevalence of GPS scores in intended ways

- The rate of “Mandatory” referrals roughly doubled from 4% to 9%
- Referrals generating no scores dropped roughly in half
- “High” scores have become the most common score range, supplanting “Medium”

Score Category	Overall Since 8/1/16	Before Nov. 29 Build	Since Nov. 29 Build
Mandatory	6%	4%	9%
High	28%	24%	34%
Medium	29%	31%	26%
Low	19%	19%	20%
No Score	17%	21%	11%
Total	N = 6,103	N = 3,603	N = 2,500

November 29th Improvements, cont.



Use of the Tool

- Since implementation, overall screening rates have remained stable with the prior year's same period
- Generally, referrals with higher scores are being screened-in more frequently

Score Category	Count of GPS Referrals	Screen-In	Screen-Out	Investigation Pct.	Assessment on Active Family
Mandatory	362	184	67	73%	102
High	1733	738	693	52%	289
Medium	1778	707	880	45%	175
Low	1186	362	739	33%	77
No Score	1044	352	681	34%	9
Total	6103	2343	3060	43%	652
Prior Year	5925	2278	3068	43%	579

Score Demographics

- **Racial disparities** have been a monitoring priority at all stages of research and implementation.
- Race was not explicitly invoked in the algorithms, but the outputs of the tool nevertheless showed a tendency for black children to receive higher scores than white children. To date this has borne out in practice as well.
- The impact evaluation will be assessing racial disparity in greater detail to see if the introduction of the tool made any positive or negative changes to biases at call screening.

Score Category	Count of GPS Children	White (N = 1096)	Black (N = 1068)	All Children (N = 2514)
Mandatory	292	10%	14%	12%
High	999	33%	49%	40%
Medium	578	24%	23%	23%
Low	433	22%	12%	17%
No Score	212	10%	3%	8%
Total	2514	100%	100%	100%

Impact Evaluation

The impact evaluation is underway, and will be focusing on:

- Accuracy of decisions
- Reduction in unwarranted variation in decision-making
- Reduction in disparities
- Overall referral rates and workload

Outcomes assessed will include:

- Overall rate of screen-ins
- Likelihood of screen-outs leading to re-referrals or other adverse outcomes
- Likelihood of screen-ins not being accepted for services
- Unwarranted variation in screening decisions
- Disparity in screening decision

Process Evaluation Findings

- 82% felt “somewhat” or “very well” prepared to use the tool following the training.
- In the early weeks of the tool, 69% reported “occasionally,” “almost always,” or “always” consciously using the tool to inform recommendations.
- Some voiced objections to the tool illustrate the tension between immediate allegation and longer-term risk propensity:
 - “the Tool does not take the human element of judgment” into account;
 - “the score frequently has nothing to do with what is actually going on with the situation at hand”

Publication Releases

- Methodology Report (Spring 2017)
- Independent Ethics Review and County Ethics Response (Spring 2017)
- Frequently Asked Questions (Summer 2017)
- Process and Impact Evaluation Reports (TBD)

OPPORTUNITY #1:
Improving Child Welfare
Decision Making

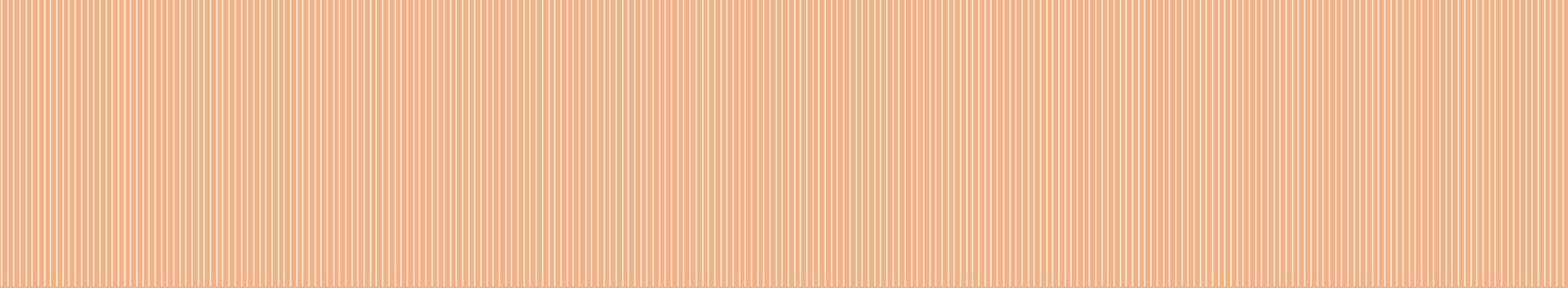
OPPORTUNITY #2:
Rethinking Prevention of
Child Abuse & Neglect





How well do our child serving systems
choose the right child at the right time?





Not very well:

4 in 5 children in this county who died (or nearly died) as a result of abuse **were never referred** to child welfare before the incident.

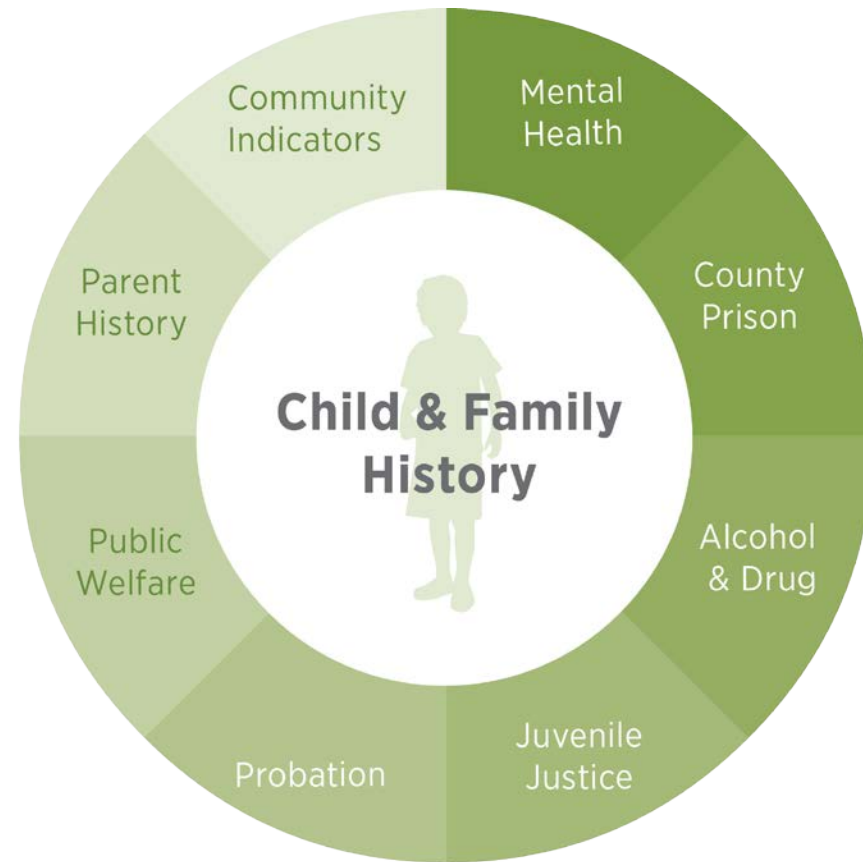


Generating a “Needs” Score at Birth

As **soon as the birth** is registered we could assign a needs score between 1 and 20

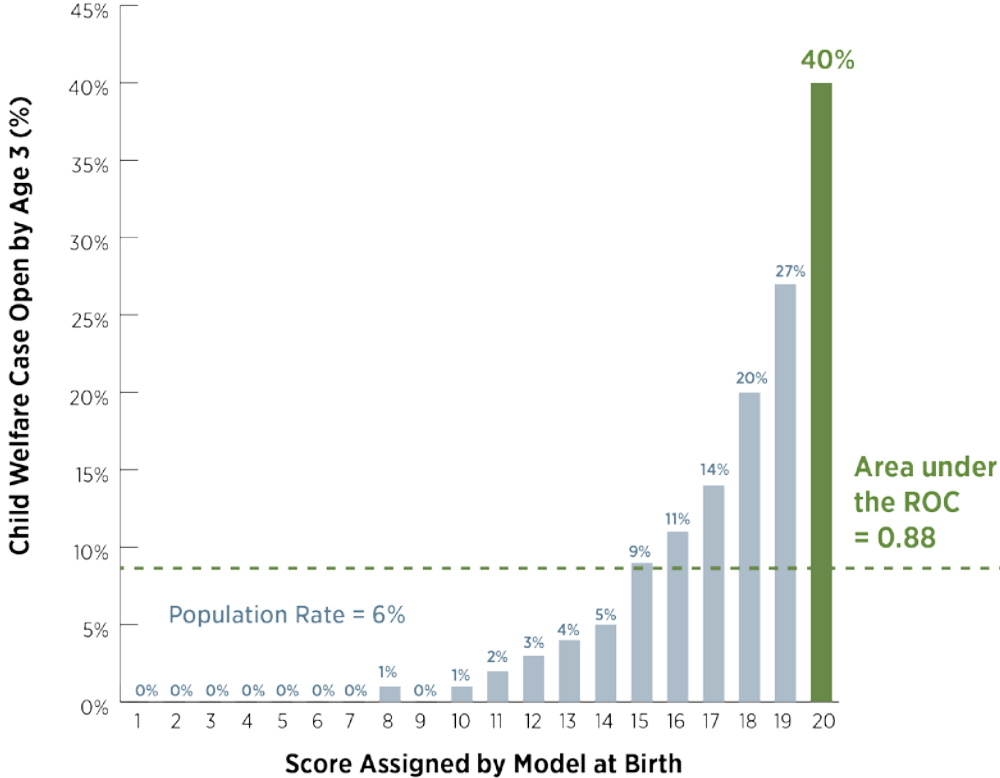
Predicting a child protection case opening **by age 3**

- Vision would be to prioritize high needs births for upstream early intervention support in the hopes of preventing the need for later child protection involvement



Generating a Score at Birth

Of those who received a risk score of 20, 40% of them resulted in an open case by age 3



Opportunities for Prevention

- Offer voluntary services at the time of birth
- Use needs score to prioritize home visiting services through coordinated intake
- Use needs score to provide extra support to families who engage at a family support center
- Proactively reach out to high-risk families who live in a catchment area for family support centers
- Build needs score into screening at Children's Hospital



Using Predictive Modeling to Improve Outcomes

For Children in Allegheny County

Key Partners

Research Team

- Rhema Vaithianathan, Auckland University of Technology
- Emily Putnam-Hornstein, USC
- Irene de Haan, University of Auckland
- Marianne Bitler, UC Irvine
- Tim Maloney, Auckland University of Technology
- Nan Jiang, Auckland University of Technology

Ethics

- Tim Dare, University of Auckland
- Eileen Gambrill, UC Berkeley

Evaluators

Process

Hornby-Zellar Associates

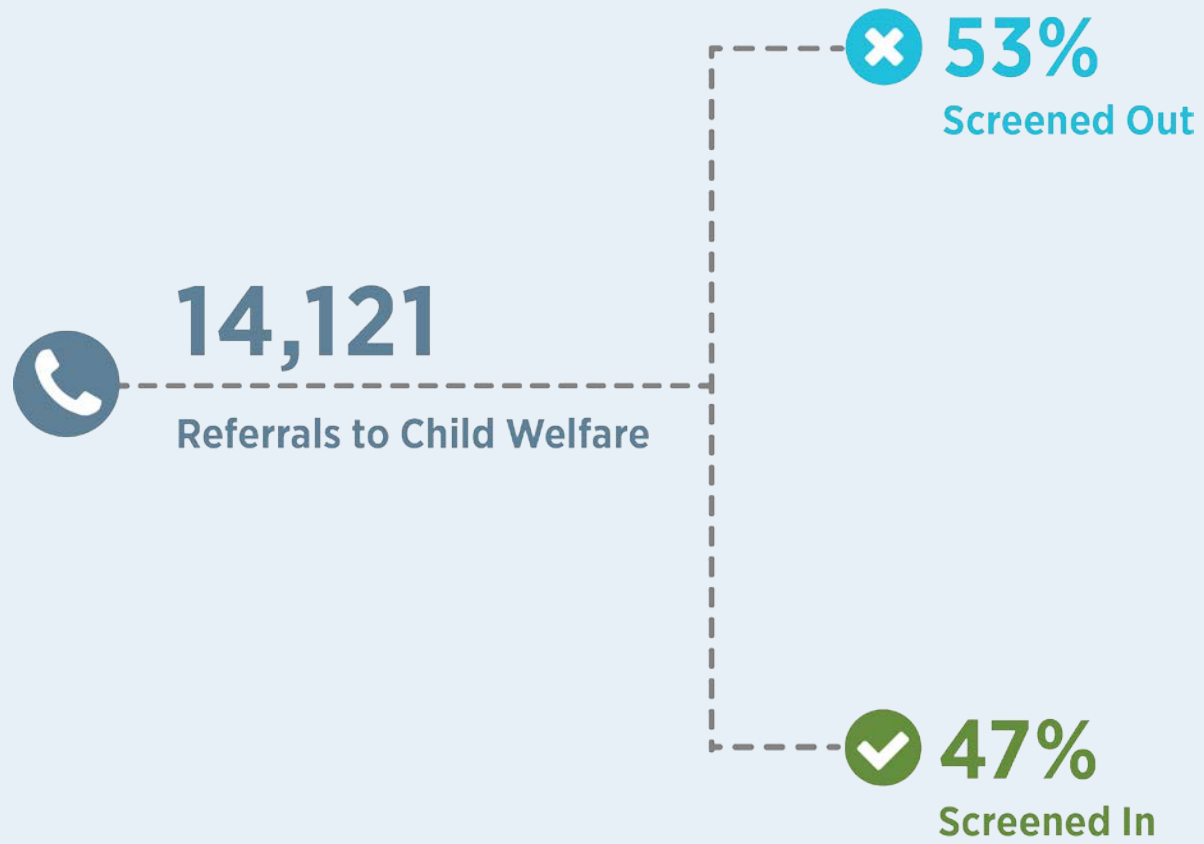
Impact

Stanford University

Technology

Implementation

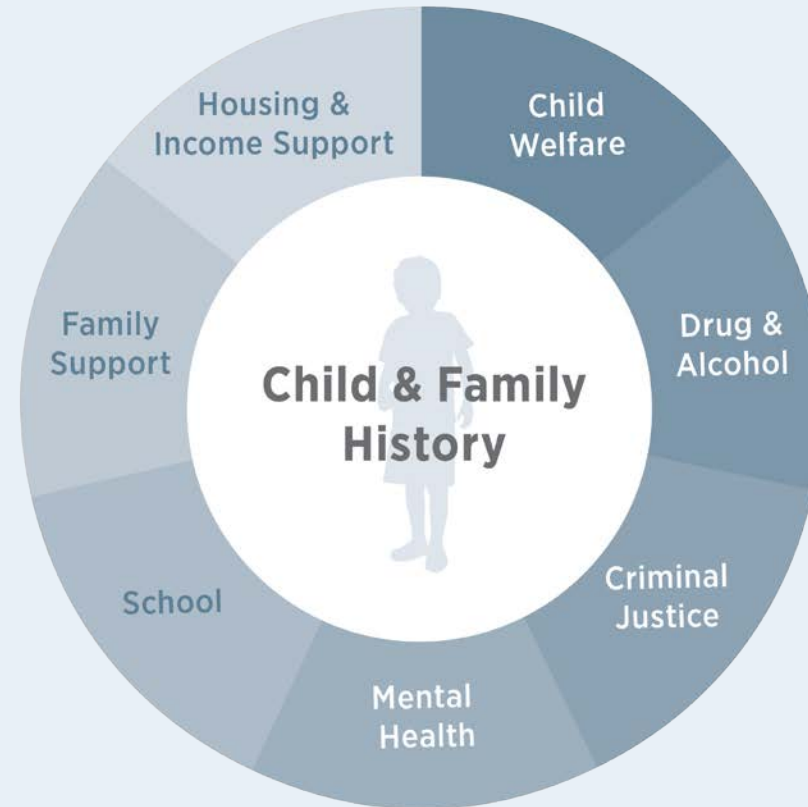
Deloitte



Today: Using Integrated Data to Inform Decision-Making

In Allegheny County, rich data are available to case workers to help inform initial maltreatment screening decisions at the child protection hotline, but

- No standardized protocols for using these data to make referral screening decisions
- No method for systematically weighting this information in an equitable manner across all referrals
- No understanding of what information is correlated / predicts future adverse outcomes for children



Developing a Screening Score

- The **screening score** is from 1 to 20
- The **higher the score, the higher the chance of the future event** (e.g., abuse, placement, re-referral) according to the data



Researchers built a screening model based on information that we already collect

They identified more than 100 factors that predict future referral or placement

To test if the model might improve the accuracy of screening decisions, we **scored thousands of historical maltreatment calls** and **then followed the children in subsequent referrals to** see how often the model was correct...



The Results: Re-Referrals



1 in 10 children

with a score of 1 were re-referred
within two years of the call.



9 in 10 children

with a score of 20 were re-referred
within two years of the call.

The Results: Out-of-Home Placements



1 in 100 children

who received a **score of 1** were placed
out-of-home within 2 years of the call

The Results: Out-of-Home Placements



1 in 2 children

who received
**a score of 20 were placed
out-of-home** within
2 years of the call

Under current practice:

27% of highest risk cases were screened out —
of these, **1 in 3 are re-referred and placed**
within 2 years of the initial screened out call

48% of lowest risk cases were screened in —
and yet only 1.4% of those
are placed within 2 years.



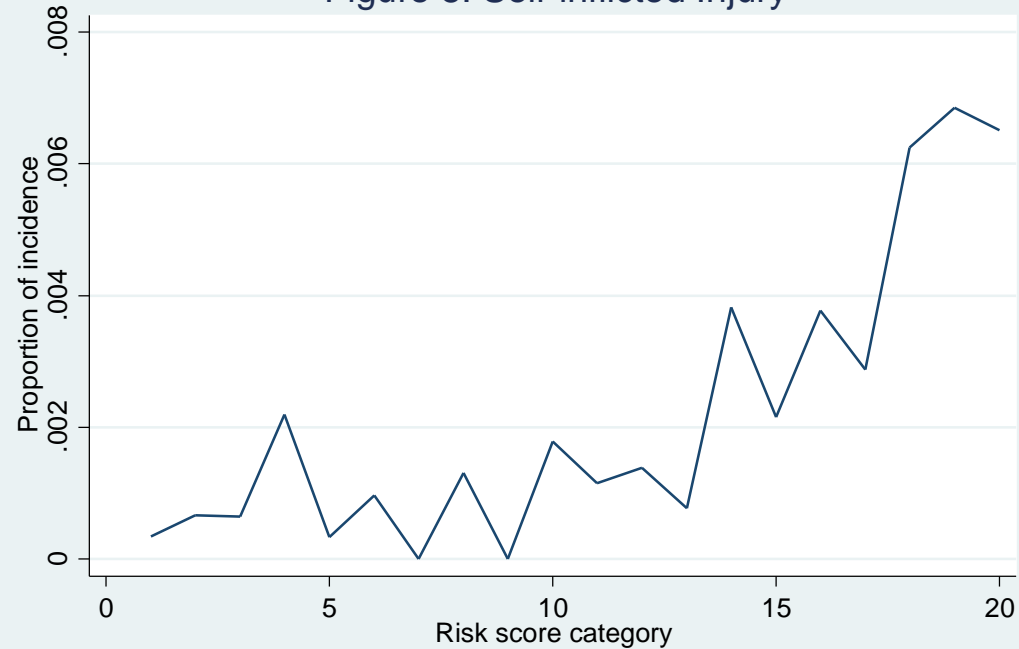
Children's Hospital Validation

- Allegheny County entered into a research agreement with Children's Hospital of Pittsburgh of UPMC into order to study relationships between the child welfare risk modeling and injury data.
- Child welfare referrals were matched with hospital event data (including emergency department visits and in-patient admissions) from February 3, 2002 to December 31, 2015.

Children's Hospital Validation

- Over a broad range of injury types *there is a positive correlation* between the scores¹ at call referral and the rate of hospital events.

Figure 5: Self-inflicted Injury



Note: age of children is restricted to between 7 and 17 for self-inflicted injuries.

¹ maximum placement risk score ever received for each child in the referral data



Preparing for Implementation

Ethics Assessment

- Tool independently reviewed by ethicists from University of Auckland and UC Berkeley
- Concluded that there would be “significant ethical issues in **not** using the most accurate risk prediction measure.”
- Among key opinions:
 - The tool does not access any data that workers were not already able to utilize in decision-making
 - It is likely more *accurate* and more *transparent* than existing decision-making processes
 - The tool may reduce burdens of stigmatization by allowing for more effective targeting of services

(cont.)

Ethics Assessment

Among key opinions (*cont.*):

- Racial disparities are already present in data at many decision points, and continued vigilance will be required to avoid reinforcement of past biases. However, the writers note that:
 - The *predicted* designation of risk is designed to prompt further in-depth investigation into the family's *actual* risk status; and
 - The resulting potential interventions are designed to assist families.
- Training and ongoing monitoring will be key to ensuring and maintaining effectiveness
- While identifying at-risk families more effectively, it is further ethically required that the eventual services offered are effective



Implementing and Evaluating Predictive Modeling

The Allegheny Family Screening Tool

Family Screening Tool Appearance

Screening Score Historical Screening Scores

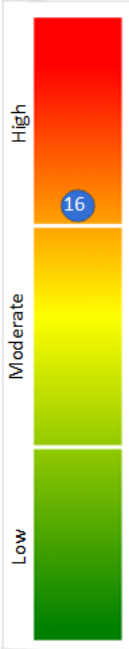
Family Screening Score

[Recalculate Screening Score](#)

The purpose of the **Family Screening Score** is to use information collected by DHS and other partners to inform screening decisions. The **Family Screening Score** is calculated by integrating and analyzing hundreds of data elements on each person related to the referral to generate an overall **Family Screening Score**. The score predicts the long-term likelihood of re-referral, if the referral is screened out without an investigation, or home removal, if the referral is screened in for investigation.

If the **Family Screening Score** meets the threshold for “mandatory screen-in,” the call must be investigated. In all other circumstances, the **Family Screening Score** provides additional information to assist the Call Screening Unit in making a call screening decision and should not replace clinical judgement.

The **Family Screening Score** is only intended to inform call screening decisions and is not intended to be used in making investigative or other child welfare decisions.



High
16
Moderate
Low

Last Run By: Jane McBeth **Last Run Date:** 4/7/2016, 10:32 AM **Algorithm Versions Used:** Re-referral v43
Placement v22

Screening Score Historical Screening Scores

Family Screening Score

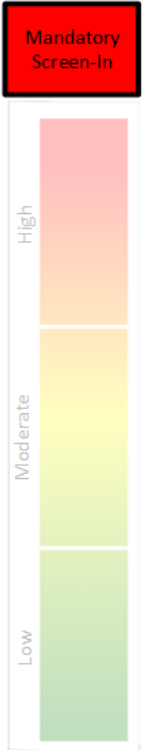
[Recalculate Screening Score](#)

Mandatory Screen-In

The purpose of the **Family Screening Score** is to use information collected by DHS and other partners to inform screening decisions. The **Family Screening Score** is calculated by integrating and analyzing hundreds of data elements on each person related to the referral to generate an overall **Family Screening Score**. The score predicts the long-term likelihood of re-referral, if the referral is screened out without an investigation, or home removal, if the referral is screened in for investigation.

If the **Family Screening Score** meets the threshold for “mandatory screen-in,” the call must be investigated. In all other circumstances, the **Family Screening Score** provides additional information to assist the Call Screening Unit in making a call screening decision and should not replace clinical judgement.

The **Family Screening Score** is only intended to inform call screening decisions and is not intended to be used in making investigative or other child welfare decisions.



High
Moderate
Low

Last Run By: Jane McBeth **Last Run Date:** 4/7/2016, 10:32 AM **Algorithm Versions Used:** Re-referral v43
Placement v22

Monitoring Performance

- 7 months of data through end of February
- Frequent internal monitoring and support activities:
 - Bi-monthly leadership meetings with updated data analyses
 - Tool modifications, functionality fixes as needed
 - Auto-generated weekly support reports regarding “high scores” screened-out
 - Informal interviews with screeners, supervisors
 - Ongoing support activities for contracted process and impact evaluations

Early Scores Differed from Expectations

As data accrued and trends materialized, the first months of the tool yielded:

- More “No Scores” than expected, including a disproportionate impact on referrals involving newborns or other very young children
- Fewer “High” scores than expected

In response to this, made an alteration of the tool to:

- Relax the tool’s requirement for a *child* to have a prior MCI (instead allowing for a score if any individual is known)
- Implemented client-matching functionality to gather data from duplicate IDs

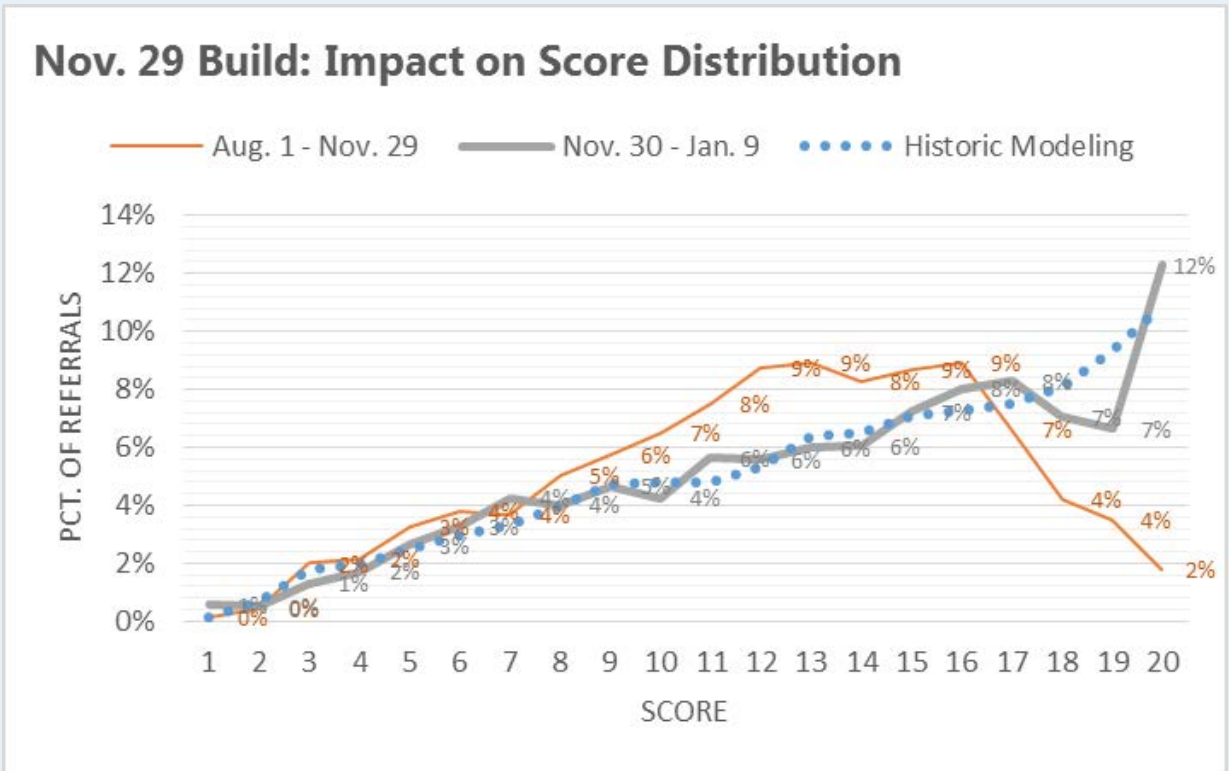
November 29th Improvements

This tool modification went live on November 29th, and changed the relative prevalence of GPS scores in intended ways

- The rate of **“Mandatory” referrals roughly doubled** from 4% to 9%
- Referrals generating **no scores dropped roughly in half**
- “High” scores have become the most common score range, supplanting “Medium”

Score Category	Overall Since 8/1/16	Before Nov. 29 Build	Since Nov. 29 Build
Mandatory	6%	4%	9%
High	28%	24%	34%
Medium	29%	31%	26%
Low	19%	19%	20%
No Score	17%	21%	11%
Total	N = 6,103	N = 3,603	N = 2,500

November 29th Improvements, cont.



Use of the Tool

- Since implementation, overall screening rates have remained stable with the prior year's same period
- Generally, referrals with higher scores are being screened-in more frequently

Score Category	Count of GPS Referrals	Screen-In	Screen-Out	Investigation Pct.	Assessment on Active Family
Mandatory	362	184	67	73%	102
High	1733	738	693	52%	289
Medium	1778	707	880	45%	175
Low	1186	362	739	33%	77
No Score	1044	352	681	34%	9
Total	6103	2343	3060	43%	652
Prior Year	5925	2278	3068	43%	579

Score Demographics

- **Racial disparities** have been a monitoring priority at all stages of research and implementation.
- Race was not explicitly invoked in the algorithms, but the outputs of the tool nevertheless showed a tendency for black children to receive higher scores than white children. To date this has borne out in practice as well.
- The impact evaluation will be assessing racial disparity in greater detail to see if the introduction of the tool made any positive or negative changes to biases at call screening.

Score Category	Count of GPS Children	White (N = 1096)	Black (N = 1068)	All Children (N = 2514)
Mandatory	292	10%	14%	12%
High	999	33%	49%	40%
Medium	578	24%	23%	23%
Low	433	22%	12%	17%
No Score	212	10%	3%	8%
Total	2514	100%	100%	100%

Impact Evaluation

The impact evaluation is underway, and will be focusing on:

- Accuracy of decisions
- Reduction in unwarranted variation in decision-making
- Reduction in disparities
- Overall referral rates and workload

Outcomes assessed will include:

- Overall rate of screen-ins
- Likelihood of screen-outs leading to re-referrals or other adverse outcomes
- Likelihood of screen-ins not being accepted for services
- Unwarranted variation in screening decisions
- Disparity in screening decision

Process Evaluation Findings

- 82% felt “somewhat” or “very well” prepared to use the tool following the training.
- In the early weeks of the tool, 69% reported “occasionally,” “almost always,” or “always” consciously using the tool to inform recommendations.
- Some voiced objections to the tool illustrate the tension between immediate allegation and longer-term risk propensity:
 - “the Tool does not take the human element of judgment” into account;
 - “the score frequently has nothing to do with what is actually going on with the situation at hand”

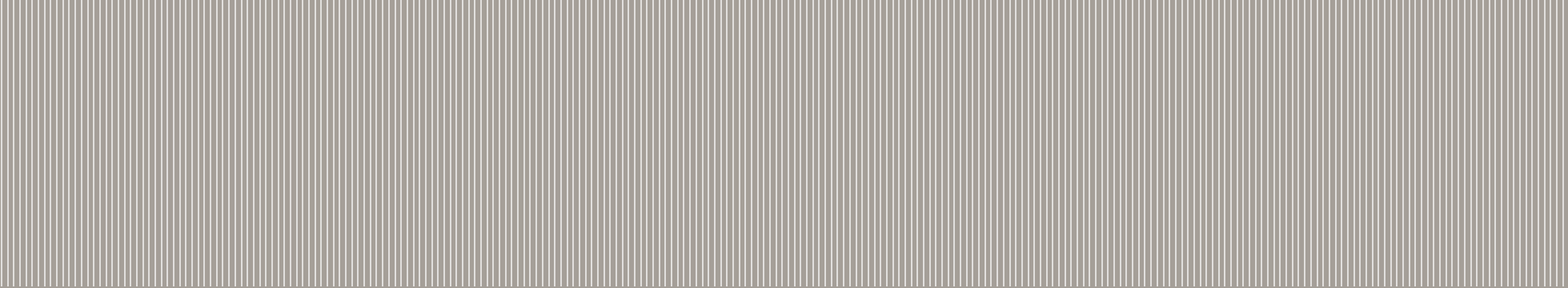
Publication Releases

- Methodology Report (Spring 2017)
- Independent Ethics Review and County Ethics Response (Spring 2017)
- Frequently Asked Questions (Summer 2017)
- Process and Impact Evaluation Reports (TBD)

OPPORTUNITY #1:
Improving Child Welfare
Decision Making

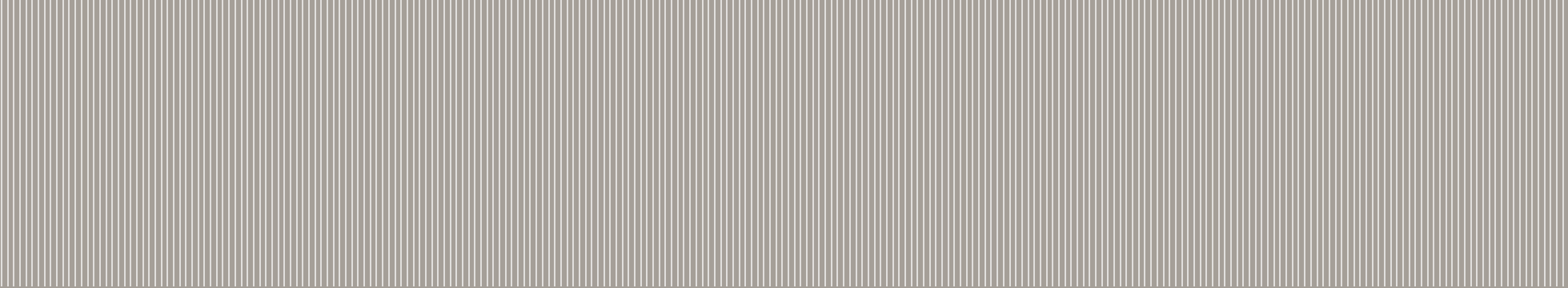
OPPORTUNITY #2:
Rethinking Prevention of
Child Abuse & Neglect





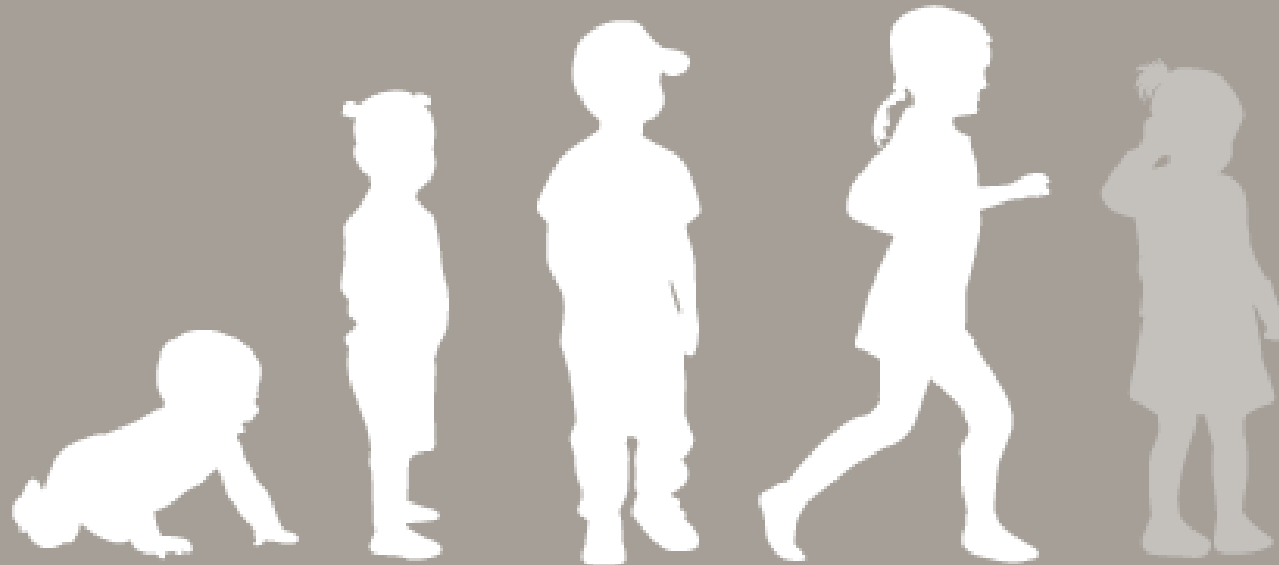
How well do our child serving systems
choose the right child at the right time?





Not very well:

4 in 5 children in this county who died (or nearly died) as a result of abuse **were never referred** to child welfare before the incident.

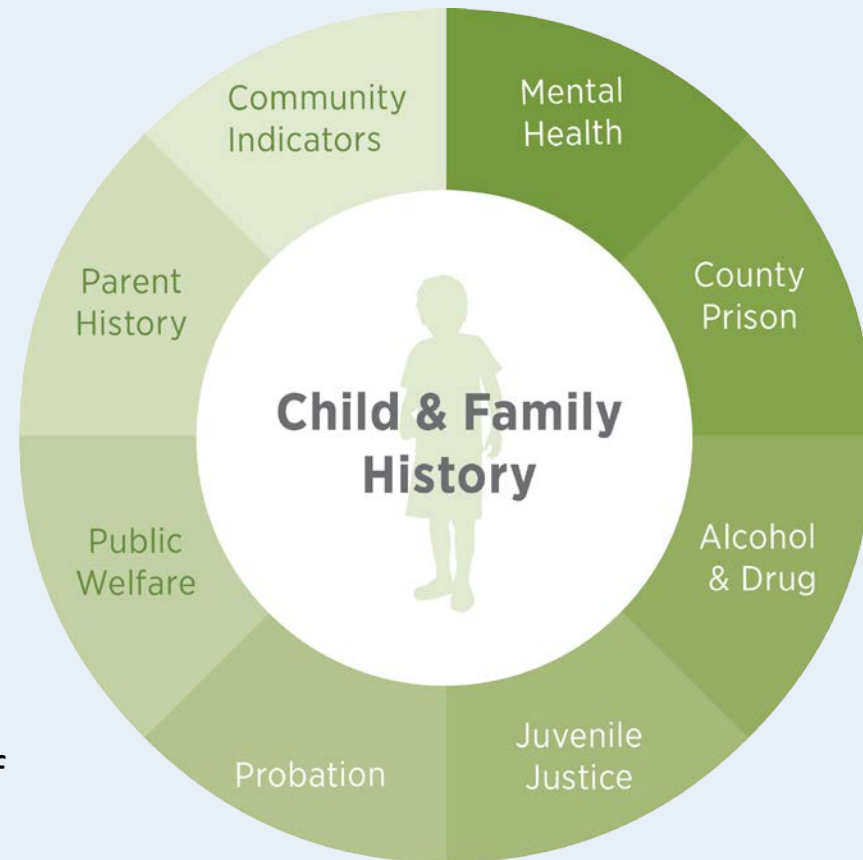


Generating a “Needs” Score at Birth

As **soon as the birth** is registered we could assign a needs score between 1 and 20

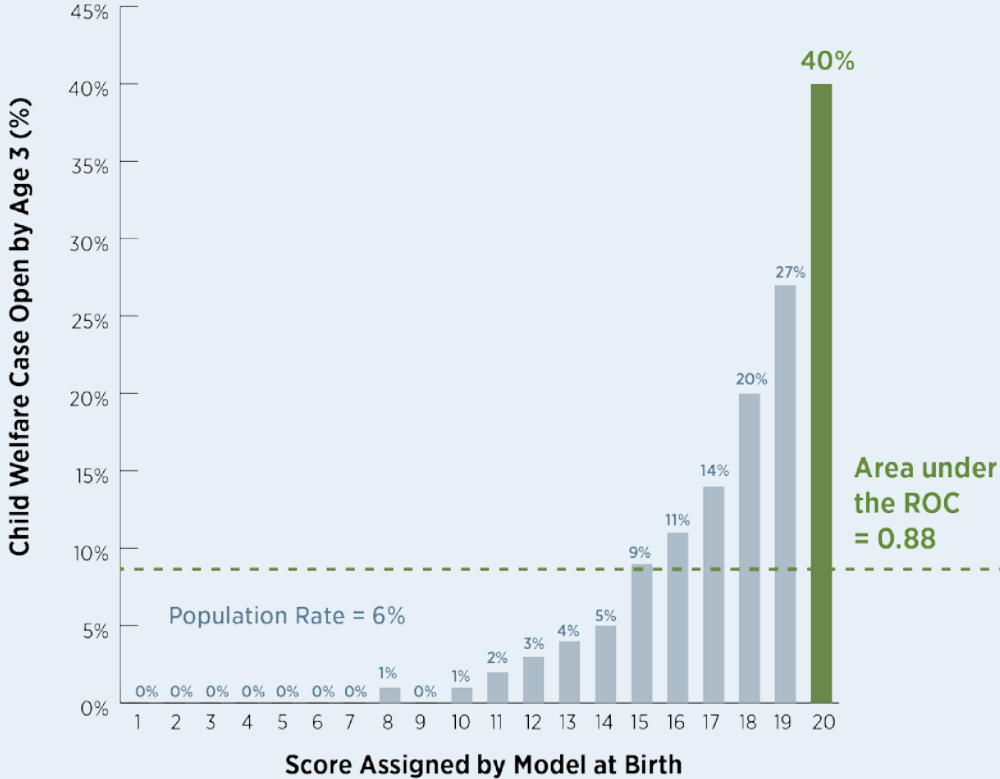
Predicting a child protection case opening **by age 3**

- Vision would be to prioritize high needs births for upstream early intervention support in the hopes of preventing the need for later child protection involvement



Generating a Score at Birth

Of those who received a risk score of 20, 40% of them resulted in an open case by age 3



Opportunities for Prevention

- Offer voluntary services at the time of birth
- Use needs score to prioritize home visiting services through coordinated intake
- Use needs score to provide extra support to families who engage at a family support center
- Proactively reach out to high-risk families who live in a catchment area for family support centers
- Build needs score into screening at Children's Hospital