

SECTION 5

Impact Evaluation Summary of the Allegheny Family Screening Tool

by the Allegheny County Department of Human Services

SUMMARY

The Allegheny Family Screening Tool (AFST) is a predictive risk model built and trained using County administrative child protection and service records.¹ Allegheny County implemented the AFST in 2016 as a decision-support tool, with the goal of improving both the accuracy and consistency of decisions made about referrals to the child maltreatment hotline.

A request for proposals was issued in December 2015 and in early 2016, the Allegheny County Department of Human Services (DHS) issued a competitive contract to Stanford University (principal investigator: Goldhaber-Fiebert) to design and conduct an independent evaluation of the impact of the AFST (along with associated policy changes) on the County's child maltreatment screening decisions.

The evaluation looks at Version 1 of the AFST and consists primarily of outcome comparisons for two groups of children: (1) the approximately 31,000 children who were referred for alleged maltreatment during the 18-month period before the AFST was implemented (January 1, 2015 through July 31, 2016, called "Pre-AFST [late]" in this report) and (2) the approximately 34,000 children referred after the AFST was fully implemented ("Post-AFST": December 1, 2016 through May 31, 2018). This report provides a summary of the findings; to read the full technical report, please see: [Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office](#). Two peer reviewers provided critical feedback on earlier drafts of the evaluation report.

Evaluation findings are detailed in the sections that follow, and emerge from a set of methodologically strong, quasi-experimental methods (i.e., interrupted time series analyses, generalized linear models). Quasi-experimental methods refer to a type of evaluation approach used when it is not possible or desirable to implement a randomized controlled trial (RCT).

¹ Vaithianathan R, Putnam-Hornstein E, Jiang N, Nand P, & Maloney T. (2017). Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County methodology and Implementation. Centre for Social Data Analytics: <https://www.alleghenycounty.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=6442457403>

While less robust than a gold-standard RCT, carefully designed quasi-experimental methods are considered the next-best approach to testing program impact. The County decided not to pursue an RCT primarily for practical reasons.

Key findings of the impact evaluation include:

- 1. Overall, the AFST did not lead to increases in the rate of referred children screened-in for investigation.** Use of the tool appears to have resulted in a different pool of children screened-in for investigation (including more children who were deemed in need of child welfare intervention or supports, see below). But from the perspective of investigative workload, there was not a substantial increase in the number or proportion of children investigated among all children referred for maltreatment.
- 2. Implementation of the AFST increased the identification of children determined to be in need of further child welfare intervention.** Use of the tool led to an increase in the screening-in of children who were subsequently determined to need further intervention or supports. Specifically, there was a statistically significant increase in the proportion of children screened-in whose child welfare case was then opened or, if no case was opened, were re-referred within 60 days. (Please note that investigators and supervisors making these case opening decisions remained blind to the score so this result reflects real change in the case-mix of families screened-in.)
- 3. Use of the AFST did not lead to decreases in re-referral rates for children screened-out without investigation.** Re-referral rates among children screened-out stayed the same for children overall, with the exception of children who were 4-6 years of age. This was the age group directly affected by County changes to mandatory field policy screening protocols by age. Unfortunately, for this age group there was a slight but statistically significant increase in the likelihood of the Post-AFST group of children being re-referred.
- 4. The AFST led to reductions in disparities of case opening rates between black and white children.** Prior to the introduction of the AFST, case-opening rates for black children were higher than for white children. During the Post-AFST period, increases in the rate of white **children determined to be in need of further child welfare intervention**, coupled with slight declines in the rate at which black children were screened-in for investigation, led to reductions in racial disparities. Specifically, there was an increase in the number of white children who had cases opened for services, reducing case disparities between black and white children.
- 5. There was no evidence that the AFST resulted in greater screening consistency within individual call-screeners.** Specifically, for the subgroup of 11 call screeners who handled a substantial volume of both Pre-AFST and Post-AFST referrals, attempts were made to assess whether the AFST led to more “between-screener” consistency. Likewise, changes

in screening consistency by referred child's age group and racial group were also assessed. No impact was detected, although it should be noted that there was likely insufficient power to identify anything other than very large shifts.

2 Allegations fall under the state of Pennsylvania's Child Protective Service (CPS) statutes (23 Pa.C.S. § 6303) or General Protective Service (GPS) statutes (23 Pa.C.S. § 6334). CPS referrals include those made for child abuse, including physical and sexual abuse. CPS referrals must be investigated and require more urgent response times, often overlap with law enforcement and medical investigations, and lead to a determination of whether abuse occurred (that may result in perpetrators being registered in the state's ChildLine registry). GPS referrals include referrals made when there is a risk of harm. For example, neglect, truancy and substance use by parents would all fall under GPS referrals. GPS referrals may be investigated or screened out without further assessment, at the discretion of call screening staff. GPS investigations assess for risk and safety to ensure well-being of children and provide families with any supports they may need. GPS investigations cannot result in registry with the state's ChildLine registry. Both CPS and GPS referrals can result in a family having a case opened at the end of an investigation for ongoing services and supports. In 2017, 21 percent of DHS referrals were CPS referrals and 79 percent were GPS referrals.

3 Of note, a referral flagged for auto screen-in does not obligate the call-screener and/or supervisor to screen-in, but rather implies that this would be the default action. The concurrence rate for referrals marked as auto screen-in for the dates December 1, 2016–November 29, 2018 was 61 percent.

METHODOLOGY

Implementation of the AFST

With the implementation of the AFST, call screeners in Allegheny County are now presented with a single Family Screening Score. The score is a standardized summary of available data, providing additional information to aid the call screener (and their supervisor) to make decisions regarding further investigation. Screening recommendations are made on any call which is classified as “general protective service” (GPS)² for all individuals currently residing in the same household as the alleged victim (alleged child/victim, biological mother and father of alleged victim, the perpetrator, other related and unrelated children in the home, and other adults in the home). Investigation recommendations are made by the County hotline staff (screeners and supervisors) and follow one of three courses: 1) Screen-out of a referral without any further evaluation or assessment, 2) Field screen of the referral to assess whether an investigation is warranted, or 3) Screen-in of a referral, which is synonymous with conducting a formal investigation. When a field screen (a home visit to assess the safety of the child[ren] and determine whether a formal investigation is warranted) is conducted, it is always followed by a decision to either screen-out or screen-in the referral.

At the time of referral, a re-referral and placement risk score is calculated for each child associated with the referral. The AFST, which is the only score the screeners see, is based on the maximum score (either re-referral or placement) across all children associated with the referral at the time of the screening call. The score ranges from 1 to 20 (where 20 is the highest “risk” and 1 is the lowest), indicating the ventile into which the AFST falls. A recommendation for “auto screen-in” occurs when the AFST falls above 18 for the placement score.³

Accompanying Protocol Changes with AFST Implementation

Several other systematic changes to the maltreatment referral screening process accompanied the full implementation of the AFST.

- **Field Screening.** First, the County's mandatory field screen policy was updated. Previously, households with at least one child under the age of 7 were required to be field screened, regardless of the family's history. With the implementation of the AFST, the maximum age for a mandatory field screen decreased from 7 to 4 years of age. In addition, the new mandatory field screen policy added the following three conditions: (1) all children who attend homeschool/cyber school receive a mandatory field screen regardless of age; (2) any family that has had 4 or more referrals in 2 years without any of the referrals being formally investigated receive a mandatory field screen; and (3) any other referrals where more information is necessary to make a final decision receive a field screen.⁴

4 An average of a 1000 GPS referrals per month have been processed since the AFST was implemented and data about reasons for field screening started being collected. During this time period, about 10% of referrals were assigned to the child welfare unit that handles field screening. Of these field screened referrals, 82% indicated a child under 4 on the referrals as the reason and 16% expressed the discretionary desire for more information. The other rules—regarding home schooled children or 4+ recent referrals screened out—were used very rarely.

- **Screener Supervision.** Second, call screeners now make a recommendation about the decision to screen-in/out to his/her supervisor who has the responsibility for the ultimate decision. Prior to this set of policy and practice changes, the primary role of the call screening staff was to gather information to inform supervisor decision-making. Call screeners collected data from several databases and resources, including internal DHS systems (e.g., KIDS, Client View), courts, public assistance and criminal justice. Call screeners also spoke with the individual making the report and other key contacts (e.g., schools, doctors). The information collected was given to supervisors for final decision-making. Although the process between screeners and supervisors was collaborative, following implementation of the AFST, call screeners took on a greater role in making recommendations for screening decisions.

Evaluation Outcomes

The evaluation team defined three main outcomes to measure underlying effects of the AFST implementation on the County's maltreatment referral screening decisions. The choice of outcomes reflected the emphasis the County placed on evidence of changes to screening accuracy, as well as potential impacts on screening decisions by race/ethnicity. The main outcomes investigated were:

- **Overall rates of children screened-in for investigation**

This outcome is intended to measure how the implementation of a predictive risk model impacted the flow of children referred for alleged maltreatment into investigations, with potential implications for workload and the system overall.

A child is considered screened-in for investigation if the referral (i.e., household) that includes the child is advanced by the hotline screener and their supervisor for further investigation. Therefore, the rate of "screened-in for investigation" was defined as equal to the total number of children in referrals assigned to further investigation (numerator) divided by the total number of children in referrals (denominator), computed for referrals falling in each calendar month and for children in different age and racial/ethnic groups.

- **Likelihood a screened-out child had no re-referrals within 2 months**

This outcome is intended to measure how AFST implementation impacted one feature of accuracy: do children who are screened-out appear in subsequent referrals? The assumption is that the absence of a near-term follow-up referral indicates there were no safety and/or well-being issues. A re-referral is assumed to indicate that there was a missed opportunity on the part of the County to have intervened with services earlier.

A child is considered screened-out if the maltreatment referral that includes the child is not advanced by the hotline screener and supervisor for an in-person investigation. If no additional referral is made within 2 months of the index referral event, then the child is considered to have been screened out without a re-referral. It should be noted that a child can have more than one screen-out and re-referral over time, but only subsequent referrals within 2 months of a specific "index event" were examined. The rate of screen-outs with no

re-referral within a 2-month time window was defined as the number of children in referrals that were not advanced for further investigation and were not re-referred within 2 months (numerator), divided by the total number of children in referrals that were screened-out without investigation (denominator), computed for referrals falling in each calendar month and for children in different age and racial/ethnic groups. These analyses were repeated using a 6-month re-referral window as a robustness check.

- **Likelihood a screened-in child had a case opened for services upon investigation, or had a re-referral within 2 months if no case was opened**

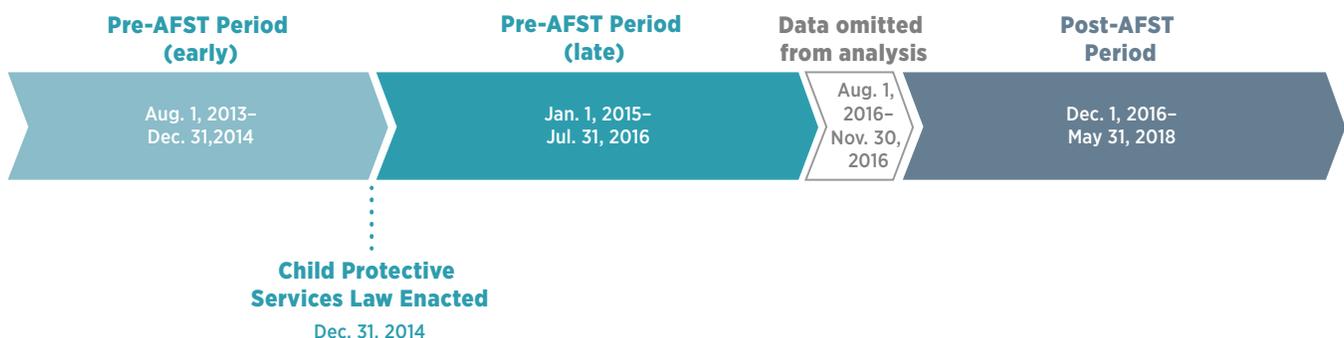
This outcome is intended to measure how AFST implementation impacted one feature of accuracy: do children who are screened-in for investigation evidence safety and service needs requiring child protective services (i.e., case opened initially, or a re-referral if not)?

A child is considered to have experienced this third outcome if a referral that includes the child is screened-in (i.e., advanced by the hotline screener and supervisor for investigation) and upon investigation, one of two things happens: (1) a child protective service case is opened by the investigating worker (indicating that safety concerns and service needs were identified); or (2) a child protective service case is not opened by the investigating worker and the child is re-referred within 2 months of the original referral (indicating safety concerns and service needs were identified by the hotline screener, but may have been improperly addressed by the investigating worker). The rate of screen-ins requiring services was defined as the total number of screened-in children meeting criteria 1 or 2 above (numerator), divided by the total number of children screened in for investigation (denominator), computed for referrals falling in each calendar month and for children in different age and racial/ethnic groups.

It should be noted that case openings were deemed to provide a good measure for the purposes of evaluation because the investigator does not see the AFST score. As such, the investigating worker's decision is made independent of the score.

Evaluation Window

The entire evaluation window spans August 1, 2013 through May 31, 2018. For the purpose of the analyses, the data were divided into multiple periods:



- **Pre-AFST Period.** The “Pre-AFST Period” spans August 1, 2013 through July 31, 2016 and was divided in two parts:

- The Pre-AFST Period (early) spans August 1, 2013 through December 31, 2014
- The Pre-AFST Period (late) spans January 1, 2015 through July 31, 2016

The evaluator’s decision to divide the Pre-AFST Period is based on a set of amendments to the State of Pennsylvania’s existing Child Protective Services Law, which became effective on December 31, 2014 and had the effect of altering a number of features of referrals to the call center.⁵ The second (late) Pre-AFST Period served as the point of most comparison in the analysis.

⁵ Child Protective Services Act, P.L.1240, No.206, 23 PA §§6301-6386 (2015).

- **Post-AFST Period.** The period after the full implementation period is termed the “Post-AFST Period” and spans December 1, 2016 through May 31, 2018. Outcomes during the Post-AFST Period are compared to outcomes in periods prior to this.

Notably, data for the period between August 1, 2016 and November 30, 2016 are omitted from all analyses. When the AFST was launched, an initial policy decision sought to restrict score generation to only individuals and families who could be substantively identified in prior county data (preventing scores from being displayed that were solely constructed from basic referral/geographic information when the family was otherwise unknown to DHS). This policy initially restricted scores to situations where a child on the call was positively identified with a prior county identifier (meaning that the child was previously known to DHS and had been assigned a unique identifying client number). Many children, most notably newborns and infants, who experience the highest rates of maltreatment and fatalities, often do not have system involvement, and therefore do not have their own county identifier, but their parents or caregivers may have significant current and prior system involvement. The initial design led to situations where known information about adults on the referral could not be used in generating a score if none of the children were recognized, and this was quickly deemed too restrictive. After November 30, 2016, scores could be generated for a referral if any individual named, child or adult, could be matched to a county identifier. Given the non-random nature of the children who did not receive AFST scores prior to December 1, 2016, analyses of this data cannot reliably attribute observed changes in outcomes pre- and post-implementation to the AFST score.

Data

All analyses use de-identified (anonymized) data relating to individuals named in maltreatment referrals made to Allegheny County’s child protective services hotline. The data consist of information about individual household members including their race, legal sex and age. Additionally, the data identify the call screeners and supervisors associated with the referral and track previous referrals and investigations with child welfare and other child-serving systems from August 1, 2013 through May 31, 2018.

The analytic dataset focuses on outcomes (described above) for children below 18 years at the time of referral.

The analytic dataset also contains several variables used in the analysis to control for child demographics (age, legal sex, race) and household characteristics (household counts and composition, socioeconomic status and maximum risk scores). For child's race, the evaluation used a categorical variable which included the category "Unable to Determine", when race was not coded as white, Black/African American, or other. Other control variables had complete data.

Analytic Approach

Three main types of analyses are reported for each of the three main outcomes. Comparisons of unadjusted population means and the Interrupted Time Series Analysis (ITSA) describe levels and changes in outcomes within the County's child maltreatment screening system overall, and for age and race/ethnic subgroups given existing trends in the Pre-AFST period (late) (January 1, 2015 through July 31, 2016) and the Post-AFST period (December 1, 2016 through May 31, 2018). Individual-level multivariate regression analyses focus on changes after adjustment for changes in referral case mix over time. The evaluation team also considered the AFST's effects on outcomes for subgroups of children defined in terms of their age group and racial/ethnic characterization. This enabled potential heterogeneity and disparities in the policy's effects across subgroups.

- **Unadjusted Population Means.** The simplest comparison performed was a comparison of unadjusted means for the Pre- and the Post-AFST periods, testing whether they are statistically different from one another using a two-sided t-test of equality of means.
- **Interrupted Time Series Analysis.** Changes in the level and trend of monthly rates of each outcome during the Pre- and Post- periods were assessed using an Interrupted Time Series Analysis. In this evaluation, the ITSA measures changes in both the level and slope of each outcome in the Post-AFST months in relation to the Pre-AFST months. The ITSA approach captures population-level changes in outcomes and trends after a policy change in comparison to the levels and trends prior to that change.
- **Child-Level Multivariate Regression Analysis.** Finally, the evaluators used multivariate individual-level regression analyses to assess the impact of the AFST on the predicted level of each outcome Pre- and Post-AFST, while adjusting for child and household characteristics. These analyses focus on estimates of the average effect of the AFST, adjusting for evolving case mix over time. The predictive margins presented in evaluation tables and figures can be interpreted as the average outcome if all children in the sample were in either the Pre-AFST or the Post-AFST time-frame, holding all other control variables constant.

RESULT HIGHLIGHTS

Overall rates of children screened-in for investigation

All children screened-in

- Prior to the implementation of the AFST, the number of maltreatment referrals had increased during the Pre-AFST period. This increase in total referrals corresponded with state law and policy changes that expanded mandatory reporting. With an increase in the number of referrals received, the fraction of all referrals screened-in for investigation began to decline. The AFST largely halted this decline in screened-in investigations for all groups. Even though the average screen-in level in the Pre-AFST period was higher than in the Post-AFST period, it is unknown whether screen-in rates would have continued to decline.

Children screened-in, by age group

- Prior to the implementation of the AFST, the fraction of referrals screened-in for investigation was declining for children in all subgroups except for ages 4 to 6, with larger declines observed in the oldest age group (13 to 17 years). The AFST largely halted these age-specific declines, most noticeably for children ages 7 years and older.

Children screened-in, by race

- Prior to the implementation of the AFST, the fraction of referrals screened-in for investigation were declining for children in all race groups, with larger declines observed for Black/African American children than for white children.
- The AFST largely halted all race-specific declines, most noticeably for Black/African American children.

Call screener consistency

- There was moderate consistency in the referral screen-in outcome across call screeners.
- Screen-in rates increased in the Post-AFST period for 7 of the 11 call screeners (4 of these were significant increases) and decreased for 4 of the 11 call screeners (none statistically significantly).
- The variance of call screener outcomes decreased for children in both Black/African American and white race groups, with a larger effect apparent in the Black/African American group (though not statistically significantly).

Likelihood a screened-out child had no re-referrals within 2 months

Re-referrals of all children

- In breaking down the AFST's effect on decisions to screen out children without investigation, there was a small increase in the overall rate at which screened-out children were re-referred.

Re-referrals by age group

- The increase in re-referral rate was concentrated among children in the 4-to-6-year-old age group. Among all other age groups, reductions in the likelihood of being re-referred after being screened out were non-significant.
- The observed increase in re-referral rates among 4-to-6-year-olds after the implementation of the AFST is likely due to corresponding changes in the County's policy regarding the maximum age for mandatory field screening. With implementation of the AFST, the County reduced the age for mandatory field screening from under 7 years of age to under 4 years of age. It may be that previous field screenings in this age group helped to identify more children for whom a screen-in for investigation was appropriate.

Re-referrals by race group

- Multiple analyses showed small increases in re-referral rates for both race subgroups, which were not significant for white children and only occasionally significant for Black/African American children.

Re-referrals by call screeners

- The absence of changes in re-referral rates for children screened-out was consistent across call-screeners, and variation between calls screeners in this outcome did not change significantly. The variance of call-screener-specific outcomes increased slightly (not statistically significantly) in the Post-AFST period compared to the Pre-AFST period.
- Results were similar when evaluators used a re-referral window of 6 months instead of two months.

Likelihood a screened-in child had a case opened for services upon investigation, or had a re-referral within 2 months if no case was opened

- The AFST increased the identification of higher-need children (measured as those children determined to be in need of further child welfare intervention, i.e., those who, after being screened-in, had cases opened for child protective services or, if no case was opened, had a re-referral within 2 months). It should again be noted that the investigating worker and supervisor, those making the decision to open a case, did not have access to the score.
- Increase in the identification of children determined to be in need of further child welfare intervention emerged across age and racial subgroups.
- While changes to screening-in higher-need children remained throughout the Post-AFST period, the initial improvement effect did attenuate somewhat over time.
- With a re-referral window of 6 months, the direction of the result was the same and there was somewhat less attenuation over time.
- The AFST had an immediate upward effect on the likelihood of screening-in children later determined to be in need of further child welfare intervention for both white and Black/African American children.

- For Black/African American children, the initial improvement effect of implementing the AFST attenuated over time, whereas for white children the effect was more persistent.
- Whereas prior to the AFST, Black/African American children had a higher rate of case openings after screen-in than white children, this disparity was reduced over time in the post-AFST period.
- The overall change in screen-ins for children determined to be in need of further child welfare intervention was consistent across call screeners, and variation between calls screeners in this outcome did not change significantly.

CONCLUSIONS

Overall, analyses documented that the AFST and associated policies:

- increased the accuracy of decisions about children screened-in for investigation and
- did not increase the number of children screened-in for investigation (as compared to the average during the pre-AFST period).

Among children screened-out without an investigation, there was a slight increase in the re-referral rates for children between 4 and 6 years of age, the group of children directly affected by corresponding policy changes to mandatory in-home assessments (i.e., field screening). The County will look at this finding carefully and will work with call screening staff to understand why some of the children in this age group who had a high AFST score, and who were later re-referred and screened-in, were screened-out at this initial referral.

One of the key topics addressed by the evaluation was the effect of the implementation of the AFST and surrounding policy changes on disparities in outcomes across race/ethnicity. Overall, changes in a type of “accuracy” measure (i.e., an increase in accuracy for children screened-in for investigation and a negligible or slight decrease in accuracy for children screened out) were consistently observed for both Black/African American and white children.

It should also be noted, however, that if community referrals are biased by race, then even appropriate screen-outs for Black/African American children might look “inappropriate” because they get re-referred. Therefore, we need to treat the racial differences in this outcome measure with caution. The fact that two-thirds of children who were defined as inaccurately screened-out in our analysis are Black/African American might be suggestive that broader (and more objective) definitions should be considered.

The AFST was hypothesized to result in greater screener consistency but the evaluation detected no such improvements. It’s worth noting that the improvements would have to have been relatively large to be detected; however, there are also other possible explanations for the lack of improved consistency here. First, with the implementation of the AFST, the County enhanced the autonomy of the call screener who now makes a recommendation to their supervisor. Prior to the implementation of the AFST, decisions were collaborative in nature

but were ultimately made by the supervisor. This practice change enhanced the opportunity for variability in decision making, perhaps reducing any improvements that the AFST might have shown. Further, there is considerable lack of concurrence with the AFST by call screeners, limiting the ability for a tool like the AFST to have effect on consistency. For the period December 1, 2016–November 29, 2018, only 61 percent of the referrals that scored in the “mandatory” screen-in range were in fact screened in (Table 1). Therefore, the County will continue to try to work with call screeners to understand why they might be making these decisions. We hope to have more data and therefore more power to measure screener variation in the next stage of the evaluation.

TABLE 1: Percent of children screened-in for investigation, by AFST risk level

	PERCENT SCREENED- IN FOR INVESTIGATION
Mandatory	61%
High	47%
Medium	42%
Low	31%
No Score	23%
Total	41.4%

The evaluation team concluded that “the effects of implementing the AFST and surrounding policy changes show moderate improvements in accuracy of screen-ins with small decreases in the accuracy in screen-outs, a halt in the downward trend in pre-implementation screen-ins for investigation, no large or consistent differences across race/ethnic or age-specific subgroups in these outcomes, and no large or substantial differences in consistency across call-screeners.”

The County is encouraged that the AFST has shown some effect on the accuracy of decision making and reductions in overall case opening disparities between black and white children, particularly in the face of implementation challenges (for a discussion of technical, practice, and policy challenges please see the [FAQ](#)). More importantly, there was no evidence of unintended adverse effects.

The evaluation aligns with what the leadership of the County have observed: that the tool has tremendous potential and that there are few, if any, unintended adverse effects given workers’ willingness to use their own discretion in the screening decision. But implementation challenges were significant and persist; these must be overcome to maximize the impact of automated risk stratification tools.

We will continue to work to improve the model and its implementation. As of November 2018, the County released a model with significant enhancements (for more information on this, see [Methodology, Version 2](#) and [FAQ](#)). We will also continue the evaluation and will ask Stanford University to consider streamlining their methods and examine:

- how previously-defined outcomes (defined in this evaluation summary) changed with the implementation of Version 2 of the tool, stratified by race/ethnicity, age, and AFST score.
- the impact of the AFST and associated policies on home removals.
- consistency of outcomes across supervisors (instead or in addition to examining call screeners, given decision making processes).
- the impact of the high- and low-risk protocols on decision making.