Allegheny Family Screening Tool

FREQUENTLY-ASKED QUESTIONS

JULY 2017

CONTENTS

INTRODUCTION

In August 2016, the Allegheny County Department of Human Services (DHS) implemented the Allegheny Family Screening Tool (AFST), a predictive risk modeling tool designed to improve child welfare call screening decisions. The AFST was the result of a two-year process of exploration about how existing data could be used more effectively to improve decision-making at the time of a child welfare referral. For more information about the AFST, see http://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx.

The process began in 2014 with a Request for Proposals and selection of a team from Auckland University of Technology led by Rhema Vaithianathan and including Emily Putnam-Hornstein from University of Southern California, Irene de Haan from the University of Auckland, Marianne Bitler from University of California – Irvine and Tim Maloney and Nan Jiang from Auckland University of Technology. Prior to implementation, the model was subjected to an ethical review by Tim Dare of the University of Auckland and Eileen Gambrill of the University of California-Berkeley. Upon the conclusion of this review, to which DHS prepared a response, the County proceeded with implementation. Concurrent with this process was the issuance of a second Request for Proposals, at the end of 2015, for an impact and process evaluation of the model. Contracts were awarded to Stanford University (impact evaluation) and Hornby Zeller Associates (process evaluation). The results of these evaluations, expected by the end of 2017, will be made publicly available.

A report on the development of the AFST,[1] prepared by Rhema Vaithianathan, PhD; Nan Jiang, PhD; Tim Maloney, PhD; Parma Nand, PhD; and Emily Putnam-Hornstein, PhD), was published in April 2017. The following Frequently-Asked Questions are presented as a quick reference for those interested in highlights from this publication and should be considered within the context of the full publication. Page numbers are provided throughout the document, indicating where the reader may find more detailed information.

1 Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening

BACKGROUND

What is the Allegheny Family Screening Tool (AFST) and how does it work?

The AFST was developed to support one key decision in the child welfare process: whether to screen-in a referral for investigation, or screen it out.

To generate the AFST scores, the AFST uses more than 100 predictive factors for each child on the referral. These factors are then weighted through a logistic regression model to calculate two AFST scores (ranging from 1–20) for each child: the risk of placement within two years if the referral is screened-in and the risk of re-referral within two years if the referral is screened-out. Call screeners and supervisors see the maximum AFST score from the referral. For example, if there are two children on the referral and one has a maximum risk score of 12 and the other has a maximum risk score of 16, the call screener will see a score of 16.

It should be noted that while in some settings machines have been used to make decisions that were previously made by humans, this is not the case for the AFST. It was never intended or suggested that the algorithm would replace human decision-making. Rather, the AFST should help to inform, train and improve the decisions made by the child welfare staff.

Who are the key partners and how were they selected?

The Allegheny County Department of Human Services (DHS) issued a Request for Proposals (RFP) in 2014, to design and implement a system of decision-support tools and predictive analytics for human services.[2]

2 Decision Support Tools and Predictive Analytics in Human Services RFP

We received 15 proposals in response to the RFP. After review by an evaluation committee, researchers from Auckland University of Technology (AUT), University of Southern California (USC), University of California-Berkeley and University of Auckland were awarded the contract and conducted the work. The research team was led by Rhema Vaithianathan (AUT).

Has the local community been involved in the decision to use the AFST?

Community engagement has been a priority for the County throughout the project. The County sought input from the community through various meetings, including six project-specific meetings. Three were held at early stages of the project to collect feedback from key external stakeholders and funders. DHS then held three open community meetings where over 30 stakeholder groups (including the Courts and the ACLU) were invited to discuss the work to date, implementation timeline and results. Additionally, DHS shared project updates with existing community networks including the Children's Cabinet and the Children, Youth and Families Advisory Board, and through the DHS Speaker Series. Feedback from these community meetings has influenced the project throughout its development.

How will the AFST be evaluated?

There are two independent evaluations of the AFST in progress; the evaluators were selected through an RFP at the end of 2015.[3] The impact evaluation is being conducted by Stanford University and the process evaluation is being conducted by Hornby Zeller Associates, Inc. The impact evaluation will focus on whether the AFST increases the accuracy of decisions, reduces unwarranted variation in decision-making and reduces disparities, and will also examine overall referral rates and workload. The process evaluation is designed to assess implementation.

3 Evaluation of a Predictive Risk Modeling Tool for Improving the Decisions of Child Welfare Workers RFP

THE MODEL

Doesn't the AFST just predict child welfare system decision-making?

A challenge is to identify outcomes to predict that are truly independent of the system and not too rare to be predicted.

The first adverse outcome predicted by the AFST is placement within two years of screen-in. Because placements are determined by a judge, and all parties (parents, children and County) are represented by attorneys, a placement outcome is reasonably independent of the County child welfare system.

The second adverse outcome that the AFST predicts — re-referral after an initial referral has been screened-out — is independent of the County child welfare system because referrals come from the community.

What data does the AFST use?

The AFST uses information from DHS's integrated data system that links administrative data from 29 sources including child protective services, publicly funded mental health and drug and alcohol services, and bookings in the County jail. Please **see page 11** of the methodology and implementation report for additional information on the data used.

Does the AFST use race as a factor?

No. The County made the decision not to include race as a factor in the AFST because including race does not improve the accuracy of the score. This doesn't mean, however, that other variables in the tool aren't correlated with race. There are other predictors that are correlated with race due to potentially institutionalized racial bias (e.g., criminal justice history) that would imply that race is still a factor. For this reason, continued monitoring of application of the model with regard to racial disparities should be undertaken.

Please **see page 29** of the methodology and implementation report for additional information on the impact of race as a predictor and *Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County.*

Does the AFST magnify race bias in data?

The data used by call screeners in the existing system, and now used to run the AFST, contain race bias. In numerous data sets, biased practices translate into bias within data that cannot be removed. For example, jail bookings aren't necessarily a function of crime but of criminal justice policies as well as arrest and jailing policy and practices.

When data contain bias, which is invariably the case, decisions based on that data will perpetuate the bias to some extent. This will be the case regardless of how the data are used. Bias in data has been an ongoing challenge within the existing Allegheny County child welfare decision-making system and any new decision-making approach must also work with this 'imperfect' data.

Therefore, as part of the AFST impact evaluation, we need to compare how the AFST works with how the existing system works. Researchers found that, in addition to the inherent race bias in data, decisions made under the existing system are affected by race.[4]

The fairness of algorithms is an ongoing issue for researchers in this field and the AFST research team will continue to monitor how that research impacts the AFST.

4 Maloney, Tim, et al. "Black–White Differences in Child Maltreatment Reports and Foster Care Placements: A Statistical Decomposition Using Linked Administrative Data." Maternal and Child Health Journal 21.3 (2017): 414-420.

Does the AFST use prior allegations of maltreatment as a factor?

Yes, because historical data tell us that previous reports of maltreatment, substantiated or not, have predictive power (there is no factor included in the model that does not have significant predictive power).

How accurate is the AFST?

Measuring the accuracy of predictive tools is not simple; however, at rollout, the accuracy of the AFST was described as comparable to a mammogram: 77 percent accuracy for predicting whether a child would be placed in care within two years after being referred and screened-in for investigation, and 73 percent accuracy for predicting whether a child would be re-referred within two years after being referred and screened-out for investigation. At six-month rebuild, we intend to add an additional flag for mandatory screen-in, which is generated by a Random Forest Model which has accuracy of 88 percent (which is substantially higher than a mammogram).

Please **see page 15** of the methodology and implementation report for additional information on model performance.

Has the AFST been validated?

In addition to assessing the accuracy of the AFST in predicting placement and re-referral, the research team also conducted an external validation looking at hospital events (emergency department visits and inpatient admissions). Findings show that over a broad range of injury types there is a positive correlation between the placement scores generated by the AFST at referral and the rate of hospital events.

For example, those children with a placement risk score of 20 (the highest possible score) have a hospital event rate for self-inflicted injury or suicide of 0.65 percent compared to 0.03 percent for those with a placement risk score of 1 (the lowest possible score). That is, a child who scores a 20 at referral is 21 times more likely to be hospitalized for a self-inflicted injury than a child who scores a 1.

Please **see page 19** of the methodology and implementation report for additional information on the hospital validation study.

What did the research tell us about existing practice?

Prior to introduction of the AFST, call screeners could access and use historical and cross-sector administrative data related to individuals associated with a report of child abuse or neglect through Client View, a front-end application to the integrated data system. Call screeners were required to review all relevant information related to a referral and provide it to the call screening supervisor so that a screen-in/screen-out decision could be made. However, it was challenging for call screeners to efficiently access, review and make meaning of all available records. The AFST provides a consistent way to access and weight the available information to predict the risk of future adverse events for each child on the referral.

Researchers found that existing practice had screened out one in four children who the model would screen-in due to their score. For these children, who the model scored as highest risk, 9 in 10 were re-referred and half were placed in foster care within two years. Forty-eight percent of the lowest-risk cases were screened-in with only one percent of these referrals leading to placement within two years.

What happens when there is missing/duplicate information?

The AFST leverages a probabilistic matching algorithm to catch as many duplicate IDs as possible. This method, however, does not capture all duplicate IDs for the same person and, thus, it is possible for an AFST score to exclude data held on a second ID. Efforts to minimize duplicate client records are ongoing.

Is the AFST score assigned to a child/family permanently?

No, because the AFST score will change as underlying data change. The County will retain AFST scores for quality assurance and evaluation purposes.

What safeguards are in place to make sure the AFST is working appropriately?

Immediately before the AFST was put into operation, researchers validated the scores generated by the DHS Data Warehouse (for individuals in historical, de-identified data) by generating scores for the same individuals in the research environment, to ensure that the Data Warehouse was accurately running the AFST. Since implementation, County child welfare leadership has been reviewing monthly quality assurance reports to monitor the performance of the AFST.

AFST scores are securely stored and cannot be manually altered by call screeners. However, as an additional quality assurance check, DHS is proposing to add functionality to the AFST that will allow workers to report when a score seems wrong/surprising to them. All reported referrals will be reviewed by the research team.

An independent impact evaluation (which will assess the effectiveness of the AFST compared to the existing approach) and process evaluation (which will evaluate how the AFST is implemented) will alert the County and research team to any concerns about the effectiveness and operation of the AFST.

Will the County improve the AFST over time?

The AFST has already been rebuilt once by the research team since it came into use in August 2016, taking learnings from practice and using those to optimize how the AFST scores are generated. The County intends to build a "Version 2" of the AFST that will include improvements identified by evaluation of the AFST.

How does the AFST compare to other approaches?

The AFST has a similar purpose to other tools like the Structured Decision Making tool (SDM), but creates a score without the reliance on manual data input that is required for SDM. For the highest category of risk, the AFST outperformed the SDM model.

Please **see page 24** of the methodology and implementation report for additional information on comparing the model to SDM and rule-based/threshold approaches.

PRACTICE

Who gets an AFST score and how?

All children involved in an allegation of maltreatment, regardless of whether they are described as the victim or not, will be included in the AFST score; that is, all children living in the same household or added to the case by the call screener. When an allegation of maltreatment is received and the call screener enters details into the child welfare case management system (KIDS), a click will automatically generate the AFST score. Call screeners and call screening supervisors are required to generate the AFST score prior to finalizing a screening decision.

Are there some children for whom an AFST score can't be generated?

Yes, those not known to the system and those for whom not enough data are held in the Data Warehouse. The County has determined that the AFST will only be used to screen for risk when data that goes beyond demography (e.g., age, gender, address) are held for one or more person associated with the allegation. If only demographic data are held for all individuals, then the allegation will be assessed using the existing approach (no AFST score will be generated). As of April 2017, approximately 10 percent of incoming referrals were not generating an AFST score.

Who gets access to the AFST score?

Only the call screener and call screening supervisor have access to the AFST score. If and when a referral moves to the investigation stage, investigations staff cannot access any AFST score. The Courts also do not have access to the AFST score.

Please **see page 26** of the methodology and implementation report for additional information on the implementation of the AFST score.

Does a certain AFST score make screening-in mandatory?

The AFST flags some scores as "mandatory screen-ins." The threshold for the mandatory screen-in was determined solely by the placement score and designed to capture as many of the children at heightened risk of abuse-related fatal or near-fatal injuries (Act 33 Events) as possible. The model includes functionality that allows call screening supervisors to override the "mandatory screen-ins" at their discretion; overrides are documented and reviewed.

Please **see page 26** of the methodology and implementation report for additional information on mandatory screen-ins.

Will caseworkers be afraid to 'defy the score?'

The only caseworkers who make screen-in/screen-out decisions are the call screening supervisors. They consider all information provided by the call screeners, including details shared during the call, by the person alleging abuse or neglect, the score generated by the AFST and recommendations from the call screener.

Screening decisions are not in any way 'dictated' by the AFST. Call screening supervisors have full discretion over call screening decisions, regardless of generated AFST scores.

How is the risk of stigma minimized in the AFST?

No system can entirely remove the chance of screening-in some of the 'wrong' children, so wrongly stigmatizing them. The ethicists suggest, however, that we must then take a comparative view: Is the proposed tool as good or better than the existing approach, when it comes to minimizing the risk of stigma? Compared to the existing system, the AFST is expected to increase accuracy and consistency of decision-making, which means wrongful stigma is expected to be reduced. The impact evaluation will assess this.

In particular, the County will work to minimize stigmatization by carefully controlling access to AFST scores and providing appropriate training that aims to reduce stigmatization and ensures that call screeners are aware of the possibility of false positives/negatives and understand the risk of confirmation bias.

Has the AFST significantly increased the number of investigations?

No. As of April 2017 (with eight months of AFST experience), average screen-in rates have been nearly identical to rates for the same period of time one year prior (approximately 43%).