

Can an Algorithm Tell When Kids Are in Danger?

Child protective agencies are haunted when they fail to save kids. Pittsburgh officials believe a new data analysis program is helping them make better judgment calls.

By DAN HURLEY JAN. 2, 2018

The call to Pittsburgh's hotline for child abuse and neglect came in at 3:50 p.m. on the Wednesday after Thanksgiving 2016. Sitting in one of 12 cubicles, in a former factory now occupied by the Allegheny County Police Department and the back offices of the department of Children, Youth and Families, the call screener, Timothy Byrne, listened as a preschool teacher described what a 3-year-old child had told him. The little girl had said that a man, a friend of her mother's, had been in her home when he "hurt their head and was bleeding and shaking on the floor and the bathtub." The teacher said he had seen on the news that the mother's boyfriend had overdosed and died in the home.

According to the case records, Byrne searched the department's computer database for the family, finding allegations dating back to 2008: parental substance abuse, inadequate hygiene, domestic violence, inadequate provision of food and physical care, medical neglect and sexual abuse by an uncle involving one of the girl's two older siblings. But none of those allegations had been substantiated. And while the current claim, of a man dying of an overdose in the child's home, was shocking, it fell short of the minimal legal requirement for sending out a caseworker to knock on the family's door and open an investigation.

Before closing the file, Byrne had to estimate the risk to the child's future well-being. Screeners like him hear far more alarming stories of children in peril nearly every day. He keyed into the computer: "Low risk." In the box where he had to select the likely threat to the children's immediate safety, he chose "No safety threat."

Had the decision been left solely to Byrne — as these decisions are left to screeners and their supervisors in jurisdictions around the world — that might have been the end of it. He would have, in industry parlance, screened the call out. That's what happens to around half of the 14,000 or so allegations received each year in Allegheny County — reports that might involve charges of serious physical harm to the child, but can also include just about anything that a disgruntled landlord, noncustodial parent or nagging neighbor decides to call about. Nationally, 42 percent of the four million allegations received in 2015, involving 7.2 million children, were screened out, often based on sound legal reasoning but also because of judgment calls, opinions, biases and beliefs. And yet more United States children died in 2015 as a result of abuse and neglect — 1,670, according to the federal Administration for Children and Families; or twice that many, according to leaders in the field — than died of cancer.

This time, however, the decision to screen out or in was not Byrne's alone. In August 2016, Allegheny County became the first jurisdiction in the United States, or anywhere else, to let a predictive-analytics algorithm — the same kind of sophisticated pattern analysis used in credit reports, the automated buying and selling of stocks and the hiring, firing and fielding of baseball players on World Series-winning teams — offer up a second opinion on every incoming call, in hopes of doing a better job of identifying the families most in need of intervention. And so Byrne's final step in assessing the call was to click on the icon of the Allegheny Family Screening Tool.

After a few seconds, his screen displayed a vertical color bar, running from a green 1 (lowest risk) at the bottom to a red 20 (highest risk) on top. The assessment was based on a statistical analysis of four years of prior calls, using well over 100 criteria maintained in eight databases for jails, psychiatric services, public-welfare benefits,

Over the course of an 18-month investigation, officials in the county's Office of Children, Youth and Families (C.Y.F.) offered me extraordinary access to their files and procedures, on the condition that I not identify the families involved. Exactly what in this family's background led the screening tool to score it in the top 5 percent of risk for future abuse and neglect cannot be known for certain. But a close inspection of the files revealed that the mother was attending a drug-treatment center for addiction to opiates; that she had a history of arrest and jail on drug-possession charges; that the three fathers of the little girl and her two older siblings had significant drug or criminal histories, including allegations of violence; that one of the older siblings had a lifelong physical disability; and that the two younger children had received diagnoses of developmental or mental-health issues.

Finding all that information about the mother, her three children and their three fathers in the county's maze of databases would have taken Byrne hours he did not have; call screeners are expected to render a decision on whether or not to open an investigation within an hour at most, and usually in half that time. Even then, he would have had no way of knowing which factors, or combinations of factors, are most predictive of future bad outcomes. The algorithm, however, searched the files and rendered its score in seconds. And so now, despite Byrne's initial skepticism, the high score prompted him and his supervisor to screen the case in, marking it for further investigation. Within 24 hours, a C.Y.F. caseworker would have to "put eyes on" the children, meet the mother and see what a score of 19 looks like in flesh and blood.

For decades, debates over how to protect children from abuse and neglect have centered on which remedies work best: Is it better to provide services to parents to help them cope or should the kids be whisked out of the home as soon as possible? If they are removed, should they be placed with relatives or with foster parents? Beginning in 2012, though, two pioneering social scientists working on opposite sides of the globe — Emily Putnam-Hornstein, of the University of Southern California, and Rhema Vaithianathan, now a professor at the Auckland University of Technology in New Zealand — began asking a different question: Which families are most at risk and in need of help? "People like me are saying, 'You know what, the quality of the services you provide might be just fine — it could be that you are providing them to the wrong families,'" Vaithianathan told me.

Vaithianathan, who is in her early 50s, emigrated from Sri Lanka to New Zealand as a child; Putnam-Hornstein, a decade younger, has lived in California for years. Both share an enthusiasm for the prospect of using public databases for the public good. Three years ago, the two were asked to investigate how predictive analytics could improve Allegheny County's handling of maltreatment allegations, and they eventually found themselves focused on the call-screening process. They were brought in following a series of tragedies in which children died after their family had been screened out — the nightmare of every child-welfare agency.

One of the worst failures occurred on June 30, 2011, when firefighters were called to a blaze coming from a third-floor apartment on East Pittsburgh-McKeesport Boulevard. When firefighters broke down the locked door, the body of 7-year-old KiDonn Pollard-Ford was found under a pile of clothes in his bedroom, where he had apparently sought shelter from the smoke. KiDonn's 4-year-old brother, KrisDon Williams-Pollard, was under a bed, not breathing. He was resuscitated outside, but died two days later in the hospital.

The children, it turned out, had been left alone by their mother, Kiaira Pollard, 27, when she went to work that night as an exotic dancer. She was said by neighbors to be an adoring mother of her two kids; the older boy was getting good grades in school. For C.Y.F., the bitterest part of the tragedy was that the department had received numerous calls about the family but had screened them all out as unworthy of a full investigation.

Incompetence on the part of the screeners? No, says Vaithianathan, who spent months with Putnam-Hornstein burrowing through the county's databases to build their algorithm, based on all 76,964 allegations of maltreatment made between April 2010 and April 2014. "What the screeners have is a lot of data," she told me, "but it's quite difficult to navigate and know which factors are most important. Within a single call to C.Y.F., you might have two children, an alleged perpetrator, you'll have Mom, you might have another adult in the household — all these people will have histories in the system that the person screening the call can go investigate. But the human brain is not that deft at harnessing and making sense of all that data."

She and Putnam-Hornstein linked many dozens of data points — just about everything known to the county about each family before an allegation arrived — to predict how the children would fare afterward. What they found was startling and disturbing: 48 percent of the lowest-risk families were being screened in, while 27 percent of the highest-risk families were being screened out. Of the 18 calls to C.Y.F. between 2010 and 2014 in which a child was later killed or gravely injured as a result of parental maltreatment, eight cases, or 44 percent, had been screened out as not worth investigation.

According to Rachel Berger, a pediatrician who directs the child-abuse research center at Children’s Hospital of Pittsburgh and who led research for the federal Commission to Eliminate Child Abuse and Neglect Fatalities, the problem is not one of finding a needle in a haystack but of finding the right needle in a pile of needles. “All of these children are living in chaos,” she told me. “How does C.Y.F. pick out which ones are most in danger when they all have risk factors? You can’t believe the amount of subjectivity that goes into child-protection decisions. That’s why I love predictive analytics. It’s finally bringing some objectivity and science to decisions that can be so unbelievably life-changing.”

The morning after the algorithm prompted C.Y.F. to investigate the family of the 3-year-old who witnessed a fatal drug overdose, a caseworker named Emily Lankes knocked on their front door. The weathered, two-story brick building was surrounded by razed lots and boarded-up homes. No one answered, so Lankes drove to the child’s preschool. The little girl seemed fine. Lankes then called the mother’s cellphone. The woman asked repeatedly why she was being investigated, but agreed to a visit the next afternoon.

The home, Lankes found when she returned, had little furniture and no beds, though the 20-something mother insisted that she was in the process of securing those and that the children slept at relatives’ homes. All the appliances worked. There was food in the refrigerator. The mother’s disposition was hyper and erratic, but she insisted that she was clean of drugs and attending a treatment center. All three children denied having any worries about how their mother cared for them. Lankes would still need to confirm the mother’s story with her treatment center, but for the time being, it looked as though the algorithm had struck out.

Charges of faulty forecasts have accompanied the emergence of predictive analytics into public policy. And when it comes to criminal justice, where analytics are now entrenched as a tool for judges and parole boards, even larger complaints have arisen about the secrecy surrounding the workings of the algorithms themselves — most of which are developed, marketed and closely guarded by private firms. That’s a chief objection lodged against two Florida companies: Eckerd Connects, a nonprofit, and its for-profit partner, MindShare Technology. Their predictive-analytics package, called Rapid Safety Feedback, is now being used, the companies say, by child-welfare agencies in Connecticut, Louisiana, Maine, Oklahoma and Tennessee. Early last month, the Illinois Department of Children and Family Services announced that it would stop using the program, for which it had already been billed \$366,000 — in part because Eckerd and MindShare refused to reveal details about what goes into their formula, even after the deaths of children whose cases had not been flagged as high risk.

The Allegheny Family Screening Tool developed by Vaithianathan and Putnam-Hornstein is different: It is owned by the county. Its workings are public. Its criteria are described in academic publications and picked apart by local officials. At public meetings held in downtown Pittsburgh before the system’s adoption, lawyers, child advocates, parents and even former foster children asked hard questions not only of the academics but also of the county administrators who invited them.

“We’re trying to do this the right way, to be transparent about it and talk to the community about these changes,” said Erin Dalton, a deputy director of the county’s department of human services and leader of its data-analysis department. She and others involved with the Allegheny program said they have grave worries about companies selling private algorithms to public agencies. “It’s concerning,” Dalton told me, “because public welfare leaders who are trying to preserve their jobs can easily be sold a bill of goods. They don’t have a lot of sophistication to evaluate these products.”

Another criticism of such algorithms takes aim at the idea of forecasting future behavior. Decisions on which families to investigate, the argument goes, should be based solely on the allegations made, not on predictions for what might happen in the future. During a 2016 White House panel on foster care, Gladys Carrión, then the

commissioner of New York City's Administration for Children's Services, expressed worries about the use of predictive analytics by child-protection agencies. "It scares the hell out of me," she said — especially the potential impact on people's civil liberties. "I am concerned about widening the net under the guise that we are going to help them."

But in Pittsburgh, the advocates for parents, children and civil rights whom I spoke with all applauded how carefully C.Y.F. has implemented the program. Even the A.C.L.U. of Pennsylvania offered cautious praise. "I think they're putting important checks on the process," said Sara Rose, a Pittsburgh lawyer with the organization. "They're using it only for screeners, to decide which calls to investigate, not to remove a child. Having someone come to your home to investigate is intrusive, but it's not at a level of taking a child away or forcing a family to take services."

The third criticism of using predictive analytics in child welfare is the deepest and the most unsettling. Ostensibly, the algorithms are designed to avoid the faults of human judgment. But what if the data they work with are already fundamentally biased? There is widespread agreement that much of the underlying data reflects ingrained biases against African-Americans and others. (Just last month, the New York City Council voted to study such biases in the city's use of algorithms.) And yet, remarkably, the Allegheny experience suggests that its screening tool is *less* bad at weighing biases than human screeners have been, at least when it comes to predicting which children are most at risk of serious harm.

"It's a conundrum," Dalton says. "All of the data on which the algorithm is based is biased. Black children are, relatively speaking, over-surveilled in our systems, and white children are under-surveilled. Who we investigate is not a function of who abuses. It's a function of who gets reported."

In 2015, black children accounted for 38 percent of all calls to Allegheny County's maltreatment hotline, double the rate that would be expected based on their population. Their rate of being placed outside their home because of maltreatment was even more disproportionate: eight out of every 1,000 black

children residing in the county were placed outside their home that year, compared with just 1.7 of every 1,000 white children.

Studies by Brett Drake, a professor in the Brown School of Social Work at Washington University in St. Louis, have attributed the disproportionate number of black families investigated by child-welfare agencies across the United States not to bias, but to their higher rates of poverty. Similarly, a 2013 study by Putnam-Hornstein and others found that black children in California were more than twice as likely as white children there to be the subject of maltreatment allegations and placed in foster care. But after adjusting for socioeconomic factors, she showed that poor black children were actually less likely than their poor white counterparts to be the subject of an abuse allegation or to end up in foster care.

Poverty, all close observers of child welfare agree, is the one nearly universal attribute of families caught up in the system. As I rode around with caseworkers on their visits and sat in on family-court hearings, I saw at least as many white parents as black — but they were all poor, living in the county’s roughest neighborhoods. Poorer people are more likely not only to be involved in the criminal-justice system but also to be on public assistance and to get their mental-health or addiction treatment at publicly funded clinics — all sources of the data vacuumed up by Vaithianathan’s and Putnam-Hornstein’s predictive-analytics algorithm.

Marc Cherna, who as director of Allegheny County’s Department of Human Services has overseen C.Y.F. since 1996, longer than just about any such official in the country, concedes that bias is probably unavoidable in his work. He had an independent ethics review conducted of the predictive-analytics program before it began. It concluded not only that implementing the program was ethical, but also that not using it might be unethical. “It is hard to conceive of an ethical argument against use of the most accurate predictive instrument,” the report stated. By adding objective risk measures into the screening process, the screening tool is seen by many officials in Allegheny County as a way to *limit* the effects of bias.

“We know there are racially biased decisions made,” says Walter Smith Jr., a deputy director of C.Y.F., who is black. “There are all kinds of biases. If I’m a screener and I grew up in an alcoholic family, I might weigh a parent using alcohol

more heavily. If I had a parent who was violent, I might care more about that. What predictive analytics provides is an opportunity to more uniformly and evenly look at all those variables.”

For two months following Emily Lankes’s visit to the home of the children who had witnessed an overdose death, she tried repeatedly to get back in touch with the mother to complete her investigation — calling, texting, making unannounced visits to the home. All her attempts went without success. She also called the treatment center six times in hopes of confirming the mother’s sobriety, without reaching anyone.

Finally, on the morning of Feb. 2, Lankes called a seventh time. The mother, she learned, had failed her three latest drug tests, with traces of both cocaine and opiates found in her urine. Lankes and her supervisor, Liz Reiter, then sat down with Reiter’s boss and a team of other supervisors and caseworkers.

“It is never an easy decision to remove kids from home, even when we know it is in their best interest,” Reiter told me. But, she said, “When we see that someone is using multiple substances, we need to assure the children’s safety. If we can’t get into the home, that makes us worry that things aren’t as they should be. It’s a red flag.” The team decided to request an Emergency Custody Authorization from a family-court judge. By late afternoon, with authorization in hand, they headed over to the family’s home, where a police officer met them.

The oldest child answered their knock. The mother wasn’t home, but all three children were, along with the mother’s elderly grandfather. Lankes called the mother, who answered for the first time in two months and began yelling about what she considered an unwarranted intrusion into her home. But she gave Lankes the names of family members who could take the children for the time being. Clothing was gathered, bags packed and winter jackets put on. Then it was time for the children to get in the car with Lankes, a virtual stranger empowered by the government to take them from their mother’s care.

At a hearing the next day, the presiding official ordered the mother to get clean before she could have her children returned. The drug-treatment center she had been attending advised her to enter rehab, but she refused. “We can’t get in touch

with her very often,” Reiter recently told me. “It’s pretty clear she’s not in a good place. The two youngest kids are actually with their dads now. Both of them are doing really, really well.” Their older brother, age 13, is living with his great-grandfather.

In December, 16 months after the Allegheny Family Screening Tool was first used, Cherna’s team shared preliminary data with me on how the predictive-analytics program was affecting screening decisions. So far, they had found that black and white families were being treated more consistently, based on their risk scores, than they were before the program’s introduction. And the percentage of low-risk cases being recommended for investigation had dropped — from nearly half, in the years before the program began, to around one-third. That meant caseworkers were spending less time investigating well-functioning families, who in turn were not being hassled by an intrusive government agency. At the same time, high-risk calls were being screened in more often. Not by much — just a few percentage points. But in the world of child welfare, that represented progress.

To be certain that those results would stand up to scrutiny, Cherna brought in a Stanford University health-policy researcher named Jeremy Goldhaber-Fiebert to independently assess the program. “My preliminary analysis to date is showing that the tool appears to be having the effects it’s intended to have,” Goldhaber-Fiebert says. In particular, he told me, the kids who were screened in were more likely to be found in need of services, “so they appear to be screening in the kids who are at real risk.”

Having demonstrated in its first year of operation that more high-risk cases are now being flagged for investigation, Allegheny’s Family Screening Tool is drawing interest from child-protection agencies around the country. Douglas County, Colo., midway between Denver and Colorado Springs, is working with Vaithianathan and Putnam-Hornstein to implement a predictive-analytics program there, while the California Department of Social Services has commissioned them to conduct a preliminary analysis for the entire state.

“Given the early results from Pittsburgh, predictive analytics looks like one of the most exciting innovations in child protection in the last 20 years,” says Drake,

the Washington University researcher. As an author of a recent study showing that one in three United States children is the subject of a child-welfare investigation by age 18, he believes agencies must do everything possible to sharpen their focus.

Even in Illinois, where B.J. Walker, the director of the state's Department of Children and Family Services, is terminating its contract with the companies that developed Rapid Safety Feedback, predictive analytics is not dead. "I still believe it's a good tool to make better informed decisions," Walker told me in December. Walker knows Cherna and Dalton and saw the long process they went through to develop the Family Screening Tool. "They're doing a careful job," she said. "Their transparency has been laudable. And transparency isn't often your friend, because you're going to make some mistakes, you're going to stumble, you're going to make changes."

Cherna and Dalton are already overseeing a retooling of Allegheny County's algorithm. So far, they have raised the program's accuracy at predicting bad outcomes to more than 90 percent from around 78 percent. Moreover, the call screeners and their supervisors will now be given less discretion to override the tool's recommendations — to screen in the lowest-risk cases and screen out the highest-risk cases, based on their professional judgment. "It's hard to change the mind-set of the screeners," Dalton told me. "It's a very strong, dug-in culture. They want to focus on the immediate allegation, not the child's future risk a year or two down the line. They call it clinical decision-making. I call it someone's opinion. Getting them to trust that a score on a computer screen is telling them something real is a process."

Dan Hurley is a science journalist and longtime contributor to the magazine. He is at work on a book about his experiences as a foster father and scientific efforts to prevent and treat child abuse.

Sign up for our newsletter to get the best of The New York Times Magazine delivered to your inbox every week.

A version of this article appears in print on January 7, 2018, on Page MM31 of the Sunday Magazine with the headline: Predicting the Worst.

