디지털 행정을 통한 아동학대 대응체계 혁신방안 연구

2024년 7월

보건복지부

배태현

차례

<u>국외훈련 개요</u>

1. 훈련국: 미국

2. 훈련기관명: 시카고대학교(University of Chicago)

3. 훈련분야: 정책학(Master in Arts in Public Policy Studies)

4. 훈련기간: 2023.8.14~ 2024.7.31.

훈련기관 개요

1. 해리스 공공정책대학원 개요

과학적인 분석을 통해 공공 정책 변화에 기여하기 위해 1988년 설립된 대학원으로 US.news에서 발표하는 미국 대학 순위에서 행정학 분야 8위 대학원으로 발표된 바 있다. . 해리스 공공정책대학원은 정부, 비영리기관, 국제기구 등에서 공공정책을 수행하는 전문가 양성을 목표로 한다. 정책 분석에 있어 다학제적, 데이터 기반으로 접근함에 따라 시카고 지역 및 기타 도시 지역에 대한 학생과 교수진의 기여를 높이고 있다.

2. 주소

- 일리노이주 시카고 이스트60번가 1307(1307 East 60th Street Chicago, IL 60637)

3. 연혁

- 1988년 설립
- 1995년 Master of Public Policy(MPP) 석사 과정 프로그램 개설
- 2000년 Master of Science in Environmental Science and Policy(MSESP) 석사 과정 프로그램 개설
- 2014년 Master of Science in Computational Analysis and Public Policy(MSCAPP) 석사과정 프로그램 개설

4. **규모**

- 2024년 2월 기준 총 1,477명이 재학 중이며 그 중 1,220명이 대학원생, 257명이 학부생이다. 매년 약 550명의 신규 입학생을 모집한다. 2023년 신입생 기준 미국 학생의 비율은 39%, 유학생 비율은 61%이다.
- 교수진은 총 78명으로 교수진 대 학생 비율은 16:1이다.

5. 주요 석사과정 프로그램

- 공공정책학 석사 (Master of Public Policy, MPP): 가장 많은 학생이 재학중인 2 년제 프로그램으로, 정책 분석, 경제학, 정치학, 통계학 등의 분야에서 심도 있는 교육을 받는다. 졸업생들은 정부 기관, 비영리 단체, 국제기구, 민간 부문 등 다양한 분야로 진출한다.
- 컴퓨터 분석 및 공공정책학 석사 (Master of Science in Computational Analysis and Public Policy, MSCAPP): 시카고 대학교 컴퓨터공학과와 공동 운영하는 2 년제 석사과정이다. 데이터 분석, 프로그래밍 등의 기술을 공공 정책 문제에 적용하는 방법을 중심으로 교육한다.
- 공공정책학 석사 (Master of Arts in Public Policy, MA): MPP 와 유사한 커리큘럼이나 1 년 과정으로 이루어져 있다. 1, 2 학기에는 필수과목을 수강해야 하며 3 학기에는 본인의 관심 분야에 따라 선택과목을 수강할 수 있다.

- 연구 방법론 인증 공공정책학 석사 (MA in Public Policy with Certificate in Research Methods): 15 개월 과정으로, 공공 정책과 함께 심화된 연구 방법론 교육을 받는 프로그램이다.

6. 주요 과목

- 모든 석사과정생들은 정책분석(Analytical politics) 1&2, 공공정책과 미시경제(Principles of Microeconomics in Public Policy) 1&2, 데이터 분석과 통계학(Statistics for Data Analysis) 1&2 등의 필수과목을 수강해야 한다. 모든 과목에서 R스튜디오를 이용한 데이터 분석 과제가 포함되는 등 실증 데이터 기반의 공공정책을 강조한다.
- 이 외에도 개발 경제, 국제 개발 협력, 의료 정책, 여성 정책, 아동가족정책, 환경정책, 도시개발 등 다양한 분야의 과목을 운영 중으로 학생의 관심 분야에 따라 전문 분야를 개발할 수 있다.
- Policy Lab 과목을 운영 중이며, 이는 시카고 지역의 공공기관, 공공 컨설팅 회사, NGO 등이 의뢰하는 문제에 대해 학생들이 교수 지도 하에 문제를 분석하고 정책 개선 방안을 제시하는 현장 연계형 과목이다.

I. Research Background

In recent years, Korea has made significant strides in strengthening its national responsibility towards child abuse and overhauling its response system. One of the pivotal changes includes the deployment of child abuse investigators in every city and district. These officials are responsible for directly conducting investigations into suspected cases of child abuse, a role previously managed by non-government child protection agencies. This shift ensures that public officials handle the initial response and investigation phases, thereby enhancing the efficiency and immediacy of interventions. The new measures also allow for the immediate separation of children from potentially abusive environments when there is suspicion of abuse, reflecting a proactive stance in early intervention.

Despite these advancements, South Korea continues to struggle with a lower detection rate of abused children compared to other developed countries. Additionally, the proportion of children experiencing repeat abuse has been increasing. These issues are primarily attributed to the inadequate systems for the prevention and early detection of child abuse.

Meanwhile, the Yoon Suk-yeol administration in South Korea is accelerating its transition to a 'Digital Platform Government,' where citizens, businesses, and the government collaborate on a unified digital platform to address social issues and create new value by connecting all data.[1] Along with this trend of digital

[1] Presidential Committee on the Digital Platform Government, https://www.dpg.go.kr/DPG/contents/DPG01400000.do

governance, the system for child abuse response should innovate itself into a more proactive, prevention-focused policy based on digitalization.

Many advanced nations have effectively utilized digital administration to enhance every stage of child abuse response. For instance, in the United States, many States and Counties have developed preventive risk modeling systems for early detection of children at risk. Given the covert nature of child abuse within families and the challenges faced by children in expressing their circumstances, data-driven approaches are expected to improve the ability to identify and manage risks early significantly.

However, we can also see the opposition against using data in child welfare services regarding ethical issues, data privacy, accuracy, and racial disparities, which made several States stop using data-based risk modeling tools. For the Korean government to innovate its child welfare system without serious social outcry, we should learn lessons from the cases abroad.

As the social expectation for administration increases and data science develops, it becomes a pressing issue to integrate digital administration into child abuse response systems. The Korean government should take steps forward to a digitalized child abuse response system to ensure no child falls through the cracks. This research aims to analyze the recent cases of data-based child abuse response abroad and give implications to the Korean policy on child abuse response.

II. The Overview of the U.S. Child Abuse Response System[2]

1. How the US Child Abuse System works

(1) The scope of CPS intervention

In the US, each State has laws defining abuse and neglect and requiring local child protective services(CPS) agencies to intervene. In general, CPS does not intervene in cases where acquaintances or strangers harm children; these are treated as criminal cases by law enforcement. In some states, certain types of abuse(sexual or physical) are not dealt with by CPS, and those are responded to by law enforcement officers.

(2) Reporting and screening abuse or neglect

Certain mandatory reporters are designated by State laws, but all people can report child abuse if they are concerned. After CPS workers receive the report, they decide whether to screen it in or screen it out. A report is screened in when there is sufficient information to suggest an investigation is warranted. If a report is screened out, CPS workers can refer the reporter to other services or law enforcement.

(3) Investigation and substantiation of a case

After a report is screened in, CPS caseworkers start an investigation – meeting parents and other adults who have contact with the child and also

[2] Child Welfare Information Gateway, "How the child welfare system works."(U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau, 2020).

speaking with the child if possible. During the investigation, if the child is in immediate danger or has continued maltreatment, the child can be removed from the home and placed in foster care, shelter, or a relative's home. As a result of the investigation, case workers decide whether the abuse is "substantiated(found)" or "unsubstantiated(unfound)". CPS workers substantiate a case if an incident of child abuse or neglect is believed to have occurred.

CPS initiates a court action if there is a need for child protection or dependency proceedings. The court can order the temporary removal of the child from home or prohibit a certain individual from having contact with the child during the investigation. After an adjudicatory hearing, the court decides whether to continue the jurisdiction of the child.

(4) CPS services for substantiated cases

After a case is substantiated, CPS provides service to the child and family, considering many factors, such as the severity of the case, the child's immediate safety, perceived future risk of abuse, and available service.

- Little or no-risk cases could be closed with no services or referred to community resources, which is not provided by the child welfare agency. The CPS assesses a case as having little or no risk if it is a one-time incident with a low risk of repetition and the child is now safe.
- If the case has low to moderate risk, CPS workers can also refer those families to community-based services for parenting training, child care, mental health counseling, or services for other needs(job, housing, etc.) to enhance the family's function.

- In moderate to high-risk cases, CPS workers often provide voluntary-in-home services. If the family refuses the offer, the juvenile court can weigh in and make the requirement for the family to cooperate with services. In-home services focus on reducing safety risks. In the case of high risks of serious harm to the child, or if the child is already threatened, the court orders the removal of the child. The child would be mostly placed in foster care or relatives' home or sometimes at group residences. During the separate placement period, the child receives medical treatment and other services, and the family also gets support to mitigate risks. Within 12 months, and every other 12 months, the agency makes a permanency plan for the child, which is mostly the reunification but sometimes adoption or transfer of custody. The court holds a permanency hearing to see if the plan secures the child's safety and decides the reunification or other methods.

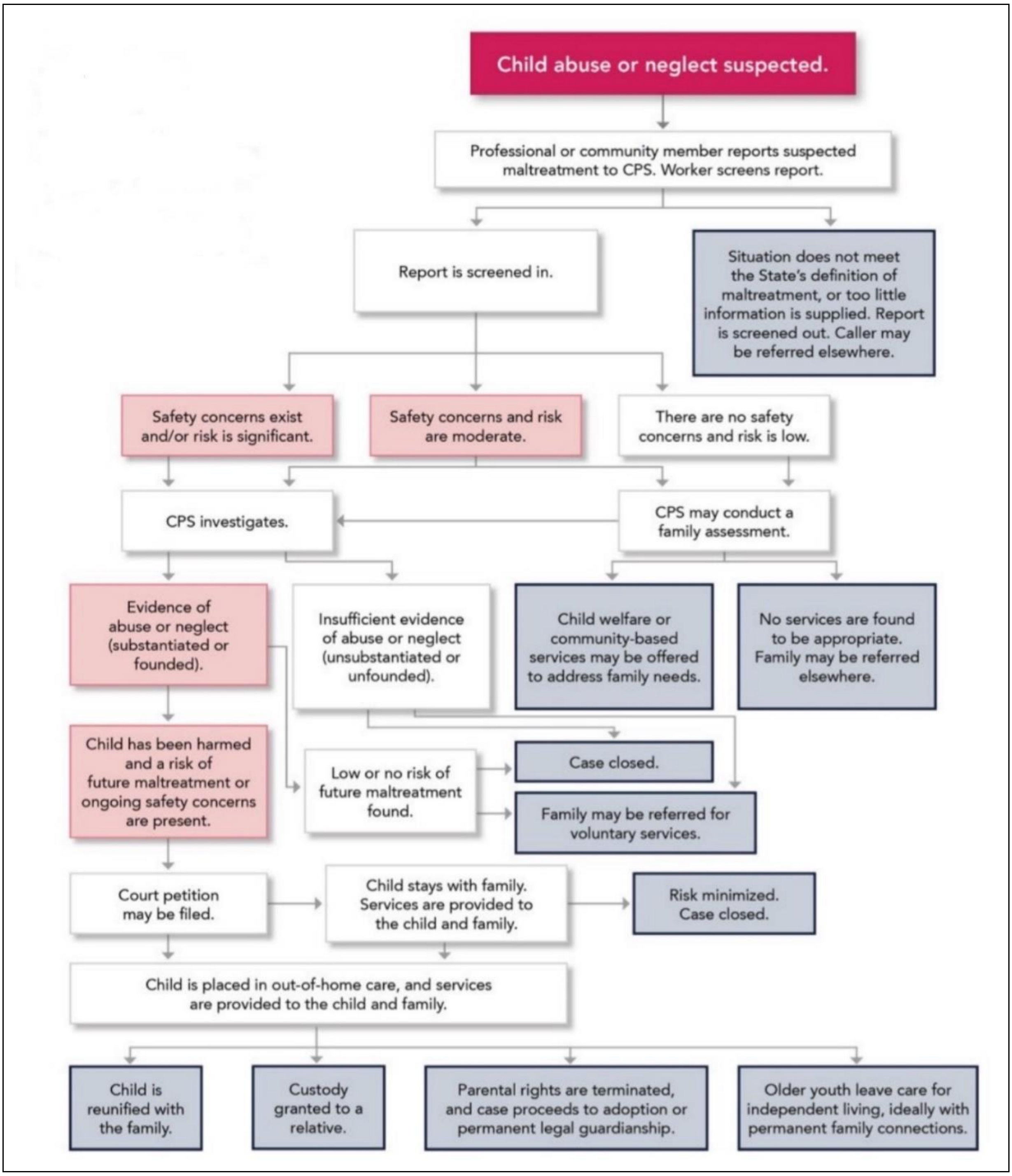
(5) Alternative Responses

Suppose the CPS decides on a low or moderate risk of abuse at the early phase. In that case, some states offer alternative approaches for low-risk cases, which do not formally determine the occurrence of maltreatment or perpetrator. Instead of investigating the evidence of abuse, CPS assesses the family's strengths and difficulties and focuses on identifying necessary support for the family. This is also called 'family assessment response' or 'differential response.'

States have varying standards and systems for alternative approaches. Some States that initiated alternative approaches do not report the cases to the NCANDS, the national database for abuse. There are also some States where the alternative approach is implemented only in particular Counties as a pilot project. In the US, Alabama, Arkansas, Georgia, Iowa, Kentucky, Maryland, Missouri, Nebraska, Nevada, North Dakota, Ohio, Texas, Washington, and Wisconsin have fully or partially implemented alternative response pathways.[3]

[3] U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau, *Child Maltreatment 2022* (2024), 17, https://www.acf.hhs.gov/cb/data-research/child-maltreatment.

Figure 1: CPS workflow[4]



[4] Child Welfare Information Gateway, "How the child welfare system works."(U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau(2020), 8.

2. Key Statistics of Child Protective Services in 2022[5]

(1) Referral, Screening, and Victim number

In 2022, Child Protective Services (CPS) agencies across the United States received approximately 4,276,000 referrals, involving about 7,530,000 children. The national rate of referrals that were screened in and accepted for investigation stood at 29.0 per 1,000 children. Among the 47 states reporting both screened-in and screened-out referrals, 49.5% were screened-in, while 50.5% were screened-out. The total number of victims amounted to 558,899. The overall victimization rate was 7.7 victims per 1,000 children.

(2) Types of Maltreatment

The types of maltreatment reported in 2022 revealed that neglect was the most prevalent, affecting 74.3% of victims. Physical abuse was the next most common, accounting for 17.0% of cases, followed by sexual abuse at 10.6%, and psychological maltreatment at 6.8%.

(3) Foster Care

Of the children reported as victims, 104,747 (19.6%) received foster care services on or after the day of the abuse report. Additionally, 40,702 non-victims (1.4%) also received foster care services within the same timeframe. The type of foster care includes family foster homes, such as those of

[5] U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau, *Child Maltreatment 2022* (2024), https://www.acf.hhs.gov/cb/data-research/child-maltreatment

relatives or others(majority), or other group settings, as well as residential facilities, such as group homes and emergency shelters.

(4) Child Fatalities

Tragically, an estimated 1,990 children died due to abuse and neglect in 2022, corresponding to a fatality rate of 2.73 per 100,000 children. Infants younger than 1 year old represented a significant portion of these fatalities, accounting for 44.7% of the deaths.

(5) Perpetrators

The number of perpetrators in 2022 was 434,090 in total. The majority of perpetrators (76.0%) were parents of the victims. Among non-parent perpetrators, relatives constituted 5.8%, and unmarried partners of parents made up 3.7%.

(6) CPS Workforce and Caseload

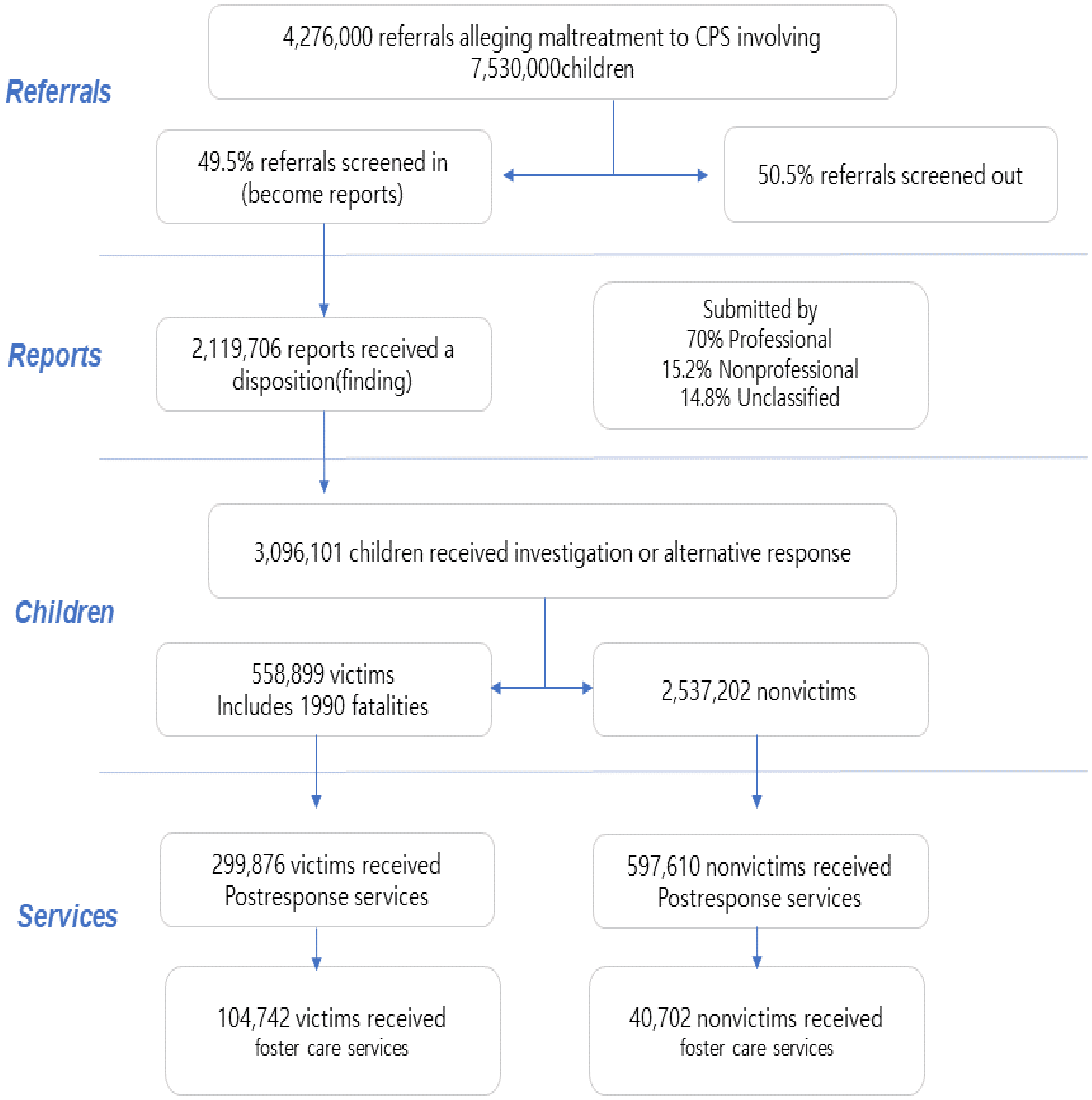
In 2022, 45 states reported a total CPS workforce of 30,750 workers. Additionally, 41 states reported having 5,036 specialized intake and screening workers. CPS investigation and alternative response workers completed an average of 69 responses per worker for the year, highlighting their responsibilities' significant workload and critical nature.

(7) Victims by race and ethnicity

In 2022, White children accounted for 227,593 victims(6.6‰), Hispanic children numbered 130,048 victims (7.0‰), and Black or African American children comprised 118,850 victims (12.1‰). Children identified as having two or more races represented 31,837 victims (9.4‰), while American Indian or Alaska Native children had the highest rate with 8,043 victims (14.3‰). Asian children had 5,283 victims (1.3‰), and Native Hawaiian or Other Pacific

Islander children had 1,459 victims (9.3‰). There is a difference in victimization rates among different racial and ethnic groups, with American Indian or Alaska Native and Black or African American children experiencing the highest rates of victimization.

Figure 2: Child Abuse in the US by Workflow(2022)[6]



[6] Ibid.

III. Recent Focus of Prevention as a Background of Data-based Approach

Current US child welfare policy is shifting towards a prevention-focused approach. After a stream of criticism against placing excessive children in out-of-home settings, mostly by child rights advocates, the US federal government officially recognized the policy priority to be the prevention of child abuse. In 2018, The Family First Prevention Services Act(FFPSA) was signed into law as a part of Public Law. The core of FFPSA is promoting child abuse prevention services in advance and avoiding unnecessary placing in congregate care. FFPSA authorized IV-E funding for States' services in evidence-based mental health programs, substance abuse prevention and treatment, in-home parent skill-based programs, and kinship navigator programs.[7]

What's notable about FFPSA is it articulated the eligibility for IV-E funding as evidence-based practice. To make judgments about which program is evidence-based, the Administration for Children and Families established a website, "Title IV-E Prevention Services Clearinghouse." Clearinghouse reviews multiple research studies on related programs and services that are currently running in a transparent and objective manner. On this website, programs and services implemented by local governments are rated into 4 categories – well-supported, supported, promising, and does not currently meet criteria. For a program to be eligible for Title IV-E reimbursement, it

---

[7] Administration for Children&Families, Children's Bureau, Title IV-E program, https://www.acf.hhs.gov/cb/title-iv-e-prevention-program

should be rated as at least promising. Also, over 50% of a State's prevention services should be rated as well-supported.[8]

Figure 3: Title IV-E Clearinghouse Visualization[9]

Program or Service Name and Date	Program or Service Rating	Program or Service Impacts
30 Days to Family® Date Research Evidence Last Reviewed: Mar 2024	Well-supported	**Subdomains with favorable impacts** Child permanency: Least restrictive placement **All Impacts:** Favorable: 17 No Effects: 1 Unfavorable: 0
A Second Chance, Inc. Kinship Navigator Date Research Evidence Last Reviewed: Mar 2024	Does not currently meet criteria	**Subdomains with favorable impacts** None **All Impacts:** Favorable: 0 No Effects: 0 Unfavorable: 0
Active Parenting of Teens: Families in Action™ Date Research Evidence Last Reviewed: Dec 2022	Does not currently meet criteria	**Subdomains with favorable impacts** None **All Impacts:** Favorable: 0 No Effects: 0 Unfavorable: 0

Since the implementation of FFPSA, States have actively pursued prevention-first programs with sufficient empirical evidence. For example, Illinois DCFS has initiated the Nurturing Parenting Program(NPP), Positive Parenting Program(Triple P), Child Parent Psychotherapy(CPP), Trauma-Focused Cognitive-Behavioral Therapy(TF-CBT), and Multi-Systemic Therapy(MST), which meet the eligibility requirements of Title IV-E.[10]

Scholars see the shift to prevention and emphasis on empirical data, combined with the rapid advancement in data science, as paving the way for the use of preventive risk assessment in child welfare services.[11]

---

[8] Title IV-E Prevention Services CLEARINGHOUSE, https://preventionservices.acf.hhs.gov/

[9] Title IV-E Prevention Services CLEARINGHOUSE, https://preventionservices.acf.hhs.gov/program

[10] Illinois Department of Children and Family Services, *Family First Newsletter*, Issue 3 (September 2023), 2.

[11] Paul Lanier et al., "Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical

IV. Predictive Use of Data in Child Protection (1) – Birth Match

1. Overview of Birth Match

The birth match system is a tool designed to protect newborns at high risk of abuse or neglect by cross-referencing birth records with child welfare and criminal data. The birth match system represents a proactive approach to child protection, aiming to prevent harm by identifying high-risk situations early.

The first adoption of birth match traces back to 2001 in Michigan after cases of brutal child death cases by parents who already had severe maltreatment records. After Michigan's adoption, several more States followed suit: Minnesota, Maryland, Texas, and Missouri. In most States other than Missouri, the implementation of the birth match system emerged in response to tragic cases where infants were harmed or killed by parents with a history of severe abuse or terminated parental rights. In 2016, the Commission to Eliminate Child Abuse and Neglect Fatalities(CECANF) recommended the expansion of birth match policies to States and Counties, which led to the adoption of birth match in Missouri.

The five states generally use similar criteria for filtering parents for birth match. They use child welfare data and criminal data regarding the termination of parental rights, severe child abuse records, or court records, including the death of children. Once a person on the list gives birth to a baby, a system matches the birth record with the list and gives an alert to local child welfare divisions.

Principles to Evidence-Based Child Welfare Policy," *Journal of Family Violence* (2020), 2.

Though the manual for response to birth match alerts varies across States, it is generally treated as similar to abuse referrals, which results in abuse investigations. Among the 5 States, Maryland is considered to have limited application of birth match, as it conducts a less invasive assessment rather than a full investigation, whereas Missouri, which implements a more immediate Newborn Crisis Assessment, is assessed to have the proactive version of birth match.

Table 1: Summary of States with birth match

State	Rooted in Law	Criteria for Birth Match	Response	Matching Period	Year
Michigan	X	Termination of parental rights, child death due to abuse/neglect, serious abuse/neglect	Regular investigation	No specific limit	2001
Minnesota	O	Egregious child abuse, involuntary termination of parental rights	Regular investigation	No specific limit	2006
Maryland	O	Termination of parental rights, child murder/attempted murder	Assessment by local DSS	Initially 5 years, extended to 10	2007, revised 2019
Texas	X	Fatal child abuse, termination of parental rights	Regular investigation	2 years	2013
Missouri	O	Termination of parental rights, certain crimes against children	Newborn Crisis Assessment	10 years	2021

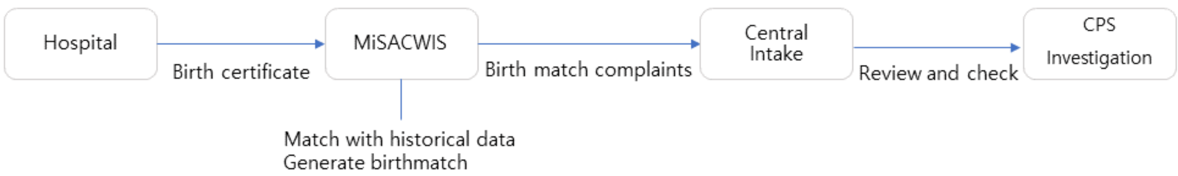
2. Birth Match in States

(1) Michigan

Michigan was the first State to adopt a birth match system as early as 2001, after the tragic death of an infant by parents whose parental right to another child was terminated. In 2001, an interagency agreement made it possible to match birth data and the termination of parental rights. When a new child is born to highly risky parents, the Michigan Department of Health and Human Services(MDHHS) automatically receives a notification. The criteria for notification are parental rights termination in a child protective proceeding, abuse or neglect-related death of a child, or if the parents are manually added to a match list due to severe child abuse or neglect cases.

When a child is born, hospitals send the birth certificate to Michigan's child welfare information system(MiSACWIS). If the parents meet the criteria above, the MiSACWIS automatically makes a birth match complaint and sends an email alert to the state's hotline call center, Central Intake. On any birth match alert, the CI must verify the accuracy. If there is an ongoing or pending case in the family, the alert must lead to the start of an official child abuse investigation. In such cases, investigations are based on the allegation of "threatened harm to the child", which is one category of child abuse in Michigan law.

Figure 4: The operational flow birth match in Michigan



In Michigan, the number of birth match complaints is known to be declining, as is the percentage of open cases. In the 2010s, the number was around 1000 a year: 1213 complaints in 2014 and 1186 in 2018. It dropped to 873 in 2020 and 515 in 2021, which no one could fully explain. At the same time, the percentage of open cases from birth matches declined from 9 percent in 2012 to 3 percent in 2020.[12] While the reason is unclear, Michigan's Department of Health and Human Services announced they will make a front-end redesign of the child protection system, including the birth match system.

(2) Minnesota

Minnesota is praised as having the most protective birth match system. It was implemented in 2006 after a related law amendment in 2001 and computer system development for several years. In Michigan, the criteria for the birth match is 'threatened harm', which includes parents who had previously done egregious harm to children or an act that led to involuntary termination of parental rights or transfer of custody.

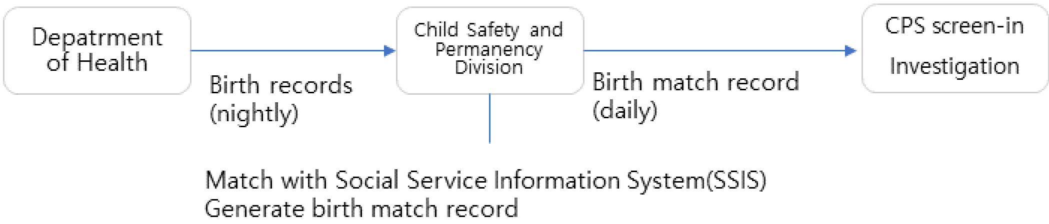
Minnesota Department of Health cooperates with the Child Safety and Permanency Division in the process of birth match. The Department of Health uploads birth records every night, which are sent to the Child Safety and Permanency Division for the matching process. The child protection records from the Social Service Information System(SSIS) are matched with birth records daily through an application. There is no time limit to go back

---

[12] Marie Cohen, "Learning from the Past: Using Child Welfare Data to Protect Infants Through Birth Match Policies," *American Enterprise Institute* (May 2022), 7.

regarding the child welfare records. According to the Minnesota government, annual birth match records go up and down by around 248 to 454.[13]

Figure 5: The operational flow birth match in Minnesota



(3) Maryland

In Maryland, the voice calling for the birth match has been present since 2004, when multiple numbers of children died from parents who had previous abuse records. Despite a child welfare reform committee's recommendation of birth match, it took long for the State to implement the system. In 2006 and 2007, a bill for the birth match was introduced, but it has been pending in the committee due to opposition regarding finance and parental rights issues.

In 2007, after the death of a 2-year-old child, the bill was passed in Congress. However, the birth match in Maryland was limited in the time frame and the criteria for choosing parents.

According to the law, the Maryland Social Service Administration(SSA) was required to provide to the health department an updated list of potentially dangerous parents whose parental right was terminated. Then, the health

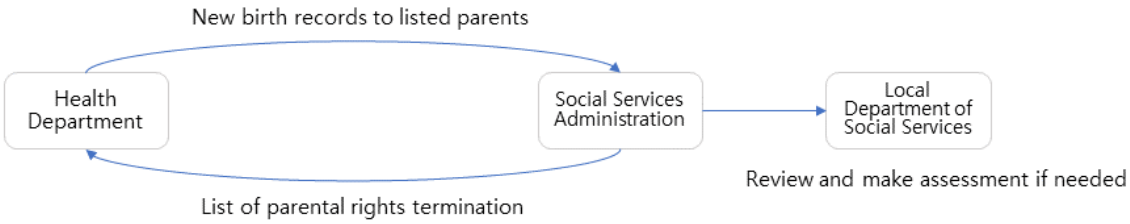
[13] Cohen, "Learning from the Past: Using Child Welfare Data to Protect Infants Through Birth Match Policies", 8.

department needs to give information back to SSA if the parents gave birth to new babies. There is a time limit of 5 years regarding the termination of parental rights. Also, the local Department of Social Services(DSS) is not always mandated to screen in or investigate the cases. Only if the agency reviews the data and has concerns will it assess the family. The assessment is a weaker step than a formal investigation, as it does not necessarily include interviewing people outside the family. Social workers are only required to contact the family within seven days and assess the risk, safety, strengths, and needs of the family, which leads to referring the family to appropriate social services.

After the Baltimore Child Fatality Review Team recommended an expansion of birth match, Maryland passed a bill to revise the birth match system up to the current standard. It extended the time period of termination of parental rights to 10 years. Also notable is it started including parents who committed murder or attempted murder of a child, which was a loophole in the past as they were technically not available to terminate their right to the children. It is also mandated that an investigation be started if the parents refuse the requested birth matching process. This expansion of the birth match system faced opposition as it was punitive to parents by adding a second punishment, and also violated data privacy.

Despite the opposition, the bill was passed in the legislature unanimously in both the Senate and the House, with a compromise of assessing the effectiveness of birth match by an independent entity. In order to conduct the impact assessment, DHS was required to build up related birth match data. According to the DHS data, the number of children involved in birth match almost doubled in 2019, from around 100 to over 200. This was due to the extension of the time period from 5 to 10 years. However, the number dropped back to 124 in 2020 for unknown reasons.[14]

Figure 6: The operational flow birth match in Maryland



(4) Texas

Texas implemented its birth match system in 2013 through an interagency agreement between the Texas Department of Family and Protective Services(DFPS) and the Texas Department of State Health Services(DSHS). The State followed a recommendation from the Texas State Child Fatality Review Team(SCFRT) and came to the implementation of the system.

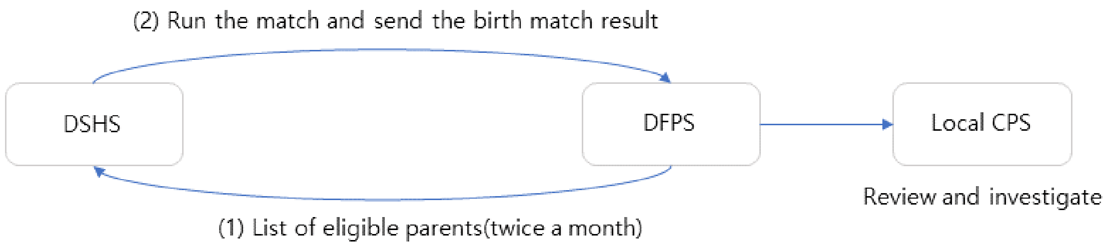
In Texas, the criteria for the birth match are if the parents committed child abuse or maltreatment, which resulted in a fatality, or had parental rights termination in the past two years. The DFPS is required to inform the list of eligible adults two times a month. Then, DSHS runs a matching between newborn data and the parent list and sends back the match result to DFPS. If the result is correct, the CPS conducts regular child abuse investigations on the family.

Compared to other States where the birth match was implemented, the Texas system is very limited in terms of looking back period. In 2018, SCFRT

[14] Cohen, 11.

recommended extending the period to 5 years, but it was rejected by DFPS. The reason was constraints in the capacity of the front-line agency and community service providers as well as the quality concern of the overall data matching process. In Texas, an annual number of birth match results stays around 800 to 1100, with 6 percent of matched families receiving in-home services and 3.5% of infants entering foster care.[15]

Figure 7: The operational flow birth match in Texas



(5) Missouri

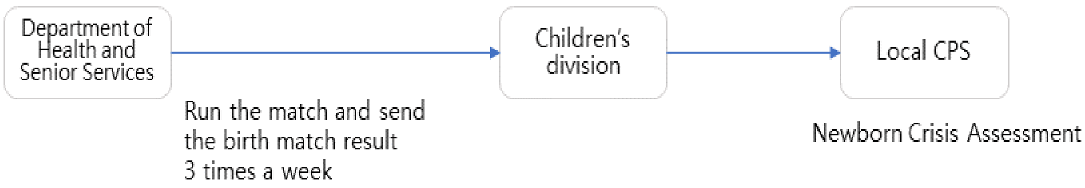
Missouri became the latest State to implement a birth match system by amending a related bill in 2021. The system came into effect in August 2021. Missouri's introduction of the birth match system followed the recommendation from CECANF, which called for the nation's adoption of birth matches.

In Missouri, the eligibility for birth match is in two categories: if parental rights have been terminated due to child abuse or neglect in 10 years or committed a certain crime against children in 10 years. The match system works by the Missouri Department of Health and Senior Services, sending birth match results three times a week to the Children’s Division. Children’s Division is mandated to inform the match result to the local County office to initiate contact with the family. Once receiving the information, Counties

[15] Cohen, 13.

should run a Newborn Crisis Assessment(NCA), which is more like an emergency response. For example, they should contact the family in 3 hours, and if the child is thought to be unsafe, the case worker should request immediate custody from the court.[16]

Figure 8: The operational flow birth match in Missouri



3. Evaluations about Birth Match

Birth Match has existed for more than twenty years, but only five States in the US have applied the policy. Some people are eager to make Birth Match a nationwide system, while others are cautious.

Cohen is one of the proactive advocates of birth match, calling for its expansion in additional states. She argues that, if adequately implemented, the birth match is a crucial tool for preventing child maltreatment. She explicitly claims that " birth match should be added to every state's arsenal of prevention programs" and that "Congress should consider mandating it on a federal level."

However, many people, including Cohen, agree that the current system, which has not improved since its first implementation in the early 2000s, needs to be improved in terms of accuracy and validity.

---

[16] Cohen, 14.

(1) Limitation of input data

First of all, people are concerned that a lot of potentially harmful adults are not captured from the current birth match system, as it only filters biological parents. It means the system fails to inform dangerous historical data about fathers who are not listed on the birth certificates or other adults the biological mom is now living with. Cohen says this loophole might potentially lead us to miss a significant number of high-risk individuals, and the scope of the birth match should be expanded for all related adults who committed severe child maltreatment.[17]

(2) Concerns about violation of privacy and parental right

Some critics also argue that the system can infringe on civil liberties by targeting individuals based on historical data. This concern includes potential violations of privacy and parental rights, as well as fears of discrimination. In some states, such as Maryland, this opposing view caused a lag between the bill's introduction and passage. In Texas, the extension of the birth match period was declined partly due to the criticism.

(3) Lack of empirical assessment of the impact

It is also worrisome that even States that have adopted the birth match system did not proactively collect the birth match data for further analysis of its effectiveness. It is more notable considering a declining number of birth match results in some States, which do not have an apparent reason. Cohen recommends, "States should collect the data needed to assess implementation and outcomes."

[17] Cohen, 19.

(4) Transparency of the system issue

Lanier pointed out the lack of transparency in birth match systems. He emphasized the importance of rationale behind input data selection; in other words, agencies should be able to explain why certain data are selected as criteria and why others are not. From the application process of birth match, we couldn't see any States clearly explaining why they picked "termination of parental rights" or why 2 years or 10 years. [18]

(5) The possibility of "black swan problem"

Lanier also raises concern about the basic feasibility of birth match systems as they attempt to predict scarce and unusual situations.[19] Some scholars argue that random or unexpected events are very difficult to predict, which is called the black swan phenomenon. Therefore, he cautiously points out the impossibility of predicting severe infant maltreatment cases from the beginning.

---

[18] Lanier et al., "Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence-Based Child Welfare Policy", 7.

[19] Lanier et al., 9.

V. Predictive Use of Data in Child Protection (2) – PRM

1. The rise of PRM in child welfare policy

(1) The risk assessment tool in child abuse before PRM

In the early days of child abuse response in the US, caseworkers had no other option but to rely on unaided worker judgment when assessing the risk of a reported child and making decisions on proper interventions. To enhance risk assessment accuracy and consistency, agencies developed a "consensus-based" assessment method. The consensus-based method is made up of factors that are thought to be predictive of maltreatment but were not proven in reality.

About 20 years ago, scientists developed a new risk assessment method called "actuarial" instruments. They distinguish themselves from consensus-based instruments by including empirically supported factors to predict maltreatment. Since their introduction, actuarial methods have outperformed traditional unaided worker judgment or consensus-based methods.[20]

The most prominent actuarial method is structured decision-making (SDM), which was first developed by the National Center of Crime and

[20] Brett Drake et al., "A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare," *Annals of the American Academy of Political and Social Science* 692, no. 1 (November 2020): 163-165.

Delinquency(NCCD) in 1998 but is still widely used across the States. This instrument assists case workers in determining the child's risk at many points of the abuse response process, such as screening, investigation, case planning, reunification, etc.

(2) Limitation of actuarial risk assessment

Although the use of SDM and other actuarial tools enhanced the accuracy and consistency of child risk assessment compared to before, it still had operational limitations. First of all, the fidelity of the tool mattered as it often had operator errors. Actuarial tools often have predictive factors that are subjective to the operator, such as whether adequate supervision is provided at home.[21] It led workers to make inconsistent ratings for identical cases. Another problem of actuarial tools is, as it has subjective features, workers are required to have a huge amount of education and training when they are already heavily burdened with caseloads.

There are several statistical limits of actuarial tools. Such tools were also rarely validated with the population or subpopulation of local governments. Moreover, they depended on static models, which could not keep up with the ever-changing population.[22]

(3) The rise of PRM

Predictive risk models(PRMs) have recently been introduced in child welfare policy. It uses "historical data to understand the relationship between myriad

---

[21] Stephanie Cuccaro-Alamin et al., "Risk Assessment and Decision Making in Child Protective Services: Predictive Risk Modeling in Context," *Children and Youth Service Review* 79 (2017): 293.

[22] Ibid.

factors to estimate a probability score for the behavior or outcome of interest."[23] PRM makes specific algorithms based on historical data, and how algorithms are created may differ from traditional regression statistics or machine learning processes. The scope of data for PRM also varies based on the local governments. Some jurisdictions' PRM is restricted to using child welfare agency's data. In contrast, other jurisdictions permit the use of a broader range of data from different agencies – education, health, birth, social benefits, criminal records, etc.

The development of PRM tools starts with defining the outcome variable, which represents the ground truth in concern. Child abuse agencies might want to know the future likelihood of child abuse occurrence. However, we are not able to capture whether the child abuse happened or not. All the historical data tells are CPS decisions: if the child was reported to be abused and it is substantiated, and whether or not the child was removed from the home. Therefore, to make the model, the ground truth is translated into one of the measurable and concrete proxies, which becomes the outcome variable.[24]

After defining the outcome variable, researchers select input variables that might predict child abuse and which modeling method might be the most powerful and explainable. Sometimes, the CPS agency requests to include or exclude some factors as input variables. Regarding the input variable, there is an alternative modeling method called natural language processing topic modeling (NLP/TP), which uses a text mining method. In NLP/TP models, the

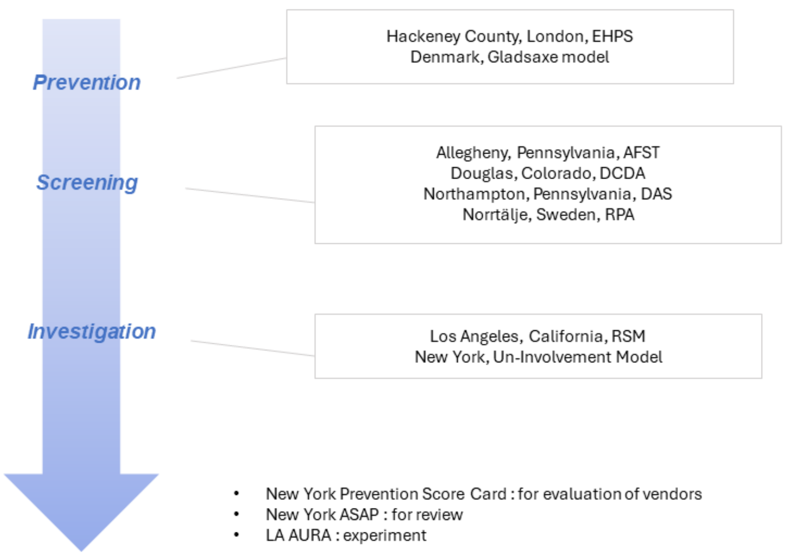
[23] Paul Lanier et al., "Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence-Based Child Welfare Policy," 2.

[24] Stephanie K. Glaberson, "Coding Over the Cracks: Predictive Analytics and Child Protection," *Fordham Urban Law Journal* 46 (2019): 328-330.

variables are defined bottom up, rather than researchers deciding which variable to put in.[25]

As the following chapter shows, PRM tools in various local governments are developed based on different modeling methods and multiple purposes. As the name indicates, the application of PRMs is mainly concentrated in the early phases of child abuse response policy, such as prevention in advance or detection of risky cases in screening.

Figure 9: The application of PRM concentrating in early stages



[25] Katarina Lappalainen, "Protecting Children from Maltreatment with the Help of Artificial Intelligence: A Promise or a Threat to Children's Rights?" in *Law, AI and Digitalization* (2022), 437-9.

2. States and counties with PRM algorithms

Table 2: Local government in the US with PRM

County, State	Title	Year of Introduction	Purpose	Output variable	Developer
Allegheny, Pennsylvania	Allegheny Family Screening Tool(AFST)	2016	Assist call screeners' decision	removal in 2 year	Vaithianathan & Putnam-Hornstein
Douglas, Colorado	Douglas County Decision Aid(DCDA)	2019.2	Assist call screeners' decision	removal in 2 year	Vaithianathan & Putnam-Hornstein
Northampton, Pennsylvania	Decision Aid Tool(DAS)	2021.3.	Assist call screeners' decision	removal in 2 year	Vaithianathan & Putnam-Hornstein
Illinois	Rapid Safety Feedback(RSF)	Implemented in 2015, dropped in 2017	Unknown	Unknown	Eckerd and MindShare Tech
Los Angeles County	Approach to Understanding Risk Assessment (AURA)	Decided not to implement in 2017	Unknown	Unknown	SAS
Los Angeles County	Risk Stratification Model(RSM)	2021	Support targeted approach	removal in 2 year	Vaithianathan & Putnam-Hornstein

Oregon	Safety at Screening Tool	Implemented in 2018, Stopped 2022	Assist call screeners' decision	removal in 2 year	Vaithianathan & Putnam-Hornstein
New York	Accelerated Safety Analysis Protocol(ASAP)	2018	Select cases for Quality Assurance Review	physical or sex abuse in 24 months	Unknown
New York	Prevention Score Card	2021	Support evaluating performances of prevention service vendors	future investiga-tion in 24 months	Unknown
New York	Un-Involvement Model	2023	select cases for early closure sooner than the typical 60 days	no future involve-ment with ACS within 24 months	Unknown

Table 3: Local governments in Europe with PRM

Region, Country	Title	Year	Implementation	Output Variable	Developer
Hackney County, London, UK	Early Help Profiling System(EHPS)[26]	Drpped in 2019	Predict risky children in advance and give early intervention	Unknown	Xantura
Denmark	The Gladsaxe model[27]	Planned but dropped in 2018	Early warning for detecting vulnerable children	Unknown	Unknown
Norrtälje, Sweden	Robotic Process Automation system(RPA)	Developed in 2020, dropped in 2021	Support the decision of initiation	Unknown	NLP/TP model Text mining
Strängnäs, Sweden	-	Developed in 2021	informs the social workers of the risk implications	Text analysis of reports	NLP/TP model Text mining

(1) Allegheny County: Allegheny Family Screening Tool(AFST)

- Overview of the AFST

Allegheny County, Pennsylvania, is one of the leading counties in the States that first developed and implemented a data-based child maltreatment prediction system called Allegheny Family Screening Tool(AFST). The Allegheny County Department of Human Services(DHS), teamed with researchers from Auckland University of Technology and Children's Data Network at the University of Southern California, created the AFST, launched in August 2016. AFST is a program based on data analysis of tens of thousands of child abuse cases in Allegheny to score the risk of the case resulting in out-of-home placement in two years. Whenever a call screener gets a new referral, the tool gives a score for each case ranging from 0 to 20, which is categorized into high-risk(score higher than 17), low-risk(score lower than 11), and others.

- The implementation of AFST

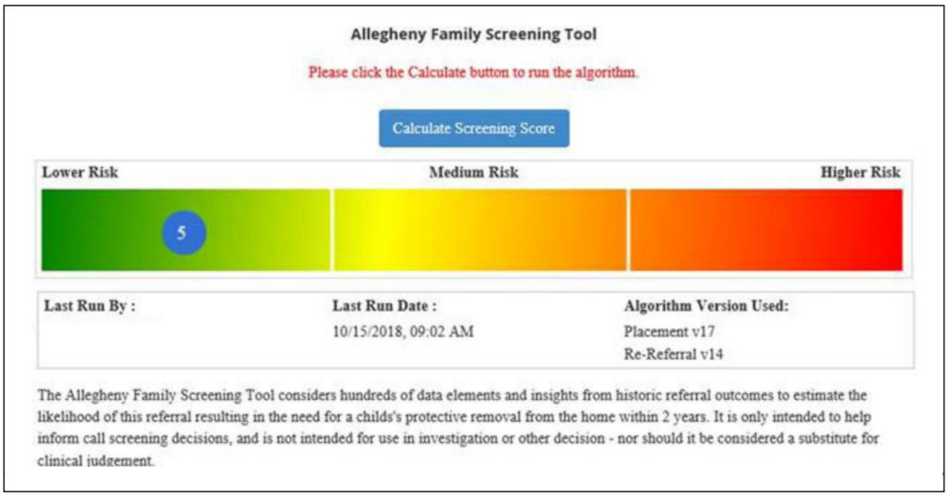
Essentially, the development of AFST was for the assist of call screeners, who get the phone call of alleged child maltreatment and decide whether to screen in or out. Call screeners, before the launch of AFST, already had access to a range of family data from over 20 years of databases, including child protection services, mental health records, and homeless services. However, with the magnitude of data and surging number of phone calls, it was nearly impossible for screeners to make meaningful use of the data. In addition, the case opening was inconsistent, relying heavily on the call screener's subjectivity. This concern led DHS to develop a data-based screening tool to assist screeners in evaluating the risk of each referral. To ensure the tool does not automatically decide the risk score but is used as an advising tool for screeners, the AFST-based risk score is visualized at the end of the whole process, and screeners evaluate the individual significance of the case.

If the AFST gives a high-risk alert and the victim child(or other children in the household) is age 16 years or younger, the protocol requires call screeners to put that case into mandatory screen-in and open a formal investigation process unless the screener overrides it. If the score falls in the low-risk scope and all children in the household are at least 12 years old, the screener is recommended to screen out the case. In any other cases, the call screeners have full discretion.

According to Rittenhouse et al.(2023), the implementation of AFST has significantly changed the rate of screening in and out based on the severity of the cases. The rate of screening-in among low-risk referrals dropped, and that among high-risks increased, aligning with the intended goal of AFST. Furthermore, the racial disparity in the screening-in ratio has declined. Before the kick-in of AFST, black children were more likely to be screened in than white children. After the usage of the tool, the racial gap in high-risk children being screened has been closed by 83%.[28]

[28] Rittenhouse et al., "Algorithms, Humans and Racial Disparities in Child Protection Systems: Evidence from the Allegheny Family Screening Tool", (2023).

Figure 10: AFST Visualization provided to call screening staff[29]



- Update of the system(AFST ver.2) and its methodology

Two years after the first introduction of AFST ver. 1(2016.4 ~ 2018.11), the Allegheny DHS has made an update to AFST ver. 2. The main changes from the original version were about target outcomes, predictors, and modeling methodology.

First of all, the AFST ver.1 used both the placement model and the re-referral model for the prediction of the risk. The placement model analyzed whether the child would be likely to experience severe maltreatment and be removed to an out-of-home setting in 2 years. The re-referral model was supposed to examine whether the one who is screened out will be referred to again as an alleged victim in 2 years. Allegheny County refused to use the re-referral model in ver.2, as it only has a limited tie to the outcome they are primarily concerned with. In many cases, children under custody disputes or

[29] Vaithianathan et al., “Allegheny Family Screening Tool: Methodology, Version2”, (Center for Social data Aanlytis, 2019).

other situations experienced repetitive calls, which often interrupted the solid analysis of the riskiness.

Secondly, the data used for the predictors has been modified. Public benefit records used in ver.1, such as receiving TANF(Temporary Aid to Needy Families) or SNAP(Supplemental Nutrition Assistance Program), dropped in ver.2. Meanwhile, birth record and the allegation data were added to the data set. Records about child welfare, jail, juvenile probation, and behavioral health continue to be used in both versions. In the AFST ver.2, a total of 82,211 referrals(45,801 observations where children were screened-in) from April 2020 to July 2014 were used as a sample. Each child-referral observation had 451 variables associated with the child, the alleged perpetrator, and the household. The sample data was partitioned into test and training sets and trained to predict the probability of placement.

Lastly, the data analyzing model has changed from simple logistic regression to add LASSO model. I will explain this further in the following chapter.

- The accuracy evaluation of AFST

The AFST model used AUC(Area Under the Receiver Operator Curve) to measure prediction accuracy. It refers to the probability of a true positive case(a high-risk case was actually assigned to a high-risk score) having a higher risk score than a true negative case. It is considered to be accurate if the AUC is over 50%. In the AFST ver.2, researchers put effort into enhancing the AUC, trying four different types of modeling methodologies(LASSO, XG-BOOST, Random Forest, and SVM). Among the models, all showed higher AUC than logistic regression in general; the researchers decided to adopt the

LASSO model in consideration of the accuracy in high-risk groups and the equivalent level of accuracy among races. The overall accuracy of the LASSO model was 75.97% AUC, while that of black children was slightly lower(74.42%) than white children(77.35%).[30]

The accuracy of the prediction was further shored up by external validation research. The child maltreatment data was linked to the hospital data from UPMC Children's Hospital of Pittsburgh, where the majority of children in Allegheny use. The external validation research looked into whether the risk score had a relation to the injury encounter. It revealed that there is a positive correlation between AFST risk scores and medical encounters(injury, suicide, abusive injury). When comparing high-risk to non-high-risk children, the odds ratio of injury(1.73), abusive injury(1.46), and suicide(1.71) all showed statistically significant ($p<0.1$) numbers, suggesting that the high-risk score from AFST might result in great medical danger in real.

(2) Douglas County: Douglas County Decision Aid(DCDA)

- Background

Douglas County, before the introduction of DCDA, made the screening decision based on the consensus of the human staff. In the majority of cases, they used the RED(Read, Evaluate, Direct) Team process, where one child protection supervisor and two or more caseworkers met to deliberate on each referral. The problem was, even though the RED team had access to TRAILS data(The state's child welfare system) and Colorado court records, it was

---

[30] Vaithianathan et al., "Allegheny Family Screening Tool: Methodology, Version2".

difficult to make a consensus based on historical data in a timely manner, and the decision to screen in or out didn't really sync with the risk of the case. From January 2016 to April 2017, 41% of cases that the Red Team screened out were re-referred in less than a year.[31]

- Data used and methodology

Like the Allegheny County AFST, invented by the same team of researchers, Douglas County Decision Aid(DCDA) measured the likelihood of the child being removed within 2 years, displayed as a score of 1 to 20. However, the type and range of data used were different from AFST. Douglas County used state-level child welfare and case management systems (TRAILS), allowing a wider data range. However, unlike AFST, where the human service and health data was accessible, DCDA only depended on child welfare data(TRAILS and SACWIS: Statewide Automated Child Welfare Information System) and welfare program data(CBMS: Client Benefits Management System). The research team collected data from January 2015 to September 2016 with 221,519 records to generate 501 predictors. Same with the AFST, the DCDA model adopted LASSO regularized logistic regression due to its efficiency and interpretability.

- Implementation

Douglas County deliberated on the detailed implementation of the tool to ensure its right use. For example, a big concern was whether to display the risk score of DCDA upfront to the screeners or at the end of the screening

[31] Vaithianathan et al., "Implementing a Child Welfare Decision Aide in Douglas County", (Center for Social Data Analytics, 2019).

process. In a 4-week randomized controlled trial of the DCDA model, researchers tried both ways and compared their pros and cons. Screeners were supposed to use the upfront display in the first two weeks and the final display in the last two weeks. As a result of the trial, they ended up using the final display, which gave the screeners room for critical thinking. Additionally, the use of DCDA was strictly limited to screeners, not allowing case workers who investigated the case to know the risk score. This also shows the County's commitment to using the tool only as an aid device, which makes a minimal change to the existing human-based system.

(3) Los Angeles County: Approach to Understanding Risk Assessment(AURA) and Risk Stratification Model(RSM)

Los Angeles County has a decade-long history of discussion around predictive risk modeling. This local government has fought with arguably the largest number of child abuse investigations, dealing with 220,000 referrals in the year of 2014 alone. It goes without saying that Los Angeles deeply felt the need to streamline the process with the application of data mining. In the early 2010s, LA launched a project called "AURA(Approach to Understanding Risk Assessment" with the prominent software firm SAS. However, the first-ever PRM tool for detecting children at risk failed to be implemented due to the lack of accuracy and public concern. However, the county did not drop the hope for efficient use of data, which finally led it to develop a new "Risk Stratification Model" put into effect in 2021.

- The development of AURA (2014)

In an attempt to predict the risk of alleged child abuse cases and make better decisions, Los Angeles County made a contract with SAS and ran an experiment of risk modeling. SAS looked into critical incidents in 2011 and 2012 in LA, which caused death or near fatalities. Data used in the experiment are known to include previous child abuse reports, mental health records, alcohol and substance records, etc. Researchers tracked down the related data of the reported child and the family and developed a 1,000 scale AURA score. The AURA model was then applied to 2013 abuse referral cases to test whether AURA accurately predicted those children who suffered severe threats.

- The decision to shelve the AURA (2017)

The AURA test with 2013 records showed partly promising but not very satisfactory results. Researchers from SAS were confident that the application of AURA could have prevented 171 severe abuse cases in 2014. According to the AURA experiment, nearly 4,000 children had higher than 900 AURA scores, which meant they were in need of special treatment. Among those flagged as high risks, 171 children proved to have experienced violent situations in 6 months, and the number was 76% of death or severe injury from abuse.[32] However, LA officials were worried about AURA's high false positive rate. Of the predicted 4,000 high-risk children, only 171 had real-life risks. The remaining 3,829 cases, if AURA were to be implemented, would put more burden on the social workers.

At the same time, grassroots organizations in LA expressed grave concerns

---

[32] Daniel Hempel, "Uncharted Waters: Data Analytics and Child Protection in Los Angeles," *The Imprint* (July 20, 2015), https://imprintnews.org/featured/uncharted-waters-data-analytics-and-child-protection-in-los-angeles/10867

about the application of AURA. Los Angeles County Community Child Welfare Coalition stated that “Predictive analytics could be used to create maps and information used to marginalize certain populations further or justify disproportionality in the Child Welfare system, based on race and bias.”. The transparency of the algorithm also came under fire. As it was developed by a software firm SAS, the way the algorithm works was confidential to anyone outside SAS. It is an explicit concern of the LA County officials. In 2017, the LA County Board of Supervisors announced a report examining AURA's strengths and weaknesses called the “Nash Report.” According to Nash Report, “because the model is proprietary, there is a lack of transparency about how its algorithms are constructed and various factors weighted(thus earning its classification as a “black box” model. This concerns users and evaluators alike, as no way exists to understand how these elements influence the decision-making process, and if systemic biases are inherent in the tool.” These concerns added up to put AURA on the shelf.

- Risk Stratification Tool(2021)

A few years after the drop of the AURA experiment, LA County came up with a more prepared risk assessment algorithm, “Risk Stratification Tool,” in 2021. A 2019 State Audit, which criticized Los Angeles County DCFS for frequent inaccuracies in safety and risk assessments[33], again made the department dust off the risk-assessing algorithm plan. The RSM model was developed by Putnam-Hornstein and Vaithianathan, the team who invented the Allegheny Family Screening Tool. Similar to other PRM models, the LA Risk Stratification Tool analyzed previous child maltreatment records and related data to assess

[33] Auditor of the State of California, *California State Auditor Report 2018-126*(May 2019), 17-21.

the risk of new child abuse referrals and support the social workers make better decisions around the investigation and treatment of the child.

The risk stratification model is supposed to filter 10% of cases that do not have immediate safety concerns now but might have the possibility of out-of-home placement in 24 months. The system designated those as “enhanced support” cases and supported the officials in providing extra care to those children. The LA county and researchers provide detailed information about which data they collected and how the tool was created. They used the state-wide child welfare service data, “Child Welfare Services Case Management System(CWS/CMS), as a source. Researchers analyzed 278,465 screened-in child abuse records from 2016-2017 in LA County and split those samples into a training set(75%) and a test set(25%). 292 features were coded as indicator variables, which include 130 features of referred child, 80 features of referral information, 48 features of other children in the household, and 34 features of parents and other adults. Races or geographical features were not included.

Table 4: Coded Features in LA Risk Stratification Model[34]

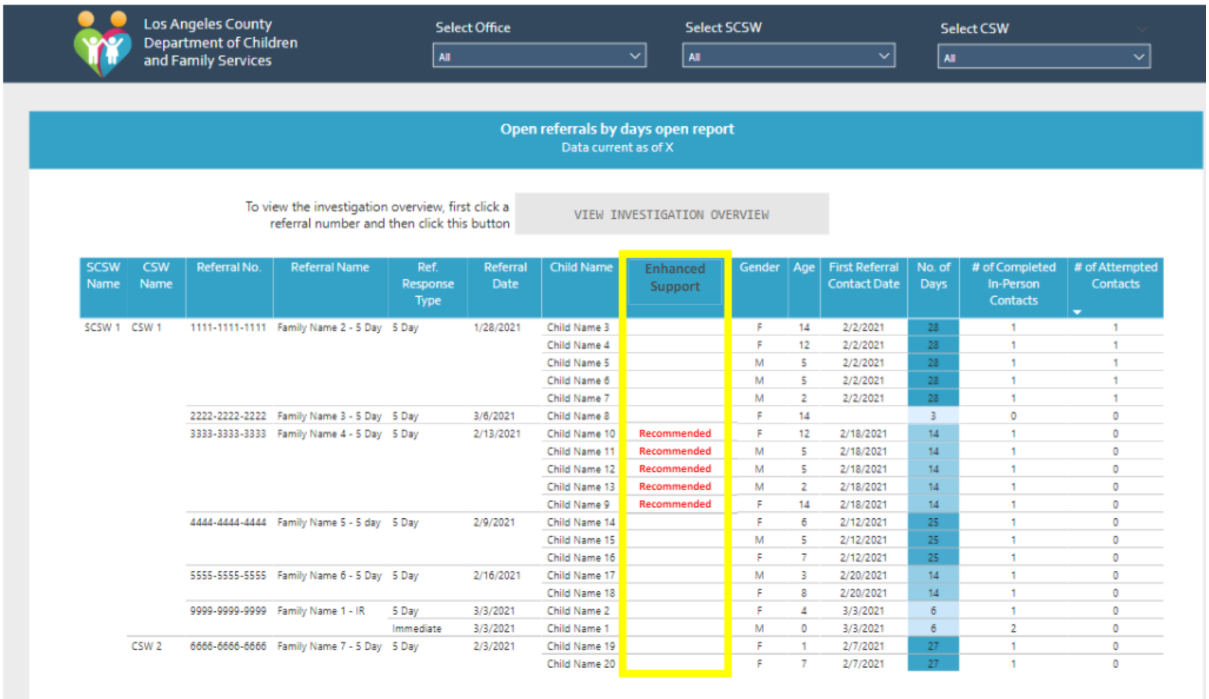
Child with allegations	130 features	· Demographic information(e.g., age, gender) · Current maltreatment allegation information(e.g., total # of allegations for child, physical abuse allegation, general neglect allegation) · Current involvement with DCFS(e.g., already open referral, already open case) · Helath information(e.g., indicator of prenatal substance exposure, indicator of developmental service or mental health needs) · Abuse/safety data(e.g., injury harm details recorded, abuse frequency recorded) · Maltreatment history(e.g., age at first ever referral, total # of prior referrals, total # of prior substantiated allegations of sexual abuse) · DCFS service and placement history(e.g., # of prior cases, age at first placement, time since last placement, prior placement with kin)
Referral information	80 features	· Reporter type(e.g., law enforcement, family member) · Day and time of report(e.g., Friday, referral received at 3am, summer holiday flag, winter vacation flag) · Number of children with allegations on report(e.g., one child, five children)

34 Putnam-Hornstein E, Vaithianathan R, and McCroskey J, *The Los Angeles County Risk Stratification Pilot : An Overview and One Year Update*(Children's Data Network, August 29, 2022), 21.

		· Number of adults associated with allegations on report(e.g., one adult, three adults) · Referral-level information(e.g., address indicates family is homeless, non-protecting parent code, infant child named in this referral)
Other children	48 features	· Demographic information for all children(e.g., age, gender) · Current maltreatment allegation information(e.g., any child with a sexual abuse allegation) · Current involvement with DCFS(e.g., any child currently in foster care) · Health information(e.g., any child w/ prenatal substance exposure, any child w/ indication of psychotropic medications) · DCFS history(e.g., any child with a termination of parental rights, any child with history of substantiated caretaker incapacity)
Parents and other adults	34 features	· Maternal information(e.g., age, adult associated with substantiated allegations) · Paternal information(e.g., relationship to child with allegations, previous terminations of parental rights) · Other adult information(e.g., adult associated with inconclusive allegations, childhood history of foster care placements)

The project result paper emphasized the supportive feature of the tool. It underlines that “the model is not being implemented as a standalone analytics tool, but it is to elevate core practices and support for investigations where the stakes are high.” The Risk Stratification Tool is used to pursue a targeted approach, where children with higher risks can use better resources promptly. This tool is utilized when caseworkers and supervisors manage already open investigations, helping them to provide enhanced support in some instances. The model runs every night using the latest reported screen-in information in CWS/CMS. The timing of the Risk Stratification Tool is unique as it gives information at the outset of the investigations. It is different from many other PRM models, such as the Allegheny Family Screening Tool, which supports the decision of screening in at the very beginning of the process.

Figure 11: LA Risk Stratification Model Visualization 1(pilot program)



Los Angeles County
Department of Children
and Family Services

Select Office: All
Select SCSW: All
Select CSW: All

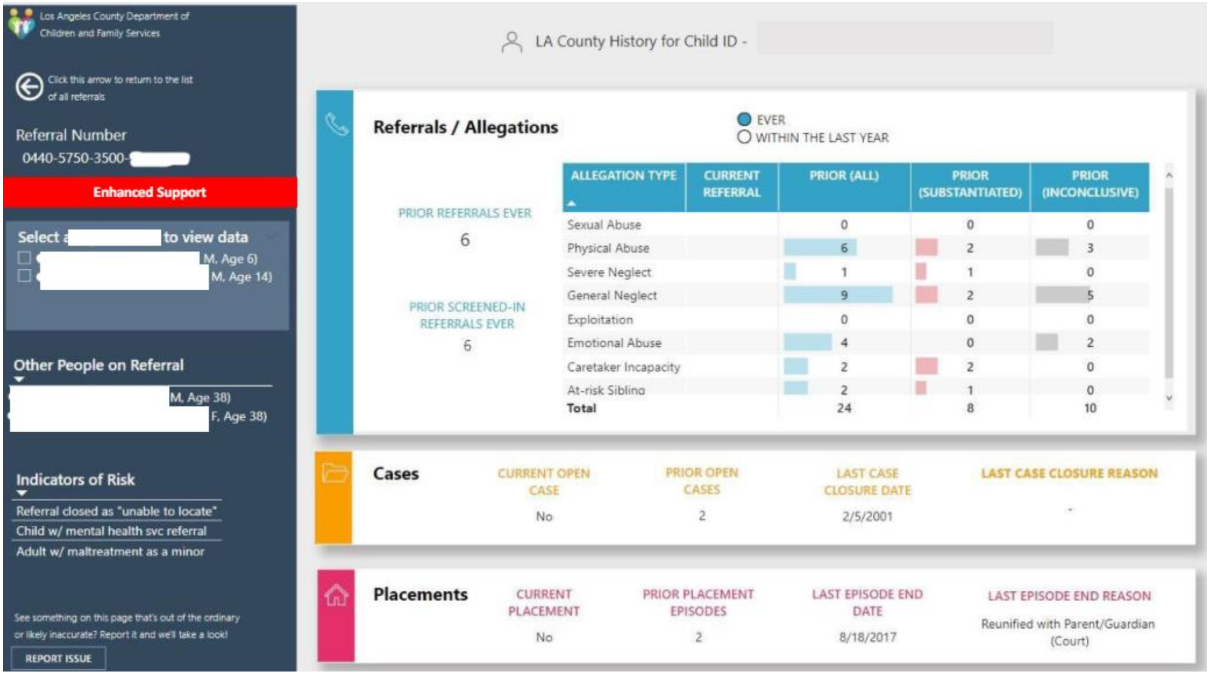
Open referrals by days open report
Data current as of X

To view the investigation overview, first click a referral number and then click this button

VIEW INVESTIGATION OVERVIEW

SCSW Name	CSW Name	Referral No.	Referral Name	Ref. Response Type	Referral Date	Child Name	Enhanced Support	Gender	Age	First Referral Contact Date	No. of Days	# of Completed In-Person Contacts	# of Attempted Contacts
SCSW 1	CSW 1	1111-1111-1111	Family Name 2 - 5 Day	5 Day	1/28/2021	Child Name 3		F	14	2/2/2021	28	1	1
						Child Name 4		F	12	2/2/2021	28	1	1
						Child Name 5		M	5	2/2/2021	28	1	1
						Child Name 6		M	5	2/2/2021	28	1	1
						Child Name 7		M	2	2/2/2021	28	1	1
		2222-2222-2222	Family Name 3 - 5 Day	5 Day	3/6/2021	Child Name 8		F	14		3	0	0
		3333-3333-3333	Family Name 4 - 5 Day	5 Day	2/13/2021	Child Name 10	Recommended	F	12	2/18/2021	14	1	0
						Child Name 11	Recommended	M	5	2/18/2021	14	1	0
						Child Name 12	Recommended	M	5	2/18/2021	14	1	0
						Child Name 13	Recommended	M	2	2/18/2021	14	1	0
						Child Name 9	Recommended	F	14	2/18/2021	14	1	0
		4444-4444-4444	Family Name 5 - 5 day	5 Day	2/9/2021	Child Name 14		F	6	2/12/2021	25	1	0
						Child Name 15		M	5	2/12/2021	25	1	0
						Child Name 16		F	7	2/12/2021	25	1	0
		5555-5555-5555	Family Name 6 - 5 Day	5 Day	2/16/2021	Child Name 17		M	3	2/20/2021	14	1	0
						Child Name 18		F	8	2/20/2021	14	1	0
		9999-9999-9999	Family Name 1 - IR	5 Day	3/3/2021	Child Name 2		F	4	3/3/2021	6	1	0
				Immediate	3/3/2021	Child Name 1		M	0	3/3/2021	6	2	0
	CSW 2	6666-6666-6666	Family Name 7 - 5 Day	5 Day	2/3/2021	Child Name 19		F	1	2/7/2021	27	1	0
						Child Name 20		F	7	2/7/2021	27	1	0

Figure 12: LA Risk Stratification Model Visualization 2(pilot program)



From the beginning of the project, external experts and community stakeholders were actively engaged in the development of the tool. For instance, one of the reasons that the county set "out-of-home placement in 24 months" as an outcome variable is the feedback from stakeholder groups. LAC community members asserted that decisions from somewhere else, such as courts, should be applied for more consistency and accountability.[35] The researchers included race and ethnicity at first in the input variable of the modeling but dropped those features based on the feedback from community partners.[36] Furthermore, after 18 months of research, the Risk Stratification Model was put into effect as a pilot program in three offices: Belvedere,

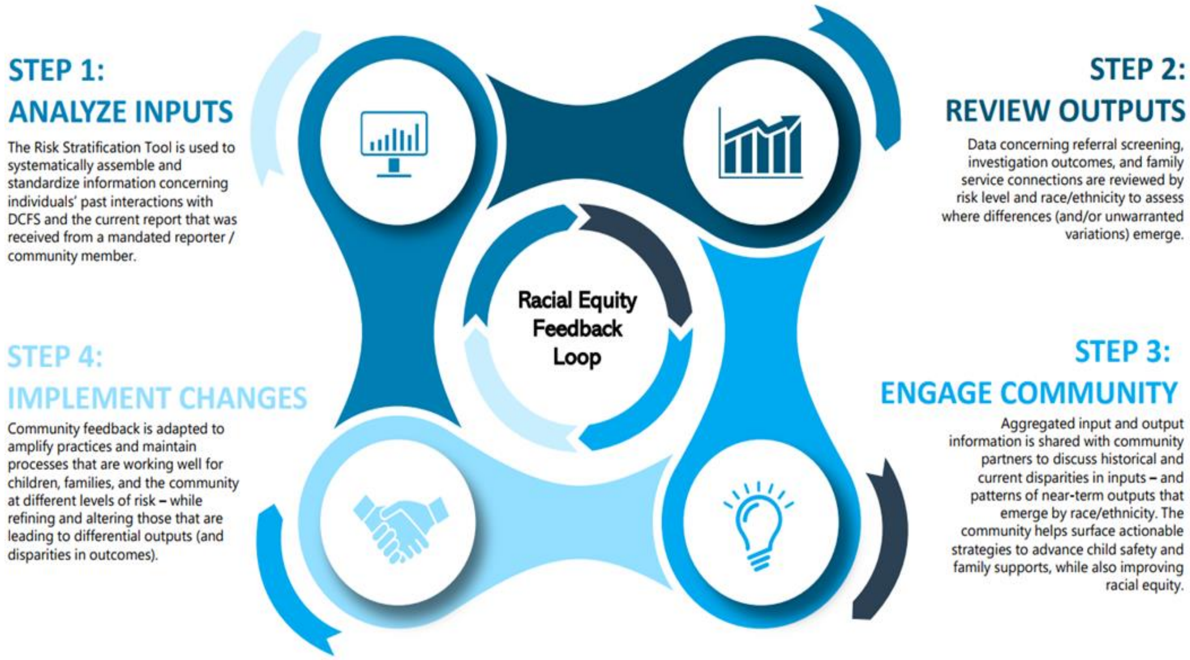
[35] Children's Data Network, *Los Angeles County Risk Stratification Model: Methodology & Implementation Report*(August 2021), 11.

[36] Children's Data Network, 12.

Lancaster, and Santa Fe Springs. All three offices launched the Risk Stratification Model in August 2021. Throughout the pilot project, an external expert group, Mathematica, worked together to collect data from three offices about how the supervisors used the data reports.

A notable feature of the LA Risk Stratification Model is its application of the “racial equity feedback loop.” The research team acknowledges that historical and current human bias can affect the referring or screening-in process. There have been disproportionately high rates of child abuse cases for Black children, which came under investigation. The model included a racial equity feedback loop to lay a foundation for reducing the subjectivity or bias towards Black households. Specifically, researchers focused on recent hotline screening-in decisions for Black children, which were seen as having a low likelihood of future involvement by the Risk Stratification Tool(lower 50% of risk), and no allegations of maltreatment were substantiated during the investigation. Hotline call reporters’ narrative text and structured data on those cases were qualitatively analyzed to figure out where the racial differences emerged. Finally, the community stakeholders, family representatives, and service providers gave feedback about ways to avoid racial bias and improve future system operations. LAC DCFS Office of Equity coordinated the racial equity feedback loop project.

Figure 13: Racial Equity Feedback Look[37]



(4) New York City: Accelerated Safety Analysis Protocol(ASAP) Tool, Prevention Score Card, Un-Involvement Model

- Accelerated Safety Analysis Protocol(ASAP) Tool

The Administration for Children's Services(ACS) in I developed a machine learning tool identifying the probability of severe harm for children involved in active investigations. The tool predicts the "likelihood of substantiated allegation of physical or sexual abuse within next 24 months. ASAP tool was developed in May 2018, using data from 132,026 closed investigation samples from April 2014 to April 2016. The input data only includes ACS administrative

---

[37] Putnam-Hornstein E, Vaithianathan R, and McCroskey J, 60.

data, such as investigations and foster care. It was also tested on 53,477 observations.[38]

This tool sounds similar to other PRMs as it predicts children at the highest risk of severe abuse, but the use is not for supporting screen-in or investigation. Rather, It is used to select cases that the ACS Quality Assurance unit in the Division of Child Protection would review. The Quality Assurance unit reviews the highest-risk ongoing cases to make sure those cases are treated with care. The unit checks whether the case has gone through relevant safety assessments and provided proper consultations and safety interventions to protect the children at high risk of physical and sexual abuse. However, the capacity of the Quality Assurance unit is limited to 3,000 cases annually for review. Therefore, the ASAP tool works to allocate the most needed cases to the quality assurance review.

According to the ACS, the ASAP model “dramatically out-performed previous approaches to identify cases for closer review, which were in fact found to be more likely to flag investigations about Black or Hispanic families for review than the predictive model.”[39]

---

[38] NYC Office of Technology & Innovation, “Summary of Agency Compliance Reporting of Algorithmic Tools” (2023): 1-2.

[39] Human Rights Watch, “If I wasn’t poor, I wouldn’t be unfit: The Family Separation Crisis in the US Child Welfare System” (2022), 159.

- Prevention Score Card[40]

In September 2021, New York ACS implemented another PRM tool called "Preventive Score Card." This tool measures the likelihood of future indicated investigation within the next 24 months for all children who are receiving prevention services from ACS. It was developed using the 158,787 observations from ACS historical data from July 2009 to June 2016. This model was tested on 84,494 observations from closed investigations from July 2009 to June 2018.

The purpose of this model is quite different from other PRMs. It is not used for allocating resources or supporting decisions. This tool makes it easier for ACS to evaluate the performance of prevention services from different vendors. As all children who receive prevention services have different levels of risk, it is difficult to compare the performance of service providers. ACS cannot simply give high rates for a vendor for not making repeated child abuse referrals, as there is a likelihood that the vendor served easier cases where the probability of repeated abuse was low from the beginning.

Therefore, ACS developed predictions of repeated maltreatment and made predictions on day ten after the prevention service started. Based on the average risk prediction of children of the year, service vendors are assigned one in four risk cohorts and get a scorecard: Very High-Risk Cohort(top 25%), High-Risk Cohort(next 25%), Medium-Risk Cohort(next 25%), Low-Risk Cohort(lowest 25%). The risk cohort makes it possible to assess the overall prevention services from different providers fairly.

---

[40] NYC Office of Technology & Innovation, "Summary of Agency Compliance Reporting of Algorithmic Tools", 3-4.

- Un-Involvement Model[41]

This is the most recent PRM tool developed by New York City ACS. The non-involvement model predicts the likelihood of no future involvement with ACS within 24 months. This tool was developed by training a model on 2012-2017 historical data containing 381,649 children from 183,516 investigations. This model was also based on only ACS data, such as time in foster care or SCR investigations.

The Un-Involvement tool is used to select cases for early closure sooner than the typical 60 days. After the ACS starts its investigation, all children take the first Un-Involvement score on the 10th day. After 30 days, they get the second Un-Involvement score at the 40th day with the updated data in 30 days. It is to ensure the safety of children. If additional data shows the child's risk is increased, the early closure decision can be revised.

The score itself does not dictate workers to automatically close the case early. It is notable that the score is not shared with caseworkers, who are at the very front when deciding the closure, so as not to create any unnecessary bias. Instead, the score is only seen by the deputy directors who manage the job of case workers. In 40 days, case workers consult with deputy directors about the decision to close cases early. In the consulting process, deputy directors can lead to the early closure of some cases based on the risk score to address the excessive workload of case workers.

[41] NYC Office of Technology & Innovation, 5.

(5) Hackney County, U.K.: Early Help Profiling System

In 2018, Hackney County in London, U.K., launched a pilot model of the Early Help Profiling System(EHPS) in pursuit of identifying vulnerable children in advance and providing proper support. EHPS is a PRM model based on administrative data to identify children who are at high risk of abuse or negligence. During the pilot implementation period, Hackney County social workers were provided with the list of vulnerable children monthly and reached out to the household with early help.

However, unlike other PRM systems like AFST or DCDA, EHPS predictive variables are not fully open to the public. This is mainly because EHPS was developed with the private firm Xantura, which was concerned that the public access to variables and modeling scheme could possibly undermine their commercial interest. Thus, I can only assume from media reports that EHPS data include socioeconomic factors such as school attendance, police records, domestic violence, social care, housing debt, substance abuse, etc. 42

According to the council, the system helped detect seven children in need who were otherwise out of the reach of social workers, and 80 percent of high-risk children identified from the system were at risk in the real world. However, the lack of transparency and accuracy backfired. There was a public outcry against using EHPS, with people voicing out frustration about the wrongful identification of EHPS and data privacy. Not only the targeted people were not informed of the use of their data, but the council refused to open the wrongful

---

[42] Lappalainen, "Protecting Children from Maltreatment with the Help of Artificial Intelligence: A Promise or a Threat to Children's Rights?" in *Law, AI and Digitalisation* (2022), 444-447.

identification rate or how it would handle it. It made the EHPS to be halted in 2019 after the pilot period as “expected benefits would not be realized due to the lack of accuracy and data.”

Regarding the drop of EHPS, Darren Martin, then vice chair of Hackney Liberal Democrats, commented, “The real issue here is the complete disregard that Hackney Council showed towards people’s privacy and their right to know how the local authority is using their data. … < UNK> We need now an assurance that any future trial of this nature needs to be put to public consultation with full disclosure of exactly what data is being collected and how it will be used.”[43] This remark strongly emphasizes how the government should handle data transparency to make the public onboard.

(6) Illinois: Eckerd’s Rapid Safety Feedback Program(ERSF)

Illinois purchased Eckerd’s Rapid Safety Feedback Program, which intended to predict the children at high risk. After children were reported through the Department of Children and Family Service(DCFS)’s hotline, ERSF analyzed the existing DCFS data to predict if the child would be likely to fall victim to severe injury or murder.

However, the severe inaccuracy of ERSF made it impossible to function as expected. The system had false positive and false negative issues at the same time. In 2015, over four thousand children were predicted to have 90%

[43] Ed Sheridan, “Town Hall drops pilot programme profiling families without their knowledge,” *Hackney Citizen*, October 30, 2019. https://www.hackneycitizen.co.uk/2019/10/30/town-hall-drops-pilot-programme-profiling-families-without-their-knowledge/

of death or injury, and 400 of the reported were classified as having a 100% probability of death, which was inconsistent with the State’s fatality rate.[44]

More surprisingly, the system failed to predict high-profile death cases related to maltreatment. The system was unable to detect the tragic death of 17-month Semaj Crosby and 22-month Itachi Boyle, which followed just a month later. Both cases were within the reach of DCFS, but the ERSF system and the following investigation filtered them as intact families.[45] Repeated death cases related to abuse significantly hampered the system’s accuracy and credibility. The ERSF system came under heated criticism, and Illinois decided to drop the use of the system in 2017.

The flaws of the system were unavoidable consequences from the very beginning. According to the Chicago Tribune, George Sheldon, then-director of DCFS, had a connection to the Eckerd and MindShare Technology and bypassed the bidding system so that they could hire Eckerd to set up the tool.[46] The DCFS was also accused of downplaying the importance of pre-tests before the system's state-wide launch. Brett Drake et al. pointed out that

---

[44] Brett Drake et al., "A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare," *Annals of the American Academy of Political and Social Science 692, no. 1* (November 2020): 162-181.

[45] David Jackson and Gary Marx, “Data mining program designed to predict child abuse proves unreliable, DCFS says,” *Chicago Tribune* (December 6, 2017), https://www.chicagotribune.com/2017/12/06/data-mining-program-designed-to-predict-child-abuse-proves-unreliable-dcfs-says/

[46] David Jackson, Duaa Eldeib, and Gary Marx, “Director of Illinois DCFS may leave amid ethics probe,” *Chicago Tribune* (May 9, 2017), https://www.chicagotribune.com/2017/05/09/director-of-illinois-dcfs-may-leave-amid-ethics-probe/

the Illinois case of ERSF "violated most or all of the key principles necessary for a practical and ethical use of such a method."[47]

(7) Case of Norrtälje and Strängnäs in Sweden

Two cases in municipalities in Sweden are worth noting because they used the NLP/TP model(natural language processing topic modeling) to capture the early risk of children being maltreated or neglected. The NLP/TP model, often used in the medical and healthcare field, uses machine learning to dig into documents' words to analyze case similarities. The NLP/TP model does not require specific predictive variables for analysis, which sets it apart from PRM models.

In 2020, the municipality of Norrtälje became the first local government in Sweden to invest in the NLP/TP-based child maltreatment prediction system. The concern of the Norrtälje government was the spiking number of child maltreatment reports, which gave an excessive workload to social workers and drained the local budget. Norrtälje developed a tool that cost 270,000 euros to assist social workers in deciding the initiation of a case after getting maltreatment referrals. The tool used a Robotic Process Automation system(RPA) to analyze the texts in previous maltreatment reports with the new case, assess the statistical similarity, and predict the child's risk.[48]

However, the Norrtälje RPA system failed to be implemented because of concerns that it could violate the Swedish data privacy law. The law prevented

[47] Brett Drake et al. 162-181.

[48] Lappalainen, 448-450.

the use of data which did not develop into investigation, as the cases might involve wrongful accuse of individuals. As both cases led to an investigation and were critical for comparing the differences in texts, the government decided not to use the RPA system in 2021. Without comprehensive access to historical access, the system's accuracy was harldy expected. There were also concerns that the system might reinforce the bias of investigation and disparities among municipalities.[49]

On the other hand, Strängnäs took a slightly different approach after the lesson of Norrtälje. They also got funding for developing an NLP/TP-based child protective system but decided not to automate the process fully. Instead, Strängnäs developed a text analysis model, which informs the social workers of the risk implications of the record but does not go further to decision-making assistance. Social worker in Strängnäs, Frida Fallstrom says, "We saw a black box effect in Norrtälje, and we didn't want to repeat it."[50]

---

[49] Alina Yanchur, "A Swedish town bought an AI to spot children at risk, but decided against deploying it," *Algorithm Watch*, August 10, 2021. https://algorithmwatch.org/en/norrtalje-children-at-risk-algorithm/#:~:text=deploying%20it%20%2D%20AlgorithmWatch-,A%20Swedish%20town%20bought%20an%20AI%20to%20spot%20children%20at,but%20decided%20against%20deploying%20it&text=The%20Swedish%20municipality%20of%20Norrt%C3%A4lje,the%20software's%20lawfulness%20and%20bias.

[50] Ibid.

VI. Controversy around the Predictive Use of Data

The implementation of PRM in child protection has long been contentious since its inception. Some experts in data science or child welfare viewed it as a novel idea to get rid of human subjectivity from child abuse responses, and make the burdensome workflow much more efficient with the support of algorithms. Others vehemently opposed the use of algorithms in child welfare as they might do nothing more than, at best, perpetuate or, even worse, exacerbate the existing human bias. They are also skeptical of the transparency and accuracy of PRM tools.

Around half of local governments in the US have developed PRM tools or are undergoing development. However, we should also note that many of them have already dropped the use of the tool despite considerable spending. The reason mainly comes down to inaccuracy, discrimination, and a lack of transparency.

Illinois is a widely known case where a severe level of inaccuracy made the PRM tool, Rapid Safety Feedback by Eckerd and Mindshare Technology cease to be used and even led to the resignation of then director of the Department of Child and Family Service, George Sheldon. The RSF system's accuracy was extremely low on false-positive and false-negative sides. It predicted over 4 thousand referred children to be 90% likely to be fatal or nearly fatal but failed to predict the real death of two infants in a row. Illinois case also came under fire as the development firm was deemed wrongfully connected to the director, passing by the official bidding system.

Illinois is not the only local government that went through messy controversy around PRM in child welfare. Oregon dropped its use of the predictive algorithm in 2022 after the racial and economic disparity was discovered in the early use of Allegheny County, where Oregon took the page out. In 2017, LA County dropped its experiment of the PRM tool with SAS due to the lack of accuracy and credibility of the tool.

The same cases happened abroad in the UK or Sweden. Hackney County in London developed an EHPS(Early Helping Profiling System) with a private company, Xantura, but the tool stopped in 2019. This was due to an inaccuracy of the tool and a lack of transparency, as Xantura refused to open the data or use the modeling method. An outcry also called for more respect for people's privacy and data autonomy. In Norrtälje, Sweden, the Robotic Process Automation system(RPA) failed to be implemented despite the cost of 270,000 euros because of the concern that it might violate Swedish data privacy law.

In regions where PRM tools are active, ongoing debates remain about whether the use is legitimate and fair. For example, Allegheny County's AFST, the most prominent PRM tool in the US, continues to be criticized by numerous academic papers or news outlets. It is said to be undergoing the Justice Department's legal scrutiny, which is casting doubt over the stability of PRM tools. In this chapter, I will walk through the controversial arguments about the use of PRM in child abuse response.

Figure 14: US Jurisdictions using/considering or dropped PRM analytics in child welfare[51]

1. Issues around Accuracy

(1) Arguments against the accuracy of PRM

- Human discretion shapes the modeling process

Virginia Eubanks points out the inherent limitation of the accuracy or objectivity of PRMs. In her book “Automating Inequality” regarding AFST, she says outcome variables, predictive variables, and validation data are set by humans, which must indicate the reflection of human decision. Specifically, the outcome variables by no means are available to capture the true abuse status but only reflect proxies for abuse, and the variables like “child placement” or

51 Adapte from, Anjana Samant, Aaron Horowitz, Kath Xu and Sophie Beiers, “Family Surveillance by Algorithm – The Rapidly Spreading Tools Few Have Heard of”, ACLU(September 2021), 8, with modifications.

“community re-referral” are made by human decisions. According to her, a model’s predictive power is harmed when outcome variables are subjective. Glaberson also argues that good proxies for child maltreatment are hard to find, as measures such as “substantiation” or “re-referral” hold human bias and inaccuracy.[52] It is also an inherent problem of PRM where rare events such as child death are not proper for outcome variables due to the lack of data, while those critical events really are that researchers and agencies want to prevent the most.[53] The statistical inability leads to setting arbitrary outcome variables, which are more common but not as important.

The decision of predictive variables and the validation data also heavily depend on human discretion, making “human bias a built-in feature of the predictive risk model.” When choosing predictive variables, governments and researchers have to only include accessible data. For this reason, many experts argue that PRM tools cannot make an objective, accurate prediction from the design.[54]

- Inaccuracy of the input(predictive) data

Some are concerned about the administrative data inaccuracy that goes into the models. Glaberson is representative for accusing the “GIGO”(garbage in, garbage out) issue, which undermines the fundamental credibility of PRM. The range and attainability of up-to-date and accurate data differ across local governments. Still, he points out that data entered by human officials can

---

[52] Glaberson, 342.

[53] Stephanie Cuccaro-Alamin et al., 295.

[54] Eubanks, 143-146.

often be wrong or outdated.[55] Actually, some failed cases, such as those in Illinois, are related to this issue. Illinois Rapid Safety Feedback program missed the death of Semaj Crosby and Itachi Boyle, and the input of flawed data took the blame for it. Chicago Tribune accused the system of "being riddled with data entry errors in both Semaj Crosby and Itachi Boyle cases"[56], and revealed that "it did not link investigations about many children to cases regarding their siblings or other adults in the same home."[57]

- The bias of the input(predictive) data

Another common concern is that the historical data is already biased, which might lead to biased results of the PRM tools. It is called the "BIBO"(bias in, bias out) problem.[58] In the worst-case scenario, human bias will be even exacerbated with time using PRM tools. Keddell worries about "Feedback loops", where people are screened in mainly because they were screened in before, and it would make a self-perpetuating loop in the system.[59] Many other experts, Drake, Samant, Cuccaro-Alamin, Yen & Heng, pointed out that if historical data were influenced by human bias, PRM would only perpetuate or amplify the bias.

Some argue that the NLP/TP model might be less problematic than PRM

---

[55] Glaberson, 337-338.

[56] David Jackson and Gary Marx, *Chicago Tribune*(December 6 2017)

[57] Ibid.

[58] Glaberson, 345-346.

[59] Emily Keddell, "Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice and the Implications for Practice", *Social Sciences 2019,8,281*(2019), 5.

models in relation to direct discrimination, which involves humans selecting the predictive variables. However, it still has the problem of "status quo bias in former decisions" by nature, which may lead to repeated bias.[60]

- Inability to reflect ongoing change

Glaberson also addresses the fact that PRM may not recognize historical changes(Zombie Prediction problem).[61] It is worrisome as the child abuse policy is rapidly changing, including the legislation of FFPSA and evolving perspectives on neglect. Specifically, there have been growing voices accusing the States of treating poverty equally to neglect and that the definition of neglect should be modified to reflect differing circumstances related to poverty. Iowa has revisited the definition of neglect as "the failure on the part of a person responsible for the care of a child to provide for adequate food, shelter, clothing, medical or mental health treatment, supervision, or other care necessary for the child's health and welfare when financially able to do so or when offered financial or other reasonable means to do so"[62] Given the changes, historical records of neglect, which make up over 60% of all substantiated cases, would not reflect the policy shift.[63]

---

[60] Katarina Fast Lappalainen, "Protecting Children from Maltreatment with the Help of Artificial Intelligence: A Promise or a Threat to Children's Rights?", Law, AI and Digitalization(2022), 463.

[61] Glaberson, 344.

[62] Jill Yordy, "Poverty and Child Neglect: How Did We Get It Wrong?," *National Conference of State Legislatures* (February 21, 2023), https://www.ncsl.org/state-legislatures-news/details/poverty-and-child-neglect-how-did-we-get-it-wrong

[63] Danielle Whicher, Emma Pendl-Robinson, Kyla Jones, and Allon Kalisher, "Avoiding Racial Bias in Child Welfare Agencies' Use of Predictive Risk Modeling"(Assistant Secretary for Planning and Evaluation Office of Human Services Policy, 2022), 6.

- The problem of over-presented false positive

When designing a PRM model, researchers face the issue of deciding the right level for risk threshold. This is because of the inverse relation of sensitivity and specificity. If the threshold is high, the model's specificity - the probability of capturing true positives among positive outcomes – is increased. However, it comes at the risk of large false negatives (or Type 2 errors). Conversely, if the threshold is too low, the model's sensitivity increases to minimize the false positive cases(Type 1 error), but the rate of false negatives increases. In this dilemma, many child welfare agencies decide to lower the risk threshold to minimize the probability of missing children at risk.[64] However, it means that innocent families are targeted for child maltreatment, possibly giving them unnecessary investigation and stigma.

(2) Arguments supporting the accuracy of PRM

On the other hand, some assert that PRM tools enhance accuracy as they hold consistency, whereas human decisions sometimes rely on their guts, leading to different choices in similar cases. For PRMs, regardless of the operator's experience or training level, the tool shows consistent results for a case. Indeed, the human assessment of abuse risk has long been criticized as inconsistent, leading to the continuous evolution of decision-making tools in the child welfare field. Cuccaro et al. argue that PRM ensures more accuracy than previous SDM tools, which are prone to subjectivity and lack fidelity.[65] In the paper on the ethical analysis of AFST, Dare and Gambrill even argued that

[64] Cuccaro et al., 295.

[65] Cuccaro et al., 293.

“PRMs are more accurate than any alternatives, as there are fewer errors than manually driven actuarial risk assessment tools.”[66] It is also the first type of tool which are validated in the specific subpopulation for which the PRM is developed and applied.

Also, PRM tools make it possible to assemble related information previously segmented in each agency. Each data might not critically impact the assessment of risk, but only when put together, give a broader and meaningful look at the whole situation. For example, in Pennsylvania, Child Protective Services addresses child abuse cases, while General Protective Services deals with neglect. The separation of abuse and neglect data into each service made it impossible for responders to assess the future risk of a child being abused even when the child has multiple reports of neglect.[67]

(3) Empirical research about the accuracy of PRM

Empirical research about the accuracy of PRM is still conflicting. Based on AFST from August 2016 to July 2018, Stanford University looked into the tool's accuracy among screened-in and-out children. The accuracy for screening-in was measured whether children screened in during the period had further action taken from the office or, if not, re-referred within several months. The accuracy for screening-out was measured if the screened-out children were re-referred within several months. The impact evaluation concluded it

[66] Tim Dare and Eileen Gambrill, “Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County”(April 2017), 4.

[67] Naomi Schaefer Riley, “Can Big Data Help Save Abused Kids?,” *American Enterprise Institute* (February 6, 2018), https://www.aei.org/articles/can-big-data-help-save-abused-kids/

enhanced the accuracy for screen-in but decreased it for screen-out compared to previous records.[68] On the other hand, From the same data set, Cheng showed that the AFST had a more accurate result predicting the intended outcome, whether the child would be removed from home within two years or re-referred in 2 months. The accuracy of AFST was 51%, but workers' final decision lowered it to 46.5.[69]

However, Eubank, in the same volume, addresses the fact that the accuracy of a model cannot be confirmed simply by the fact that the model predicts its intended outcome. She explains that a high score of AFST will lead the social workers to keep an eye on the case and likely place the child in foster care more often.[70]

2. Issues around Discrimination

(1) Arguments accusing PRMs of worsening disparity

- Concern about racial disparity

The accuracy issues, especially BIBO, lead to a significant concern around socioeconomic disparity. Many experts raise concerns that if certain racial group was targeted disproportionately as abusing children, it could be

---

[68] Jeremy D. Goldhaber-Fiebert and Lea Prince, 19-21.

[69] Hao-Fei Cheng et al., *How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions*(2022),14.

[70] Eubank, 169.

reflected in the PRMs and be repeated. Historically, a much larger number of Black and Hispanic families have been targeted for child abuse response. In New York City(2012), Black children made up 59.8% of foster care while their proportion in the population was only 25.9%.[71]

- Concern about racial disparity in AFST and the drop in Oregon's PRM tool

The racial disparity is undoubtedly a big concern enough to make some local governments stop using PRMs. In 2022, a Carnegie Mellon University team researched the 2016-2018 AFST data, revealing that AFST screening decisions would make larger racial disparities than human decisions. The research compared the difference in screen-in rates between Black children and white children and concluded that AFST-only decisions gave a bigger disparity(20%) than workers' final screening decisions(8.9%). Specifically, according to AFST's recommendation, 71% of referred Black children should be screened in, whereas the figure for white counterparts was 51%. The disparity(20%) decreased to 8.9% after workers made final decisions: 61.8% for Black children and 52.8% for white children. In the research, social workers said they adjusted for limitations of the AFST, where systemic racial biases could have played out, and made holistic and contextual assessments.[72] Oregon dropped its tool right after the accusation. It is said that AFST has gone under legal scrutiny by the Justice Department following concerns that it could be discriminatory against families with disability and mental health

[71] Naomi Schaefer Riely

[72] Hao-Fei Cheng et al., 15-22.

issues.[73]

- Racial disparity in accuracy

It is important to note that some focus on the disparity of likelihood of having high PRM scores, and others focus rather on differences in levels of accuracy. Cheng et al. claimed that despite the overall higher accuracy of AFST, AFST-only decisions had higher racial disparity in accuracy than worker-AFST decisions. AFST-only decisions showed 57.5% accuracy in white children and 44% in Black children(13.5% disparity). Worker-AFST decisions showed 48% accuracy in white children and 42.6% in Black children(5.4% disparity).[74]

- Concern about socioeconomic disparity

There is also a concern that as data from the PRM model depend heavily on public service records, poor people who cannot afford private service are likely to be overrepresented. Virginia Eubank argues AFST makes "poverty profiling." This is to say, that AFST wrongfully targets poor families as it includes public service use in its predictive data. As upper-class households that use private nannies, insurance, therapists, and rehabilitation centers do not have their data shared with the DHS, the system itself puts a thumb on a scale in favor of rich people. She also claims that among 131 predictive variables, nearly half represent household poverty and the use of means-tested programs such as TANF, Supplemental Security Income, SNAP, and

---

[73] Sally Ho and Garance Burke, "Child welfare algorithm faces Justice Department scrutiny," *AP* (January 31, 2023), https://apnews.com/article/justice-scrutinizes-pittsburgh-child-welfare-ai-tool-4f61f45bfc3245fd2556e886c2da988b

[74] Ibid, 25.

county medical assistance.[75] It raises concern that the system might exacerbate the current trend of accusing poverty of child neglect.

(2) Arguments denying the relation of PRM and disparity

- Arguing the disparity is not solely attributed to PRM

Proponents of PRMs argue that racial disparities can also be problems with alternative decision methods—operators' instinct, consensus-based decision, and actuarial decisions—and PRMs are available to "provide an opportunity to openly and systematically track disparities and correct for them that may have remained hidden under alternative approaches."

- Arguing PRMs can help overcome the disparity

Some experts believe the objective nature of PRM can actually mitigate the human bias in child abuse response. After several years of PRM's implementation, some recent studies have shown that PRM reduced racial disparity rather than expanded it. Goldhaber-Fiebert and Prince, Rittenhouse et al., Grimon and Mills found that using PRM tools reduced the racial disparity gap in several child protective services, including screen-in, substantiation, and removal. Specifically, according to Rittenhouse et al., in Allegheny County, the Black-White gap in screen-in rates decreased by 46%, and that for the highest risk referrals by 83%(from 10.6% to 1.8%). The removal rate among screened-in children also showed a similar trend. The Black-White gap of removal rate decreased by 73%(from 4.3% to 1.2%).[76]

---

[75] Eubanks, 125.

[76] Rittenhouse et al., "Algorithms, Humans and Racial Disparities in Child Protection Systems:

Table 5: Impact of Child Welfare Algorithms on Racial Gaps[77]

Study	Location	Screened In	Accepted for Service	Home Removal	Hospitalization
Goldhaber-Fiebert and Prince	Allegheny County	32%	92%	100%	N/A
Rittenhouse	Allegheny County	46%	91%	73%	N/A
Grimon and Mills	Larimer County	50%	N/A	N/A	56%

- Arguing that PRM does not exacerbate socioeconomic disparity

There is an opposing view to Virginia Eubank's accusation of PRM's poor profiling feature. According to an executive statement published by the Allegheny County Department of Human Services in response to Eubank's "Automating Inequality," the prediction using public service data actually showed the opposite result. In reality, 45% of families' receipt of public services lowered the AFST score. There was a positive correlation between the use of public service and the likelihood of child abuse.

At the same time, Samuel Oh argues that families from lower socioeconomic status can improve their accuracy because their use of public services provides additional data, leading to accurate PRM prediction.[78]

Evidence from the Allegheny Family Screening Tool", (2023), 15-17.

[77] Allegheny County Department of Human Services, *Summarizing Recent Research On Predictive Risk Models in Child Welfare*(Apr 2024), 7.

[78] Samuel Oh, "An Ethical Evaluation of the Use of Predictive Risk Models in Health and Human Services: A Case Study of the Allegheny Family Screening Tool", *Journal of Politics and Society*

3. Issues around Data Privacy

(1) Arguments accusing PRMs of violating data privacy

Some experts show grave concern that using PRM might be intrusive regarding data privacy. It is because of the intrinsic nature of PRMs. Unlike previous actuarial models, PRM tools usually include a wide range of data, and a large part comes from fields other than child welfare. In many cases, the data from hospitals, school attendance, getting public service, and others seems to have little connection to child abuse. This means that different agencies collect various input data from the end-users of PRMs. Though people might have consented to the provision of personal data to a government agency, it is hardly possible they expected those data might be used in the prediction tool for child abuse. Gladberson raises concern that using personal data in an algorithm that scores child abuse risk, whether or not consent was obtained, could be unexpected in most cases.[79] An information science scholar, Helen Nissenbaum, explained the importance of "contextual integrity" in data privacy, which means "what raises privacy hackles is not the sharing of data in and of itself, but the moments when personal information flows in ways we may not expect, flouting established informational norms."[80]

The problem gets bigger when there is substantially no choice but to give

---

XXX, No.1, (2020), 16.

[79] Glaberson, 348.

[80] Helen Nissenbaum, *Privacy in Context: Technology, Policy and Integrity of Social Life*(2010), 127.

consent in exchange for essential services.[81] As public service providers require consent for personal data sharing, it can be regarded as a transaction of data and public service, which is not very satisfactory for the public. The fact that PRM tools primarily depend on public service data even adds to another concern that “poor mothers are deprived of privacy rights.”[82]

(2) Arguments denying the data privacy issue

At the same time, there is a counterargument saying the tools are only supporting to “do a better job crunching the numbers that are already collected.”[83] Actually, it is true in many cases, including Allegheny County, where social workers already had access to all data put into the PRM. Still, they didn’t have enough time and capability to analyze the data to get a meaningful insight into children’s risks. According to Dare&Gambrill, the model does not create new rights of access to that information –a diligent child protection official would already have been entitled to gather the information now to be accessed by the tool.[84]

---

[81] Tim Dare and Eileen Gambrill, 2.

[82] Dorothy E. Roberts, *The Only Good Poor Woman: Unconstitutional Conditions and Welfare*(1995), 939.

[83] Naomi Schaefer Riely

[84] Tim Dare and Eileen Gambrill, 3.

4. Issues around Implementation

(1) Arguments that PRMs can harm human judgment

Some argue that the PRMs can do more harm in implementation than they are designed to do. It is because of "automation bias"[85] – people tend to give excessive credibility to automated results, even if the credibility is low. It leads people to give outsize authority to machines and over-rely on them. This is very worrisome because the PRM tools have limited accuracy by nature and because they hamper the human ability or efforts to look into each case and make their own clinical judgment. Citron said, "Seeing automated systems as authoritative and neutral, users tend to be less likely to search for information that would contradict a computer-generated system."[86] Glaberson and Eubanks also show deep concern about over-crediting the tool, as it could degrade or dehumanize clinical social work. Eubanks figured out that in the early application of AFST, even experienced call screeners tended to revise their risk assessment after getting the score from AFST.[87]

This phenomenon is amplified in a high-stakes or high-blame environment.[88] Elish coined the term "moral crumple zone" to describe the blame on humans when AI malfunctions; if a highly automated system makes

---

[85] Danielle Keats Citron, *Technological Due Process*(2008), 1271.

[86] ibid.

[87] Virginia Eubanks, Automating Inequality: How High-teck Tools Profile, Police and Punish the Poor(St.Martin's Press, 2018), 141-42.

[88] Emily Keddell, 14.

a mistake, human operators with limited authority tend to fall for it, as a crumple zone in a car.[89] In an environment where social workers are easily placed in a moral crumple zone, they could become passive in overriding the decisions of PRMs. Some screeners actually said they feel pressure to follow the suggestions of PRMs and are reluctant to override the decisions.[90]

(2) Arguments against PRMs harm human judgment

Some experts say PRM tools are prudently used in the field. Empirical evidence shows that human subjection to algorithms is not a worrisome issue. De-Arteaga figured out from 2016 to 2017 AFST data that call screeners did not always follow what AFST recommended. Especially when there was a technical error in AFST in the early days, human screeners did not defer to the AFST scores but assessed more accurate and desirable results.[91] Also, Chouldechova found that in the early implementation of AFST, supervisors revised about a quarter of mandatory screen-in cases recommended by AFST.[92]

---

[89] Madelein Claire Elish, "Moral Crumple Zones: Cautionary tales in human-robot interaction", *Science, Technology, and Society5*(2019), 40-60.

[90] Anna Kawakami et al., "Improving Human-AI Partnerships in Child Welfare: Understanding Worker Practices, Challenges, and Desires for Algorithmic Decision Support", *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*(2022)

[91] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova, "A Case for Humans-in-the-loop: Decisions in the Presence of Erroneous Algorithmic Scores", *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*(2020)

[92] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan, "A Case Study of Algorithm-Assisted Decision Making in Child Maltreament Hotline Screening

There are also a view that the de-humanization is nothing to worry about. Concerns of de-professionalization along with automation has long existed, even before the PRM was developed. When the actuarial decision making tool was introduced, some scholars concerned that it would hamper the human workers' critical thinking, but it was proved to be a wrong accusation. For the same reason, a paper from NIH shows a confidence in human workers' professionalism.[93]

Decisions", *Proceedings of the 1st Conferences on Fariness, Accountability and Transparency*(2018), 134-48.

[93] Brett Drake et al., 162-181.

Table 6: Key Issues in PRM in Child Abuse Response

Issues	Arguments against PRM	Counter-arguments
Accuracy	- Human discretion shapes modeling process, leading to subjective outcomes - Inaccurate input data (GIGO issue) - Bias in historical data (BIBO issue) - Inability to reflect ongoing changes (Zombie Prediction problem) - Over-representation of false positives due to low risk thresholds	- Consistent results regardless of operator experience - Validated in specific subpopulations - Integrates segmented data for comprehensive risk assessment
Discrimination	- Historical racial bias may be perpetuated or amplified (BIBO issue) - Racial disparity in accuracy persists - Socioeconomic disparity concerns due to reliance on public service data	- Provides an opportunity to track and correct disparities systematically - Some studies show reduced racial disparity with PRM use(AFST)
Data Privacy	- Intrusive due to inclusion of data from various fields - People don't expect their data in agency be used for potential surveillance - Poor families may be more forced to trade their data authority for public sevice	- PRMs do not create new rights of access to information, only support existing data analysis capabilities
Implementation	- Risk of automation bias, leading to over-reliance on PRM tools - Human judgment may be undermined, leading to passive behavior in overriding PRM decisions - High stakes environment may place undue blame on human operators when AI malfunctions (Moral Crumple Zone)	- Supports decision-making with consistent data analysis - Can integrate information across different agencies for comprehensive risk assessment

VII. Pathway for Better Use of PRM in Child Welfare

1. Enhancing Transparency

The transparency of the data and methodology used in the algorithm is a key factor in ensuring the credibility of PRM systems. Eckerd and Mindshare, the company that developed Rapid Safety Feedback, refused to make the information public even after the system failed to detect two lethal cases in Illinois.

On the other hand, the Allegheny Family Screening Tool, developed by academia, is owned by the county, and all the details of the data mining process are made public. Several public meetings to prove the system's accountability were held before the launch of the AFST. Legal experts, family representatives, foster kids, and advocates were invited to the hearings. Erin Dalton, Allegheny County director general, expressed how seriously the county took the transparency issue, saying, “We are trying to do this the right way, to be transparent about it and talk to the community about these changes. It’s concerning because public welfare leaders 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.”[94] In a recent paper, the County reiterated its belief in transparency: "The most important principle in any

[94] Dan Hurley, “Can an Algorithm Tell When Kids Are in Danger?,” *New York Times* (January 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html

technology implementation is transparency."[95]

Additionally, the open discussion involving a wide range of stakeholders made the Allegheny case successful, which didn't happen in many counties. For example, when the DHS in Allegheny County wanted to share information with schools, it faced experts' objections. However, the DHS put a lot of effort into meeting legal aid and civil liberties groups to discuss implications and tried to get consent from parents' rights groups.[96]

This teaches an important lesson about how transparency is important to garnering public support for PRM in the child welfare field. Some say that, given the openness of all data, analyzing processes, and evaluation results, the transparency of PRM tools actually far outweighs that of other alternatives, such as actuarial decision methods.[97]

- Make data and algorithms open to public

From the case of Allegheny County, we learned that disclosing all the related data and methodology to the public is essential to gaining public credibility and external vigilance for the tools. Making all information public is not only important for ethical reasons but also essential to promoting the tool's performance, as openness ensures the process of knowledge building.[98]

On this basis, some experts argue for setting a rule or legislature for

---

[95] Allegheny County Department of Human Services, *Summarizing Recent Research On Predictive Risk Models in Child Welfare*(Apr 2024), 10.

[96] Naomi Schaefer Riely

[97] Tim Dare and Eileen Gambrill, 4.

[98] Brett Drake et al., 162-181.

information disclosure. Glaberson argues that for transparency, public agencies should make the source code and detailed algorithm open to the public('open source'), which is quite common in the field of data science. He also suggests that states and municipalities take action to mandate disclosure by laws or regulations.[99]Abrams Institute suggests a legislature that requires affirmative disclosure, including on their websites, about certain information, which is opposite to request-based disclosure.[100] Concerning that external scholars may not know the existence of certain information, request-based disclosure might not work for the purpose.

As open source transparency is essential, sharing all the rationales and principles of PRM tools in plain language also holds a significant meaning. Data variables, outcomes, accuracy, and algorithm methodology are often highly complicated and require expert-level knowledge to comprehend the meanings. Therefore, agencies should deliver the rationale, effect, evaluation results, and potential risks in a plain manner, which is essential to gaining public support.[101]

- Avoid partnering with institutes that make proprietary ownership

Scholars and agencies arguably reached a consensus to avoid the proprietary use of PRM tools, which impede disclosing related data and modeling methods to the public. To what extent should public transparency

[99] Stephanie K. Glaberson, 358.

[100] Abrams Institute, "Algorithmic Accountability: The Need for a New Approach to Transparency and Accountability When Government Functions Are Performed by Algorithims" (Media Freedom & Information Access Clinic, Yale Law School, 2022), 28.

[101] Stephanie K. Glaberson, 359.

outvalue trade secrecy? There are slightly different views.

Whicher et al., in a report of the Assistant Secretary for Planning and Evaluation, said that when developing PRM tools, agencies should make a contract that obligates all vendors to waive proprietary information. Some argue that to ensure the transparency of PRM tools, it is better to work with non-profit organizations to avoid the proprietary issue from the beginning. Brett et al. express a strong belief that “all algorithms must be public, and proprietary agencies that refuse to share their models should never be used.”[102]

At the same time, some experts do not call for an unconditional surrender of proprietary rights. Abrams Institute, while opposing the current legislatures in favor of private parties’ trade secrecy, argues that “the legislature could balance these factors by requiring private parties to disclose the source code and training data only for algorithms that exceed a certain risk threshold.”[103]

- External engagement

Cuccaro mentions that to ensure the openness of the data-analyzing process, agencies should consider forming advisory panels, releasing model input data, and evaluating performance results, preferably by external experts or subject to external review.[104]

[102] Brett Drake et al, 162-181.

[103] Abrams Institute, 29.

[104] Cuccaro, 295.

2. Improving Accuracy

Needless to say, the accuracy of PRM tools should be guaranteed before they are implemented and continuously checked during the operation. We've learned the hard lesson from the Illinois case of how vital accuracy is in child risk assessment.

- Set a good proxy outcome.

First, setting a concrete and appropriate outcome variable that represents the result of interest is crucial. Setting a good proxy outcome variable is indeed difficult which is representative of "ground truth", whether the child was abused or maltreated in real. In the case of rare events like death, it is more difficult. However, researchers can externally validate the relation of proxy outcome and real outcome of interest, using other external indicators.

The case in Allegheny County is a good example to follow. When they developed AFST2, they validated the outcome – child removal in 2 years – by comparing the AFST score with hospitalization rates. This external validation is thought to support the accuracy and credibility of PRM outcome.[105]

- Balance the specificity and sensitivity.

It is also necessary to ensure the inverse relation of specificity and sensitivity and try to make additional concerns to minimize the consequence of a type of error by implementation.[106]

[105] Vaithianathan et al., "Allegheny Family Screening Tool: Methodology, Version2", 7.

[106] Danielle Whicher et al., 14.

- Use data rich in breadth, depth, and quality

The performance of PRM tools is only as good as their input data. Scholars emphasize that the data should be rich in breadth and depth and high in quality. It is essential to guarantee the accuracy and external validity of models. Additionally, many efforts must be made to promote buy-in and ownership of frontline data entry workers.[107] This is critical in improving model performance and the predictive power of the PRM.

Table 7: Features of Data Appropriate for Predictive Analytics[108]

Concept	Description
Breadth	- Availability of data on many variables related to the outcome for a sufficient population. - The breath of data ensures data generalizability - Proper level of breadth should be locally determined with data quality issues addressed before analysis
Depth	- Availability of data over a sufficient duration to observe outcomes over time - Allows examination of long-term outcomes like maltreatment re-reports and out-of-home care re-entries
Quality	- Reliability, validity, and comprehensiveness of data. - Depends on consistent, informed data collection by staff. Requires regular review for missing data, entry errors, and efficient data collection practices. - staff training is crucial.

107 Cuccaro, 295.

108 Chapin Hall& Chadwick Center, "Making the Most of Predictive Analytics: Responsible and Innovative Uses in Child Welfare Policy and Practice"(San Diego, CA & Chicago, IL, 2018), 6-7.

- Evaluate the model

Before and after the implementation of PRM tools, agencies should always monitor their performance and whether it is working well over time. If the performance deteriorates, it should be recalibrated or remodeled with new variables. Many experts call for proper evaluations of PRM algorithms in and out of the child welfare field. Abrams Institute takes it seriously that many government agencies are failing to assess the bias or effectiveness of algorithms and suggests a mandatory algorithmic assessment- before the implementation and at regular intervals while the tools are deployed.[109] The institute also addresses the required assessments that should be carried out both by the agency and external experts. Chapin Hall & Chadwick Center also emphasized the importance of "examining the validity and reliability of models on an ongoing basis."[110]

Legislators are also showing keen interest in this issue. In 2020, U.S. Senator Sherrod Brown of Ohio introduced the Data Accountability and Transparency Act("DATA Act"), which is still referred to the Committee on Commerce, Science and Transportation.[111] The bill requires public agencies to operate automated decision system risk assessment to address accuracy and bias issues before implementation and on annual intervals.

---

[109] Abrams Institute, 26.

[110] Chapin Hall& Chadwick Center, 11.

[111] Ibid.

3. Minimizing the Disparity Problem

We should carefully consider the input variables. When choosing available input variables, evaluate whether those data are riddled with historical and human cognitive bias. We should also check if the data is overrepresenting certain subgroups of the population. If it is the case, we should consider the reason for missing data and address measures to reduce disparity. This might possibly include reweighting variables.

When developing the PRM, agencies should run the model to check the predictive performance of the overall system and that of subgroups. The performance should be remodeled or discarded if it differs significantly across race, ethnicity, or other groups.

It is helpful to engage community members. The perception of fairness can differ across communities, so it is essential to have a selected definition of fairness when planning for the PRM. (Bias report 9) Make sure the ethics review committee is engaged with diverse representation.[112]

It is also essential to incorporate protective and positive factors in input variables, as the negative variables might be potentially biased against certain groups.[113] Protective and positive factors, representing strong family bonds and community support, provide a more comprehensive view of each case. Relying solely on negative variables, such as previous abuse reports or parental substance abuse, can introduce biases, potentially targeting certain groups disproportionately due to systemic issues like socioeconomic

---

[112] Chapin Hall& Chadwick Center, 11.

[113] ibid.

disparities and racial discrimination. By incorporating a balance of protective, positive, and negative factors, the model can achieve greater fairness, ultimately leading to more equitable and effective interventions.

4. Protecting Data Privacy

Most scholars disagree with the specific resolution of data privacy issues. It is a complex issue because the violation of data privacy might fall in the realm of subjection. Though people might have conceded the use of their data to DHS, their expected use might differ for all individuals. Some might have higher bars than others. Therefore, it is hard to set a certain standard for data privacy, which should be generally used for PRMs in child welfare. Instead, it should be considered based on the tool's intended purpose and society's requirement for data privacy, which might vary a lot across every society. Considering these aspects, Cuccaro said, “The needs for privacy and due process must be balanced with the agencies' duty to ensure the safety of children.”[114]

NIH views the use of data already possessed by the agency as largely exempt from the consent requirements. NIH says this is analogous to hospitals analyzing their patients' health data for research and experimentation.[115] However, considering the context of child abuse response, which could be very sensitive and intrusive to some households, we cannot expect the same level of endurance or generosity from the public.

---

[114] Cuccaro, 295.

[115] Brett Drake et al., 162-181.

5. Ensuring Efficient Implementation

- Decide when and how to use the PRMs

When it comes to implementation, it is essential to decide at which point of the abuse response process and for which purpose the tool will be used. In most cases in the US, the tools are used at the outset to support the decision to screen in or out. In some cases, they are applied even in an earlier phase to prevent and check unreported maltreatment. In cities like New York, the tools are applied across many phases for the efficient use of resources.

Dare and Gambrill emphasized implementing the PRM tools for preferably non-intrusive purposes. It is to minimize the negative consequences arising from the nature of prediction; tools cannot predict with 100 percent accuracy. In the field of child welfare, where agencies tend to prioritize sensitivity to specificity, it is essential to avoid punitive intervention, such as child removal, as a response to the prediction. It is better to implement the tools to start an investigation with caution or provide additional services, considering there might be false positive cases from prediction.[116]

Glaberson also supported the preventive use of PRM tools in the child welfare field. Specifically, he argues that PRMs can make predictions on a community level and be used to direct more attention or preventive resources rather than intervene for a certain family after a crisis happens.[117] He explicitly opposes the potential use of PRMs in the decision of child removal. This is not only because child removal from the original household is intrusive but also

---

[116] Tim Dare and Eileen Gambrill, 4.

[117] Glaberson, 361-2.

because the intervention requires an “imminent threat” to the child, which the PRM tools can never assess.[118] The removal of children should only be decided based on each child’s unique and specific circumstances.

- Adequate training of end users

It is never enough to emphasize the appropriate implementation of workers who use the tools. We should always train all the staff members to have enough knowledge about PRM tools. The training should include the “intention and development method of PRMs, the model outcomes, accuracy, its benefit, and limitation.” Also, training them in the appropriate use of PRM tools is essential. All staff members should comprehend the purpose of PRM tools is no more than “supporting” their work, and their judgment on cases should be prioritized over PRMs.[119]

Allegheny County acknowledges the importance of end-user training. In a recent paper, DHS in Allegheny shared its future tasks to improve human-algorithm interaction, which involves “building onboarding training modules that cover the science and implementation of the AFST” and “providing informed, tailored feedback and guidance to call screeners on situations in which they override the algorithm in ways that may either improve or weaken performance.”[120] Allegheny County's future pathway explains the significance of giving a better training opportunity to enhance the intended implementation of PRM tools.

---

[118] Glaberson, 360.

[119] Danielle Whicher et al., 15.

[120] Allegheny County Department of Human Services, 10.

Table 8: Features of Data Appropriate for Predictive Analytics[121]

Aspect	Key Considerations
Transparency	- Make data and algorithms open to the public - Avoid proprietary partnerships - Clearly communicate PRM rationale, effects, evaluation results, and potential risks in plain language - Engage with the community and hold public meetings
Accuracy	- Ensure the PRM tools are validated and continuously checked - Set appropriate and representative proxy outcome variables - Use data rich in breadth, depth, and quality - Balance specificity and sensitivity
Fairness	- Carefully select input variables to avoid historical and cognitive biases - Ensure data represents all subgroups fairly - Incorporate protective and positive factors - Engage community members and ethics review committees
Data Privacy	- Balance the need for data privacy with the duty to ensure child safety - Consider societal requirements for data privacy and the intended purpose of the tool - Address privacy concerns and obtain necessary consents
Implementation	- Make a concrete purpose about when and how to use the PRM tools (e.g., at the outset of cases for screening decisions, preventive use) - Use PRM tools for non-intrusive purposes to minimize negative consequences - Provide adequate training for end users on the proper use of PRM tools - Emphasize that PRM tools are to support, not replace, human judgment

121 Chapin Hall& Chadwick Center, "Making the Most of Predictive Analytics: Responsible and Innovative Uses in Child Welfare Policy and Practice"(San Diego, CA & Chicago, IL, 2018), 6-7.

VIII. Implication for the Korean Child Abuse Response Policy

1. The current standing of digitalization in Korean child abuse policy

Since 2018, the Korean Ministry of Health and Welfare(MOHW) has operated the “E-Child Happiness Support System”, the first PRM-based system in Korea to protect children at risk. “E-Child Happiness Support System” aims to analyze 44 social security data points to early detect marginalized children. The data is not only confined to child welfare or child protective service, but also include various social security and health information; school attendance, mandatory health check-up records, domestic violence, etc.

Every quarter, MOHW and the Korea Social Security Information Service (SSIS) run the system and list out 30,000 children who are predicted to be at risk. The lists are sent to local government social service offices, and officials in charge go for home visitation for each child, to check their safety. If they suspect the child is at risk of maltreatment, it is referred to as abuse report for investigation. If they conclude the child or the family are in need of social services, they connect them with proper service program.

However, the performance of E-Child System is continuously on public criticism. In the 2021 National Assembly audit, Shin Hyun-young of the Democratic Party pointed out, "Over the past three years, there have been about 40,000 cases of child abuse, but only 134 cases, or 0.3%, were identified by the e-Child Support Happiness System operated by the Korea Social Security Information Service."[122] In 2023, a Segye news criticized the same problem again, “The rate of child abuse detection through the E-child Happiness Support System is remarkably low. There are more than 30,000 cases of child abuse occurring annually. This stands in stark contrast to the mere 98 cases detected by the e-Child Happiness Support System so far.”[123]

2. Suggestion for digital governance in child protection system

The E-child system's accuracy could be improved, but first, it is crucial to define its intended purpose clearly. Currently, the E-child system appears to serve multiple roles, including child abuse prevention, early detection, and providing social services to vulnerable families. To avoid unnecessary criticism regarding its accuracy, it is essential to establish a concrete purpose for the E-child system. Clarifying the tool's primary target is the first step toward enhancing the model's accuracy.

If the model's intended use is early detection of children likely to be abused in the near future, it should be adjusted to include variables with stronger

---

[122] 이재혁, “(국감현장) 신현영, 사회보장정보원 e 아동행복지원시스템 발굴지표 개선해야“, *메디컬투데이*(21.10.19), https://mdtoday.co.kr/news/view/1065597721566349

[123] 조희연, “(단독) 부모가 애 숨기면 조사 못해, 위기아동 발굴시스템 헛바퀴”, *세계일보*(23.11.19), https://www.segye.com/newsView/20231119508077?OutUrl=naver

predictive power, and its predictive performance should be rigorously evaluated. However, we must always be cautious about the proper use of the model, especially since it is used before any referrals or suspicions of abuse arise. Unlike other PRM tools in the US, such as the Allegheny Family Screening Tool (AFST), the E-child system operates in a much earlier time frame. Any punitive use of the tool at this stage is likely to cause resistance from targeted families and face backlash. In the US, the Birth Match system operates similarly to the E-child system in terms of the time frame used, as both identify at-risk children before any referrals are reported. However, the Birth Match system only targets parents with extremely severe issues, such as those who have had their parental rights terminated or have killed children. In contrast, the E-child system targets a broader range of families and checks a significantly larger number of children (30,000 every three months for E-child, compared to a few hundred per year for Birth Match in a state). This suggests that the E-child system might be better suited to providing support and services to vulnerable families with children.

Second, apart from the E-child system, the Korean government could consider developing additional PRM tools for use in child abuse investigations, taking inspiration from the Allegheny Family Screening Tool. Such PRM tools can prioritize children who are in imminent, severe threat, enabling caseworkers to respond more efficiently. This approach is expected to protect at-risk children more quickly and alleviate the excessive burden on caseworkers. The Korean environment is more suitable for the application of PRM tools. In the US, child welfare service data is often limited to individual states or even counties, with some counties, like Douglas, only having access to data within child and family agencies. In Korea, a wide range of social service data across agencies is available nationwide. Additionally, the social perception towards PRM tools is likely to be more accepting in Korea, whereas in the US, concerns about racial disparity have been a significant opposition to PRM tools in child welfare. Furthermore, Korean governments typically avoid proprietary ownership of software, alleviating transparency concerns that have arisen in states like Illinois.

Third, there should be a stronger push toward child welfare prevention. Unlike the punitive, responsive approach in investigations or child removal, assessing the real outcomes of abuse prevention is challenging. Therefore, the US approach of the Family First Prevention Services Act (FFPSA) is a valuable model to follow. FFPSA shifted the focus to data-based, empirically supported preventive services. Since its introduction in 2018, states have been eager to assess their preventive services with empirical data and expand those with proven statistical outcomes. The federal government played a crucial role by establishing a Clearinghouse, a digital platform for objectively evaluating numerous preventive service studies. The Korean government can implement a similar system, enabling the evaluation of small and separate preventive services by the Ministry of Health and Welfare or local governments in an objective manner, thereby enhancing the use of truly productive services.

IX. Conclusion

Integrating predictive risk models (PRMs) into child welfare systems offers the potential to enhance early detection and intervention for at-risk children, but it requires careful development. Transparency is vital; data and algorithms should be open to scrutiny, avoiding proprietary constraints. Additionally, ensuring accuracy through careful selection of proxy outcome variables, continuous model evaluation, and high-quality data use is critical. Addressing disparities is also essential for the stable use of these tools. Models must mitigate biases and perform consistently across demographic groups. Moreover, data privacy must be respected, balancing child safety needs with privacy rights.

Effective implementation of PRM tools requires thorough training of frontline users, emphasizing that these tools support, not replace, human judgment. Focus should be placed on non-intrusive applications, such as early intervention and preventive services, to harness PRMs' potential while minimizing negative impacts.

For Korea, adopting best practices from successful international models and focusing on preventive services can significantly enhance child welfare interventions. By learning from global experiences and tailoring approaches to local needs, Korean government can develop a robust, data-driven child protection system that ensures the safety and well-being of all children.

-Bibliography-

1. Abrams Institute. "Algorithmic Accountability: The Need for a New Approach to Transparency and Accountability When Government Functions Are Performed by Algorithms." Media Freedom & Information Access Clinic, Yale Law School. 2022.

2. Administration for Children & Families, Children’s Bureau. Title IV-E program. https://www.acf.hhs.gov/cb/title-iv-e-prevention-program.

3. Allegheny County Department of Human Services. "Summarizing Recent Research On Predictive Risk Models in Child Welfare." April 2024.

4. Auditor of the State of California. California State Auditor Report 2018-126. May 2019.

5. Chapin Hall & Chadwick Center. "Making the Most of Predictive Analytics: Responsible and Innovative Uses in Child Welfare Policy and Practice." San Diego, CA & Chicago, IL. 2018.

6. Cheng, Hao-Fei, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. "How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions." In CHI Conference on Human Factors in Computing Systems (CHI ’22), April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA. 2022. https://doi.org/10.1145/3491102.3501831.

7. Children’s Data Network. "Los Angeles County Risk Stratification Model: Methodology & Implementation Report." August 2021.

8. Citron, Danielle Keats. *Technological Due Process*. 2008.

9. Cuccaro-Alamin, Stephanie, Regan Foust, Rhema Vaithianathan, Emily Putnam-

Hornstein. "Risk assessment and decision making in child protective services: Predictive risk modeling in context." *Children and Youth Service Review 79.* 2017.

10. Dare, Tim and Eileen Gambrill. "Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County." April 2017.

11. De-Arteaga, Maria, Riccardo Fogliato, and Alexandra Chouldechova. "A Case for Humans-in-the-loop: Decisions in the Presence of Erroneous Algorithmic Scores." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

12. Drake, Brett, Melissa Jonson-Reid, Maria Gandarilla Ocampo, Maria Mrrinson, Darejan Dvalishvili. "A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare*." Annals of the American Academy of Political and Social Science 692, no. 1*. November 2020.

13. Eubanks, Virginia. *Automating Inequality: How High-tech Tools Profile, Police and Punish the Poor*. St. Martin's Press. 2018.

14. Glaberson, Stephanie K. "Coding Over the Cracks: Predictive Analytics and Child Protection." *Fordham Urban Law Journal 46.* 2019.

15. Hempel, Daniel. "Uncharted Waters: Data Analytics and Child Protection in Los Angeles." *The Imprint*. (July 20, 2015). https://imprintnews.org/featured/uncharted-waters-data-analytics-and-child-protection-in-los-angeles/10867.

16. Human Rights Watch. "If I wasn't poor, I wouldn't be unfit - The Family Separation Crisis in the US Child Welfare System." 2022.

17. Illinois Department of Children and Family Services*. Family First Newsletter Issue 3.* September 2023.

18. Jackson, David, and Gary Marx. "Data mining program designed to predict child abuse proves unreliable, DCFS says." *Chicago Tribune*. December 6, 2017.

https://www.chicagotribune.com/2017/12/06/data-mining-program-designed-to-predict-child-abuse-proves-unreliable-dcfs-says/.

19. Jackson, David, Duaa Eldeib, and Gary Marx. "Director of Illinois DCFS may leave amid ethics probe." *Chicago Tribune*. May 9, 2017. https://www.chicagotribune.com/2017/05/09/director-of-illinois-dcfs-may-leave-amid-ethics-probe/.

20. Keddell, Emily. "Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice, and the Implications for Practice." *Social Sciences 2019, 8*. 2019.

21. Lappalainen, Katarina Fast. "Protecting Children from Maltreatment with the Help of Artificial Intelligence: A Promise or a Threat to Children's Rights?" *Law, AI and Digitalization.* 2022.

22. Lanier, Paul, Maria Rodriguez, Sarah Berbiest, Katherine Bryant, Ting Guan, and Adam Zolotor. "Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence-Based Child Welfare Policy." *Journal of Family Violence.* 2020.

23. Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and Integrity of Social Life*. 2010.

24. NYC Office of Technology & Innovation. "Summary of Agency Compliance Reporting of Algorithmic Tools." 2023.

25. Oh, Samuel. "An Ethical Evaluation of the Use of Predictive Risk Models in Health and Human Services: A Case Study of the Allegheny Family Screening Tool." *Journal of Politics and Society XXX, no. 1.* 2020.

26. Putnam-Hornstein, E., R. Vaithianathan, and J. McCroskey. "The Los Angeles County Risk Stratification Pilot: An Overview and One-Year Update." Children's Data Network. August 29, 2022.

27. Riely, Naomi Schaefer. "Can Big Data Help Save Abused Kids?" American Enterprise Institute. February 6, 2018. https://www.aei.org/articles/can-big-data-help-save-abused-kids/.

28. Roberts, Dorothy E. "The Only Good Poor Woman: Unconstitutional Conditions and Welfare." 1995.

29. Rittenhouse, Katherine, Emily Putnam-Hornstein, and Rhema Vaithianathan. "Algorithms, Humans and Racial Disparities in Child Protection Systems: Evidence from the Allegheny Family Screening Tool." June 2023.

30. Samant, Anjana. "Automated Decision Systems: Child Welfare Predictive Analytics Tools." ACLU Washington. July 2023.

31. Sheridan, Ed. "Town Hall drops pilot programme profiling families without their knowledge." *Hackney Citizen*. October 30, 2019. https://www.hackneycitizen.co.uk/2019/10/30/town-hall-drops-pilot-programme-profiling-families-without-their-knowledge/.

32. Title IV-E Prevention Services CLEARINGHOUSE. https://preventionservices.acf.hhs.gov/.

33. U.S. Department of Health & Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. *Child Maltreatment 2022*. 2024. https://www.acf.hhs.gov/cb/data-research/child-maltreatment.

34. Vaithianathan et al. "Allegheny Family Screening Tool: Methodology, Version 2." Center for Social Data Analytics. April 2019.

35. Whicher, Danielle, Emma Pendl-Robinson, Kyla Jones, and Allon Kalisher. "Avoiding Racial Bias in Child Welfare Agencies' Use of Predictive Risk Modeling." Assistant Secretary for Planning and Evaluation Office of Human Services Policy.

2022.

36. 이재혁. "(국감현장) 신현영, 사회보장정보원 e아동행복지원시스템 발굴지표 개선해야." *메디컬투데이.* 2021.10.19. https://mdtoday.co.kr/news/view/1065597721566349.

37. 조희연. "(단독) 부모가 애 숨기면 조사 못해, 위기아동 발굴시스템 헛바퀴." *세계일보.* 2023.11.19. https://www.segye.com/newsView/20231119508077?OutUrl=naver.