Final Report on Douglas County Decision Aid (DCDA)
Predictive Risk Modeling Randomized Control Trial Experiment

Maria D. Fitzpatrick
Cornell University and National Bureau of Economic Research

Christopher Wildeman
Duke University and ROCKWOOL Foundation Research Unit

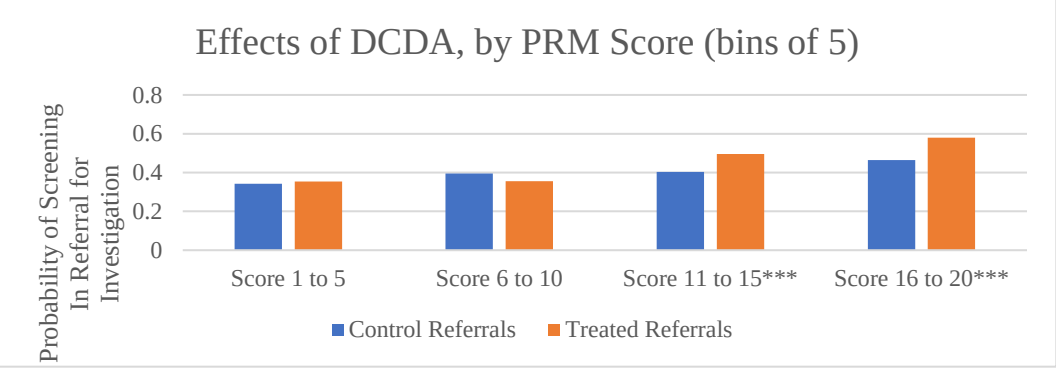
February 11, 2021
Contact info: maria.d.fitzpatrick@cornell.edu

Executive Summary

The Douglas County Decision Aid (DCDA) is a tool adopted by the Douglas County, Colorado Department of Human Services for incorporating information from a predictive risk model of the likelihood a child will be removed by Child Protective Services (CPS) within 24 months into its process for determining whether an investigation in response to alleged child maltreatment should be opened. We conducted a randomized controlled trial (RCT) of whether the availability of the DCDA altered human decisions about whether to screen-in maltreatment referrals.

As indicated in the figure below, we find clear patterns that it does, but that the pattern depends on the risk score of the referral (i.e., there are targeted effects in specific parts of the score distribution but no average effect across the entire score distribution). Without the DCDA (i.e., in the control condition), there is some relationship between predicted risk score and screen in among referrals that do not already have cases for services underway. Specifically, even without knowledge of the predicted risk score, the highest predicted risk cases without a case open are about 12 percentage points more likely than the lowest predicted risk cases without a case open to be investigated. Findings suggest that the DCDA decreases the likelihood that referrals with predicted risk below the median are investigated and increases the likelihood that referrals with higher predicted risk are investigated. The results are strongest for the highest risk referrals. For example, referrals with predicted risk scores in the top quartile are 11 percentage points (33 percent) more likely to be screened in for investigation when the DCDA is available than when it is not available. Those in the second highest quartile of the predicted risk score are 9 percentage points (27 percent) more likely to be screened in for an investigation. These results are statistically significant at the five percent level.

Effects of the Douglas County Decision Aid on Probability of Screening-in a Referral for Investigation, referrals without a case for services open on the referral date



Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and March 4, 2020 that had not had cases or investigations open within the past 60 days. The treatment of viewing the DCDA score was randomly assigned to RED Teams based on their meeting date, so we have included calendar day fixed effects for the RED Team meetings. The estimates of the effect of seeing the predicted risk score on the probability of a referral being screened in for investigation are from an instrumental variables specification where we instrumented for a score being viewed with the random assignment to a RED Team where scores were assigned to be viewed. The dependent variable is a dummy variable for an investigation opening before the end of the study period. *, **, *** indicate statistically significant estimates at the 10, 5, and 1 percent level, respectively.

Introduction and Background

Predictive risk modeling (PRM) uses complex mathematical algorithms and often-massive datasets to provide plausibly more reliable and less biased estimates of the probability some event will occur. Although PRMs will sometimes only have information on the specific system they seek to predict contact with (e.g., a PRM predicting the risk of criminal justice contact will only have data on prior criminal justice contact), some promising PRMs have used linked administrative data from a range of systems to predict the risk of experiencing a future event.

In recent years, PRM has come to be used in settings in which individuals are at risk of experiencing an adverse outcome we would be interested in preventing (e.g., incarceration, high school dropout, sexual abuse) in a context in which: (1) there are far more individuals at risk of experiencing the event than could be reasonably targeted for any given intervention (e.g., all high school students, all individuals recently released from prison); (2) there is limited information available to assess the possibility of an event; or (3) there are concerns about biased assessments of individuals on the basis of characteristics that may not be directly linked to the risk of experiencing some event (e.g., race/ethnicity, immigration status). Although PRMs have started to be applied in many fields, there has been significant interest—both positive and negative—for the application of these methods in the child welfare and criminal justice arenas.

The goal of this study was to evaluate the effect of the use of information from a PRM in Douglas County, Colorado from March 2019 through April 2020 on the probability that an allegation of maltreatment would be investigated. Because of the complex nature of the experiment, a brief overview of the steps is in order. First, a separate external research team, in collaboration with county administrators, developed an algorithm trained to predict children's risk of experiencing removal from the home within 24 months of a referral (Vaithianathan et al. 2019). The PRM scale was devised to stratify referrals into risk ventiles, ranging from a low of 1 (5% of referrals estimated to have the lowest risk of removal) to a high of 20 (5% of referrals estimated to have the highest risk of removal). It was based on linked administrative data from county and state data child welfare and public benefits systems. Second, our team worked with the county to develop an evaluation that would simultaneously provide a rigorous assessment of how the algorithm influenced child protective services (CPS) caseworkers' decision making, while also not disrupting the system. We ultimately settled on randomly varying which of the two or three evaluation teams operating on a given day received the DCDA. The teams that saw the scores saw them only at the end of the team meeting and discussion.[1] Third, we conducted analyses to estimate how the PRM score affected the decision to launch a maltreatment investigation.

Before providing a summary of the results, three caveats are in order.

First, and maybe most importantly, fallout from the Covid-19 pandemic struck Colorado in the midst of our data collection. Not only is it likely to have changed maltreatment reporting decisions, but it also likely had an effect on business processes and the ways that teams used the DCDA information. As a result, we have focused our main evaluation on the original one-year period from March 5, 2019 to March 4, 2020. We report results for the entire experiment in an

[1] More details are available in the section entitled "The Experimental Manipulation."

appendix, but given the massive shock Covid-19 presented both the CPS system and the broader nation (and world), we focus on the results from the first year of the study because it seemed the most reasonable way to deal with the portion of the data that was affected by the COVID-19 pandemic.

Second, this report is final in the sense that it includes the full period in which access to the DCDA was randomly provided to the RED teams and that it includes a complete analysis of one child welfare indicator of interest: screening-in a referral for a CPS investigation.

Third, and relatedly, this report is also preliminary in at least two ways. (a) Many of the most significant child welfare outcomes—being a victim of severe or fatal abuse, being placed in foster care, aging out of the foster care system—and many of the other relevant outcomes—school attendance, graduation from high school, juvenile justice contact, teenage pregnancy—simply cannot be assessed during the relatively short evaluation window that has passed to date. As a result, this report only provides a preliminary picture of the global effect of this intervention. (b) Many of the other system-level outcomes that we would ideally consider, such as how the availability of the DCDA scores shaped meeting length and how caseworkers and supervisors adapted to the availability of the DCDA score, have not yet been analyzed. As a result, even for outcomes we measured during the study period, this report is only preliminary.

Those caveats aside, our evaluation found support for three conclusions.

First, the availability of the DCDA score does affect the probability of launching an investigation.

Second, as expected, this effect is dependent on the score range. Point estimates suggest viewing and using the DCDA score decreased the probability of an investigation for the referrals with the predicted risk scores below the median (though not for referrals with the lowest predicted risk scores) and increased it for higher risk referrals, though the estimates are often underpowered, particularly in the full sample.

Third, in the full sample, we find evidence that the highest risk referrals are less likely to be investigated with or without the DCDA tool. In fact, in the control group of the full sample there is no relationship between the predictive risk score and investigation other than a lower investigation probability among the referrals with the highest predicted risk (those with predicted risk scores between 16 and 20). This may be because the business process makes it standard to send new referrals to the investigating caseworker when an investigation is already open. Since high-risk referrals are more likely to have a case for services open at the time of referral, their screen-in rate is lower than it would be without this process. Because of this, we also present estimates focused on the subset of referrals without a case for services open around the time of the referral (which is what we described in the previous paragraph). When we focus on this sample, the likelihood of screening in for investigation increases with the predicted risk score in the control group and the use of the DCDA decreases the likelihood of screen in for referrals with predicted risk scores below the median and increases it above the median. Specifically, referrals with predicted risk scores in the top quartile are 11 percentage points (33 percent) more likely to be screened in for investigation when the DCDA is available than when it is not

available. Those in the second highest quartile of the predicted risk score are 9 percentage points (27 percent) more likely to be screened in for an investigation. These results are statistically significant at the five percent level.

The Douglas County Decision Aid Predictive Risk Model

The DCDA is a tool that presents caseworkers with information about referrals they are evaluating for a decision about whether to start an investigation. The tool includes a score, which is based on a predictive risk model. The model uses the predicted probability of removal from the home within 24 months as the outcome of interest. It combines child welfare and case management data with social services data from the Denver metro area and nearby Larimer County to generate the prediction using LASSO regularized logistic regression. The model includes 460 predictors, which do not include any race-related measures. Researchers building the model used standard procedures for building the model on a training data set and testing its predictive power on a separate data set.

For more information about the model, see Vaithianathan et al. (2019).

The Experimental Manipulation

The goal of this study was to determine whether the use of the DCDA changed human decisions about whether to investigate a referral alleging abuse or neglect. The standard process used for 85 percent of referrals in Douglas County is through a RED (Read, Evaluate, and Direct) Team process.[2] This is a process in which a team of CPS employees reviews and discusses information about the referral with the aim of coming to a consensus decision about whether the referral should be investigated. Prior to the use of the DCDA, the information reviewed by the RED Teams in Douglas County included referral information and any additional information they accessed from a database of all previous relevant child protection system interactions and Colorado Court records.

In Douglas County, there are 3 RED Team meetings on Mondays, Wednesdays, and Fridays and 2 RED Team meetings on Tuesdays and Thursdays (13 each week). Therefore, there are about 700 RED Team meetings each year. Each RED Team is comprised of at least three members: one supervisor and two caseworkers. Each supervisor staff member from the pool of 12 sign up for seven RED Teams per month; the 32 casework staff each sign up for five RED Teams per month.

In the experiment, each RED Team is randomized to either treatment or control. We randomized such that one RED Team per day was assigned to the treatment group. This leaves either one or two control groups, depending on the day.[3]

---

[2] The remaining 15 percent of cases are either screened out because they did not meet the conditions required for assessment, they involved a Youth-in-Crisis, or they required an immediate response.

[3] We considered randomizing each referral to be in the treatment or control group. However, one concern we had in setting up this experiment was learning-by-doing on the part of caseworkers. In other words, we were worried that the caseworkers would essentially create their own mental model versions of the predictive risk model (based on the

RED Teams assigned to the treatment group have access to the DCDA score. These treatment group RED Teams first conduct their review and discussion of the prior history and current allegation with the aim of deciding whether to investigate. In this discussion, RED Team members are blind to the score. They then make an initial decision about whether or not to screen-in the referral for investigation. At the end of the discussion of each referral, the supervisor then accesses the DCDA score and shares it with the entire RED Team. The RED Team then discusses and decides whether to change the screening decision based on the score.[4]

In theory, the DCDA should have been seen and discussed by each team in the treatment group and left unseen by each team in the control group. In practice, there are a few reasons that the scores were not seen by the treatment group. In a few cases, the team either forgot to look at the score or forgot to record confirmation that they looked at the score. Also, there were technical difficulties that led the scores to be unavailable during the meeting, i.e. the "system was down". In these instances, treated teams were unable to access any scores during their meeting. Finally, the system calculated scores once each day, in the evenings. Referrals that came into CPS after that point did not have DCDA scores available for viewing by RED Teams the next day. We consider all of these to be situations where the score was unseen. This is unlikely to be a randomly selected sample of referrals (e.g., referrals that come in overnight are more likely to be high risk). For this reason, we present results from different models where we make use of the data on viewing scores and the random assignment to view scores in different ways.

Covid-19

The RCT launched March 5, 2019. It was scheduled to run for one year, with a two month follow up to observe investigation decision outcomes. At the end of February, a decision was made to extend the RCT through April 30, 2020 in order to increase the statistical power of the experiment. The decision to continue the experiment was followed by the global coronavirus pandemic, which shut down the US and led to a massive shift in maltreatment allegations and assessment processes. In the main text below, we describe results of the experiment run from March 5, 2019 to March 4, 2020. In Appendix Table 1, we present information on the results for the entire study period. There are some differences in the size of point estimates and statistical power, but overall the estimates from the full sample tell a similar story to those focused on the initial year of the study.

Validation of the Randomization Process

The goal of an RCT is to assign the treatment randomly so that the treatment and control groups are essentially the same other than the fact that one group received the treatment and the other did not. In order to verify that this is true, it is standard to examine the characteristics of the

scores they saw) as their exposure to the PRM score grew. This would have happened more quickly if scores were visible in every RED Team meeting. Additionally, we decided that a setting where all referrals in a RED Team meeting either had scores visible or did not more appropriately mirrored the comparison between two possible actual work environments where caseworkers either are or are not able to access scores for all referrals.

[4] If a RED Team decides to screen out a case with a score of 18 or higher, the case is reviewed by a supervisor.

treatment and control groups observable before treatment. If the randomization "worked" then these characteristics should be balanced across the two groups.

In Table 1, we present average characteristics of the treatment and control groups separately, as well as p-values for a t-test of whether the characteristics are the same across the two groups. Unfortunately, we have very few characteristics in the administrative data that are not possibly affected by the treatment. This is because many of the recorded observable characteristics of children and referrals are adjusted as investigations take place and as caseworkers interact with families to learn more about their identity and the situation underlying the referral. We therefore focus on the two characteristics that are fixed at the time of the referral: the predicted risk score and the child's age. We do not find a statistically significant difference between the means of these characteristics across the control and treatment groups. Moreover, any differences also appear to be small. This supports the notion that the randomization worked to create balanced comparison groups.

The average DCDA score in both the control and treatment group is just over 8. This is probably due to the fact that the DCDA score is predicted using data on all of the metro Denver area while the sample only includes Douglas County, which has a population with a slightly lower risk of future system involvement. It may also be driven by the fact that the highest risk cases in Douglas are assigned to a different process from the RED Team evaluation. The average age of children in the referral sample is 10.

Methodology for Analyzing the RCT

We now turn to analyzing the data to determine the effect of seeing the DCDA on the decision to initiate an investigation. Each observation in the data is a child-by-referral observation. (There can be multiple children on a referral and there can be multiple referrals for any child.) The relevant score for each observation in this analysis is the highest score for any child on the same referral as the child of interest on the date of interest.[5] We divide the referrals into four groups based on their score: 1 to 5, 6 to 10, 11 to 15, and 16 to 20.[6]

Different types of referrals may come in at different times of the year or on different days of the month or the week. This is the reason to randomize the treatment to RED Teams each day – to make sure the treatment and control groups are comprised of similar types of referrals. We control for the differences in types of referrals each day using fixed effects in our regression for each calendar day of the RED teams. With this statistical technique, referrals are essentially being compared to other referrals evaluated on the same day, which helps us attribute any differences that may be observed between screening decisions for treatment and control groups to the viewing of the score.[7]

[5] We assigned each child the highest score in the referral because the score in the DCDA is the highest score on the referral.

[6] In an appendix, we also report results using six roughly equally sized bins of 3 to 4 points on the score scale. These results generally confirm the pattern of results using four bins.

[7] We have also conducted analyses using referral date fixed effects. Results are very similar.

As discussed above, for various reasons, some referrals in the RED Teams randomly assigned to the treatment group may not have had scores available or the RED Team did not view them. To account for the potential bias this may cause in our estimates of the effect of the DCDA on the decision to screen-in a referral for investigation, we present estimates from three different models, each of which relies on different assumptions and has a different interpretation.

First, we present comparisons of the probability of investigation for referrals where scores were actually seen and discussed versus those that were not. The equation we estimate is the following:

$$Y_{ird} = \beta_0 + \beta_1 Score6to10_{ird} + \beta_2 Score11to15_{ird} + \beta_3 Score16to20_{ird} + \beta_4 Score\ seenXScore1to5_{ird} + \beta_5 Score\ seenXScore6to10_{ird} + \beta_6 Score\ seenXScore11to15_{ird} + \beta_7 Score\ seenXScore16to20_{ird} + \delta_d + \varepsilon_{ird}$$

In the equation, the outcome, Y, is a dummy variable measuring whether the decision was made to investigate the referral *r* for child *i* made on date *d*.[8] Each of the variables starting with "Score" indicates a score category as described. The omitted score category is those between 1 and 5. The variables starting with "Score seen" are variables measuring whether a score was viewed by the relevant RED Team, interacted with the score as indicated. Betas 1, 2, and 3 measure whether higher score referrals in the control group are more likely to be investigated than those with scores ranging from 1 to 9. Betas 4, 5, 6, and 7 measure whether seeing the score made it more likely that referrals within each score group were investigated.

The comparison of outcomes in this empirical specification is between referrals where predicted risk scores were seen versus those where predicted risk scores were not seen. Interpreting this as the causal effect of seeing the predicted risk score would require assuming that other differences in these two groups did not exist. However, as discussed above, because the process that provided scores did not provide them for referrals that came in overnight, it is possible that referrals with scores were different from the others. Therefore, although we present estimates from this model, we prefer estimates from the other two models below.

In the second model, we make use of the randomization to compare the probability of investigation for referrals that were assigned to RED Teams that were randomly assigned to have access to and view the DCDA scores or to not view them. Here, the estimation equation is the following:

$$Y_{ird} = \beta_0 + \beta_1 Score6to10_{ird} + \beta_2 Score11to15_{ird} + \beta_3 Score16to20_{ird} + \beta_4 TreatedXScore1to5_{ird} + \beta_5 TreatedXScore6to10_{ird} + \beta_6 TreatedXScore11to15_{ird} + \beta_7 TreatedXScore16to20_{ird} + \delta_d + \varepsilon_{ird}.$$

[8] In the main results presented here, we use all decisions to investigate that occurred by June 30, 2020. Results are similar if we limit the decision to investigate to the 60 days after the referral. This indicates that there is no shift in the timing of investigations due to use of the score, at least not in the somewhat long-run horizon of over 60 days. In future work, we will examine whether investigation timing shifted forward within the initial 60 days.

The comparison is between a set of referrals in the control group – where predicted risk scores were not seen – and those in the treatment group – where some scores were seen and some were not seen. Because the treatment was randomized, the assumptions underlying the interpretation of this as the effect of having a score available are extremely likely to have been met, as confirmed by our investigation that the randomization worked to create similar groups of referrals. These estimates can be considered as an "intent-to-treat" estimate. In other settings, similar computer and process constraints would prevent scores from being seen for each referral, as was the case here. Given that, results from this specification have the nice property that they can speak directly to what would happen if we were to adopt a similar process more widely in Douglas CPS or other CPS agencies outside of the specific setting of our RCT.

In our third and final specification, we make use of the fact that there is exogenous variation in which referrals had predicted risk scores viewed and which did not. This exogenous variation is driven by our randomization of the access of scores to RED Teams. We therefore use an instrumental variable specification to estimate effects of seeing the scores for those referrals where scores were seen due to their assignment to a treated RED Team. In other words, we estimate the effect of the "treatment on the treated." In this specification, we instrument for the interaction terms between the score categories and the measure of whether a score was seen with the interaction terms between the score categories and whether a referral was in the treatment group. If our random assignment worked, and the seeing of scores is strongly determined by the assignment to a treatment RED Team, then we can interpret the coefficient estimates from these analyses as the effect of seeing a predicted risk score on the investigation decision for the group of referrals whose scores are seen due to the random assignment. These are our preferred estimates and the ones we spend the most time summarizing below and that we focus on in the Executive Summary.

Results

We now turn to the results of the RCT. In Table 2, we present estimates of the coefficients on each score category dummy variable, as well as on each variable measuring the interaction between score category and whether a referral was either seen (columns 1, 2, 5, and 6) or in the treatment group (columns 3 and 4).

As mentioned in the introduction, when RED Teams review referrals to determine whether they should be screened-in for investigation, one of the things they consider is whether an investigation or case is already open. If one is already open, rather than open another investigation, it is typical for a RED Team to alert the caseworker investigating the original referral or working with the family on services about the existence of a new referral. For this reason, we present estimates of the effects of seeing the DCDA on the decision to open an investigation for the sample of clients without a case for services open within 60 days before the referral (column 1) and for the whole sample (column 2). For each of the other model specifications, we similarly present two different coefficients, one for each of these samples.

From the first three rows of the first column, we can see that in the group of referrals without an existing case for services open, when the score was not viewed there is a positive relationship between the predicted risk score and the probability of screening in for investigation. This

makes sense given that the predictive risk model includes some of the same information that caseworkers are using to evaluate the decision to screen in a referral when the DCDA is not available. However, in the sample of all referrals (column 2), the estimates in the first three rows suggest that all referrals where the score is not viewed are similarly likely to be screened in except those with the highest predicted risk scores (those in the 16 to 20 range), which are 4.6 percentage points less likely to be investigated. That this pattern is so different across the samples suggests that it is driven by the tendency to not screen in referrals where there is already a case open or investigation underway. A similar pattern exists in the remaining columns of the table for the referrals in the control group or for which scores were not seen.

The estimates in the next eight rows of Table 2 help us understand the effects of the RCT. The coefficient estimates of Table 2 suggest that viewing the DCDA tool did not have a statistically significant effect on the decision to investigate among the referrals with scores between 1 and 5, nor did being assigned to the treatment group. The point estimate suggests the DCDA decreased the probability of opening an investigation in the lowest scoring group by 0 to 1 percentage points (between 0 and 3 percent), but the point estimate is not statistically significant at conventional levels. However, having access to the DCDA and/or viewing it decreases the likelihood of screening in cases that the predictive risk model estimates are between the 26th and 50th percentile of the predicted risk distribution. Depending on the specification, this decrease is somewhere between 3 and 27 percent. It is only statistically significant at conventional levels in some specifications.

In the upper part of the predicted risk score distribution, the DCDA increases the likelihood of opening an investigation among the cases with predicted risk scores, but the size of the estimates varies and the estimates are not statistically significant at conventional levels. Among referrals in the 51st to 75th percentiles of the predicted risk score distribution, seeing the score increased the probability of being screened in for investigation by 9.1 percentage points (27 percent). The intent-to-treat estimate of being assigned to see the predicted risk score is 7.6 percentage points (23 percent). Similarly, among referrals with predicted risk scores between 16 and 20 that did not already have cases or investigations open, using the DCDA increased the likelihood of opening an investigation by 11.4 percentage points (33 percent). This treatment-on-the-treated estimate is statistically significant at the five percent level. The intent-to-treat estimate of just having the DCDA available increased the probability of opening an investigation for these referrals with high predicted risk scores by 10.3 percentage points, or 31 percent.

Conclusion

The main finding is that making the DCDA available to caseworkers assessing whether to screen-in a referral for investigation has the expected effect on the probability of screening in across the distribution of predicted risk. Access to the DCDA during the team meeting to evaluate the referral for screening in likely decreased the likelihood of screen in for referrals with predicted risk scores below the median and increased the likelihood of screening in for referrals with predicted risk scores above the median. For a graphical representation of these estimates, see Figure 1. For referrals at the highest predicted risk of removal, seeing the DCDA changed the likelihood of screen in by 33 percent, a large and meaningful impact. Even just adopting a

business practice where the score was available increased the probability of investigating the highest predicted risk referrals by 31 percent.

One area we would like to examine is whether the DCDA availability altered the screen in decisions differentially by race and ethnicity of children on the referral. Although the DCDA does not explicitly use race as a factor, it is important to understand the relationship between race and the other measures that contribute to the model and to understand if the use of the predictive risk score had any differential effects on children of different races. Because race is missing from about 30 percent of referrals, we are working to get a better understanding of what the data can tell us about this important dimension of the DCDA's effects.

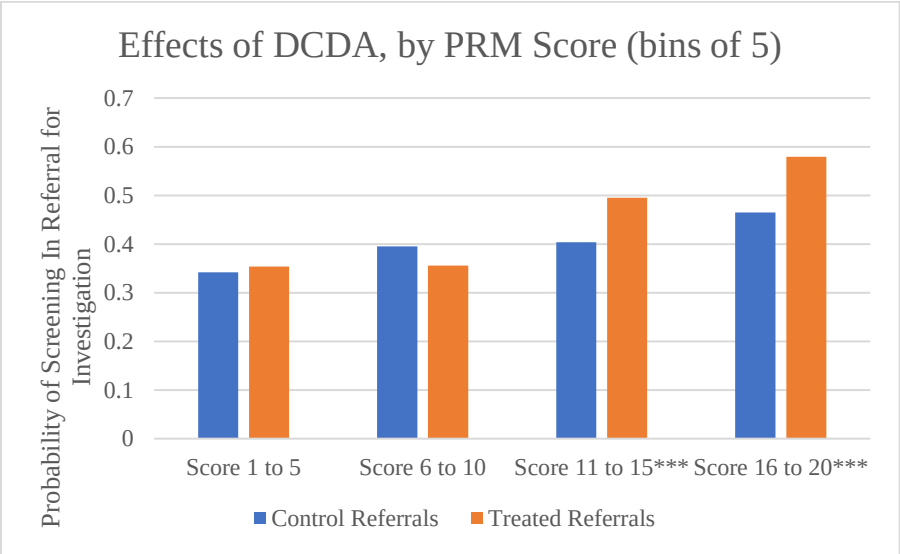
We will continue to analyze the data so as to understand the results from this study. We will analyze data we collected from RED Teams on the length of meetings, length of discussion of cases, and recorded content of a sample of cases to determine how the availability of the DCDA changed the RED Team process or discussion. Also, we will soon reach the one-year mark of the end of the RCT, which will be a natural time to begin looking at whether DCDA availability had any effect on removal from the home or placement in foster care.

References

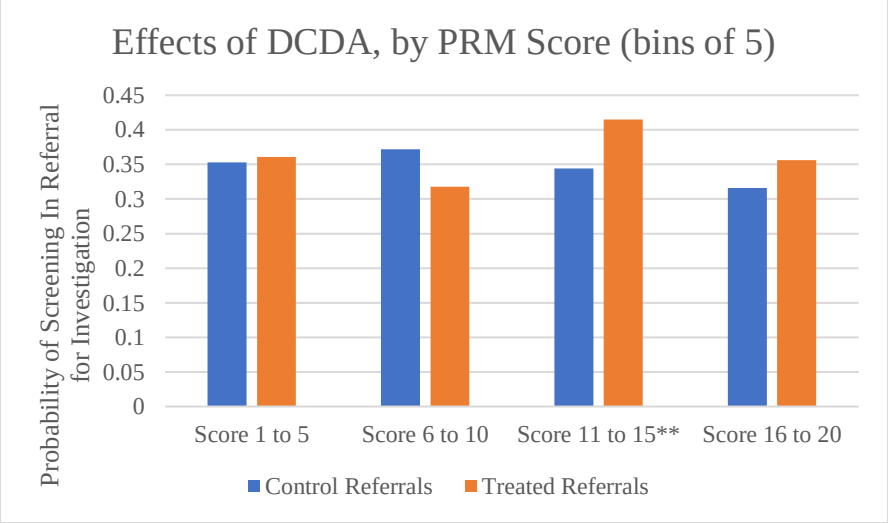
Vaithianathan, Rhema, Haley Dinh, Allon Kalisher, Chamari Kithulgoda, Emily Kulick, Megh Mayur, Athena Ning and Diana Benavides Prado. 2019. “Implementing a Child Welfare Decision Aide in Douglas County: Methodology Report.” https://csda.aut.ac.nz/__data/assets/pdf_file/0009/347715/Douglas-County-Methodology_Final_3_02_2020.pdf

Figure 1. Effects of the Douglas County Decision Aid on Probability of Opening and Investigation

Panel A. Referrals without cases for services open on the referral date



Panel B. All referrals



Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and March 4, 2020. The treatment of viewing the DCDA score was randomly assigned to RED Teams based on their meeting date, so we have included calendar day fixed effects for the RED Team meetings. The estimates of the effect of seeing the predicted risk score on the probability of a referral being screened in for investigation are from an instrumental variables specification where we instrumented for a score being viewed with the random assignment to a RED Team where scores were assigned to be viewed. The dependent variable is a dummy variable for an investigation opening before the end of the study period. *, **, *** indicate statistically significant estimates at the 10, 5, and 1 percent level, respectively.

Table 1. Baseline Characteristics by Treatment

	Control	Treatment	p-value
DCDA Score	8.27	8.20	0.70
Age	10.10	10.09	0.88
Number of Observations	2,549	3,203	

Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and March 4, 2020. The treatment of viewing the DCDA score was randomly assigned to RED Teams. Mean characteristics of observations at the child by referral level are presented in the table. The last column contains p-values for a test of the sample means being identical.

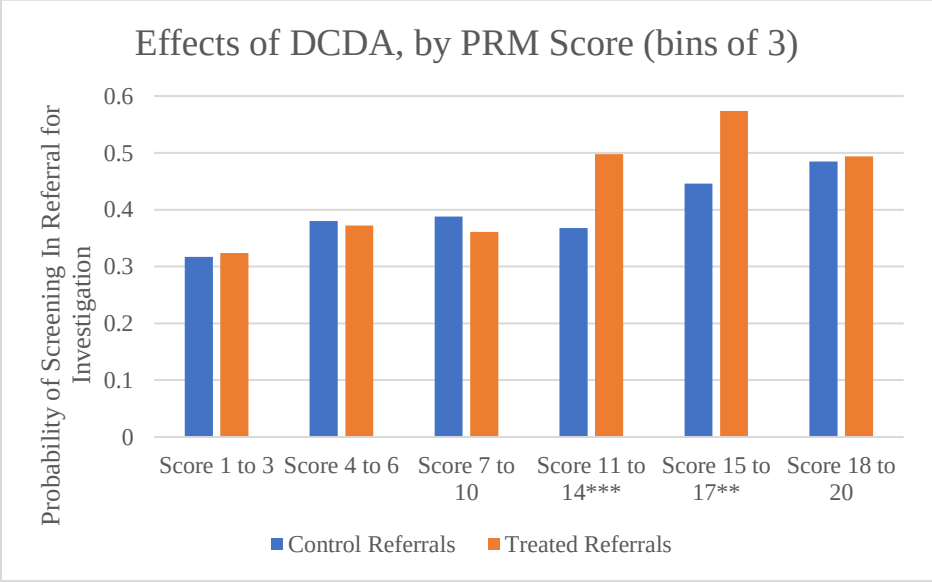
Table 2. Effects of Douglas County Decision Aide on the Decision to Investigate Child Maltreatment Referrals

	(1)	(2)	(3)	(4)	(5)	(6)
	(OLS)		(ITT)		(TOT)	
Score between 6 and 10	0.071***	0.031	0.054*	0.020	0.053*	0.019
	(0.026)	(0.024)	(0.029)	(0.027)	(0.029)	(0.027)
Score between 11 and 15	0.083***	-0.003	0.062**	-0.008	0.062**	-0.009
	(0.028)	(0.025)	(0.031)	(0.028)	(0.031)	(0.028)
Score between 16 and 20	0.128***	-0.046**	0.123***	-0.037	0.123***	-0.037
	(0.033)	(0.023)	(0.035)	(0.025)	(0.035)	(0.025)
Score seen X Score between 1 and 5	0.001	-0.002			0.012	0.008
	(0.024)	(0.023)			(0.028)	(0.026)
Score seen X Score between 6 and 10	-0.091***	-0.090***			-0.039	-0.054
	(0.033)	(0.029)			(0.040)	(0.037)
Score seen X Score between 11 and 15	0.036	0.050*			0.091**	0.071**
	(0.034)	(0.029)			(0.041)	(0.035)
Score seen X Score between 16 and 20	0.099**	0.050*			0.114**	0.040
	(0.040)	(0.026)			(0.045)	(0.030)
Treated X Score between 1 and 5			0.010	0.007		
			(0.024)	(0.023)		
Treated X Score between 6 and 10			-0.031	-0.042		
			(0.032)	(0.029)		
Treated X Score between 11 and 15			0.076**	0.059**		
			(0.034)	(0.029)		
Treated X Score between 16 and 20			0.103**	0.035		
			(0.041)	(0.026)		
Constant	0.335***	0.347***	0.330***	0.343***	0.330***	0.343***
	(0.016)	(0.015)	(0.017)	(0.017)	(0.017)	(0.017)
Observations	4,278	5,752	4,278	5,752	4,278	5,752
R-squared	0.020	0.004	0.020	0.003		
Open cases included?	NO	YES	NO	YES	NO	YES

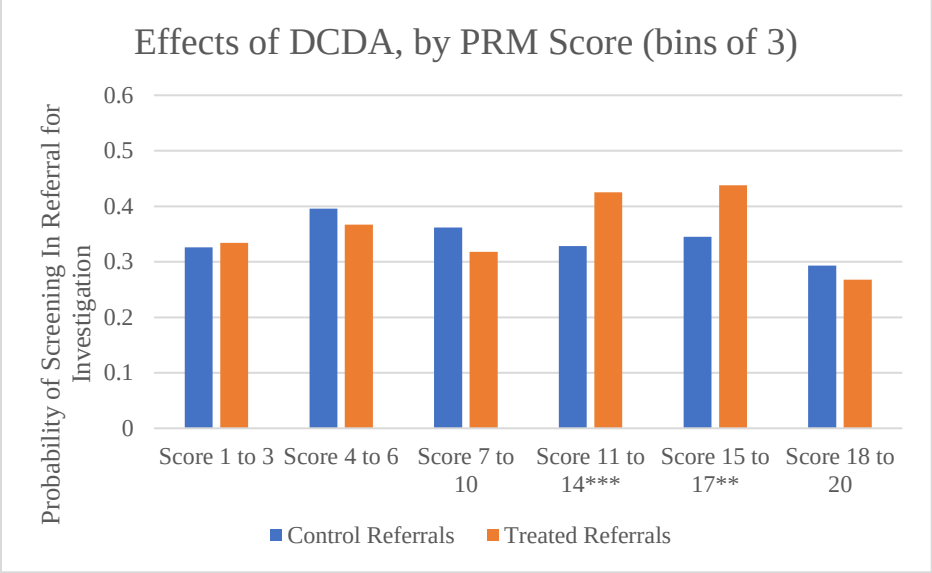
Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and March 4, 2020. The treatment of viewing the DCDA score was randomly assigned to RED Teams based on their meeting date, so we have included calendar day fixed effects for the RED Team meetings. The dependent variable is a dummy variable for an investigation opening before the end of the study period. *, **, *** indicate statistically significant estimates at the 10, 5, and 1 percent level, respectively.

Appendix Figure A1. Effects of the Douglas County Decision Aid on Probability of Opening and Investigation

Panel A. Referrals without cases for services open on the referral date



Panel B. All referrals



Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and March 4, 2020. The treatment of viewing the DCDA score was randomly assigned to RED Teams based on their meeting date, so we have included calendar day fixed effects for the RED Team meetings. The estimates of the effect of seeing the predicted risk score on the probability of a referral being screened in for investigation are from an instrumental variables specification where we instrumented for a score being viewed with the random assignment to a RED Team where scores were assigned to be viewed. The dependent variable is a dummy variable for an investigation opening before the end of the study period. *, **, *** indicate statistically significant estimates at the 10, 5, and 1 percent level, respectively.

Appendix Table A1. Effects of Douglas County Decision Aide on the Decision to Investigate Child Maltreatment Referrals, through April 30, 2020

	(1)	(2)	(3)	(4)	(5)	(6)
	(OLS)		(ITT)		(TOT)	
Score between 6 and 10	0.067***	0.021	0.052*	0.009	0.050*	0.008
	(0.025)	(0.023)	(0.027)	(0.025)	(0.027)	(0.025)
Score between 11 and 15	0.102***	0.012	0.087***	0.010	0.087***	0.010
	(0.027)	(0.023)	(0.029)	(0.026)	(0.029)	(0.026)
Score between 16 and 20	0.127***	-0.033	0.124***	-0.028	0.123***	-0.028
	(0.031)	(0.022)	(0.032)	(0.023)	(0.032)	(0.023)
Score seen X Score between 1 and 5	0.025	0.019			0.036	0.029
	(0.023)	(0.022)			(0.026)	(0.025)
Score seen X Score between 6 and 10	-0.085***	-0.079***			-0.035	-0.040
	(0.031)	(0.028)			(0.037)	(0.034)
Score seen X Score between 11 and 15	0.010	0.033			0.054	0.048
	(0.033)	(0.027)			(0.039)	(0.033)
Score seen X Score between 16 and 20	0.108***	0.043*			0.122***	0.040
	(0.039)	(0.025)			(0.043)	(0.029)
Treated X Score between 1 and 5			0.032	0.025		
			(0.022)	(0.022)		
Treated X Score between 6 and 10			-0.028	-0.031		
			(0.030)	(0.027)		
Treated X Score between 11 and 15			0.045	0.040		
			(0.033)	(0.027)		
Treated X Score between 16 and 20			0.111***	0.035		
			(0.039)	(0.025)		
Constant	0.326***	0.341***	0.321***	0.336***	0.321***	0.336***
	(0.015)	(0.014)	(0.016)	(0.015)	(0.016)	(0.015)
Observations	4,774	6,390	4,774	6,390	4,774	6,390
R-squared	0.019	0.003	0.018	0.002		
Open cases included?	NO	YES	NO	YES	NO	YES

Note: Based on the authors' calculations using administrative data from Douglas County. Sample includes referrals assigned to RED Teams between March 5, 2019 and April 30, 2020. The treatment of viewing the DCDA score was randomly assigned to RED Teams based on their meeting date, so we have included calendar day fixed effects for the RED Team meetings. The dependent variable is a dummy variable for an investigation opening before the end of the study period. *, **, *** indicate statistically significant estimates at the 10, 5, and 1 percent level, respectively.