# Examining the Effects of Antidiscrimination Laws on Child Welfare: Law on the Books

Netta Barak-Corren, Yoav Kan-Tor, and Nelson Tebbe[[1]](#footnote-2)\*

# Introduction

Conflicts between religious freedom and civil rights law are persisting in the United States. Equality protection for women and LGBTQ+ people has run up against traditional religious commitments concerning contraception and marriage, to take only two examples.[[2]](#footnote-3) Such disputes have affected various areas of civil rights law, including public accommodations statutes that protect customers against discrimination,[[3]](#footnote-4) employment discrimination laws that shield workers,[[4]](#footnote-5) and rules that require government officials to serve everyone equally.[[5]](#footnote-6) .

One specific area in which this tension has manifested recently is child welfare. In virtually all jurisdictions in the United States, governments place children in foster care and adoption with the assistance of nongovernmental child welfare agencies. Governments contract with these agencies to match children with qualified foster homes or adoptive parents. Many of these nongovernmental child welfare agencies are run by religious organizations, and some of the organizations have beliefs that prevent them from placing children with certain kinds of families--including those headed by same-sex couples. Because some governments have laws or policies that require child welfare agencies to refrain from discrimination on the basis of sexual orientation, conflicts arise.

The most salient example came in the 2021 United States Supreme Court decision in *Fulton v. City of Philadelphia*.[[6]](#footnote-7) There, the Court confronted a claim by Catholic Social Services, a religious child welfare agency that refused to certify same-sex couples as foster parents. Because the government required agencies to serve same-sex couples equally, it terminated its contract with the agency. The Supreme Court ruled for the agency, but on grounds that were arguably specific to that dispute. It focused on a peculiar provision of Philadelphia’s contract with the agency that allowed the city to grant individualized exemptions to its nondiscrimination rule, and it held that the city could not make exemptions available to others without also granting an exemption to the religious agency.

Because the Supreme Court did not decide the general question of whether religious child welfare agencies are entitled to exemptions from antidiscrimination rules in all cases, the issue is sure to return to the Court in the future. Moreover, several state legislatures have enacted laws that grant religious child welfare agencies exemptions from antidiscrimination rules that exist on the state and local levels and other jurisdictions are considering enacting such laws.[[7]](#footnote-8) So the issue has not been resolved, but continues to be of pressing importance.

Important to the legal question, and to the policy issue, is how deciding one way or another is likely to affect children. If children are likely to be harmed when agencies exclude certain classes of prospective parents, then the government will have an easier time showing that it has a “compelling state interest” in enforcing equality laws.[[8]](#footnote-9) Conversely, if children are harmed when antidiscrimination rules are enforced uniformly, that will strengthen the arguments of religious agencies seeking exemptions from those rules.[[9]](#footnote-10)

In real cases, including *Fulton*, objecting religious agencies have been claiming that children will be harmed if antidiscrimination laws are applied to them. The argument is that if agencies are not granted exemptions, they will close their doors rather than violate their religious beliefs. And if agencies close--or shut down their child placement operations--that is likely to negatively affect children: fewer families will be available--beds will go empty--because some families will only work with agencies that share their religious beliefs or identity, and because social workers will stop serving children in need.[[10]](#footnote-11) These effects will worsen the shortage of foster care and adoptive homes for needy children.

These are extremely concerning arguments about empirical facts and they matter greatly for the resolution of these important conflicts. Unfortunately, however, there is little existing empirical evidence to support the claims of either side. Governments and religious actors have cited evidence, but it is almost always anecdotal, descriptive, or limited in geographic scope.[[11]](#footnote-12) No nationwide study has been conducted using reliable empirical methods on the outcomes for children as a result of legal changes in civil rights laws affecting child welfare agencies.

Yet it ought to be possible to conduct such a study because many governments have enacted antidiscrimination rules for child welfare agencies, and several have enacted exemptions for religious agencies. In fact, over the past 20-some years, the majority of states gradually enacted laws in this area. Moreover, disputes have erupted, several of which have resulted in agencies shutting down their child placement operations. The earliest and most prominent dispute concerned Catholic Charities of Boston, which shocked many when it closed down its child placement operation in 2006 rather than comply with a requirement to serve same-sex couples that followed after Massachusetts became the first state to recognize same-sex civil marraige. But Boston was not alone--since then, several other disputes have arisen, and some of them have resulted in agency closures.

Against this background, we set out to research the question of how children are affected by antidiscrimination rules for child welfare agencies. Our study is important for at least two reasons. First and most obviously, answering the empirical question correctly is necessary for resolving the legal question in a manner that is most just. However the balance between religious freedom and equality law is struck by courts and legislatures, that should be done on the basis of real evidence about the effects on children, whose welfare should be at the center of the debate.

Second, establishing a reliable empirical foundation can help to lower the temperature of constitutional and political battles within the culture wars (Barak-Corren 2020, 2021). Values are central to those debates, of course, and so are legal rules. But facts can help to bring the parties closer together by establishing a common basis of evidence against which those questions can be debated (Barak-Corren 2021). For example, if it turns out that antidiscrimination laws harm children, even the most committed egalitarians will think twice before insisting that they be enforced uniformly. Conversely, if the evidence shows that equality laws benefit children, then even the most committed religious traditionalists will pause before insisting on accommodations for child welfare agencies. Not every aspect of these difficult questions can be resolved in this way, but some can -- or so we hope.

Our approach to the question of whether children are harmed by antidiscirmination laws for child welfare agencies has two parts. The first part, which is presented in this Article, looks at the “law on the books” and asks how antidiscrimination laws and religious exemptions at the state level affect outcomes for children in both foster care and adoption. Here, we look only at written rules and only at the state level. That cannot tell us everything about the effects of equality law on child welfare, but it can provide a nationwide view on whether and how legal changes in this area influence children.

In the second part, which we develop in a separate article, we examine the “law in action.” In particular, we study places where disputes have in fact broken out between religious child welfare agencies and governments, including but not limited to places where agencies have in fact closed down their child placement operations. We look at the effects on children in those areas, not only by examining the quantitative data but also by interviewing child welfare professionals who were involved in each conflict on the ground.

In this Article, our main finding is that state antidiscrimination rules generally are not associated with harm to children. For most children, such laws have no effect in one direction or the other. However, on average, antidiscrimination laws have a positive effect on children’s success in finding both foster homes and permanent homes, and antidiscrimination laws also reduce the average length of time it takes to place children in both foster and permanent homes. We also find that the effects vary among subgroups, such that often children who are hardest to place benefit from larger positive effects (heterogeneous treatment effects). In particular, we find that the positive effects of antidiscrimination laws increase for older children, who are often the most challenging to place with families. We also find positive effects for other vulnerable groups.

Alongside these positive effects, we find that state antidiscrimination laws have mixed effects in some states, for example increasing permanent home placements yet also increasing the time to achieve such placements. While these mixed effects pertain to a minority of the compared states and most effects are consistently positive, this heterogeneity provides important grounds for additional research in this area. Notably, we find that it is too early at this point to estimate the effect of religious exemption laws, as they were mostly enacted very recently and overall apply to fractions of the data (2-3% of the children). We plan to delve into this question in future studies, data permitting.

This Article unfolds in four parts. In Part I, we briefly review the limited literature on the question. Then in Part II we describe and defend our methodology for studying the nationwide evidence of child outcomes. In brief, we deploy a machine learning, boosted causal forest approach to estimate the effect of antidiscrimination rules on certain measures of child outcomes for both foster care and adoption. To do this, we pair an original dataset of equality laws and religious exemptions at the state level with the most comprehensive existing dataset on child welfare and additional socio-economic data. Third, we present our results, including robustness checks and exploratory analyses. Finally, we discuss the significance of our findings for ongoing legal and political disputes concerning these important issues. A brief conclusion points toward further work, including our study of “law in action” and other planned research.

# Current Knowledge and State of the Debate

Little literature exists on the subject that we study here. On the one hand, some authors have suggested that children are harmed when religious child welfare agencies close rather than comply with antidiscrimination rules. Stephanie Barclay, a law professor who was involved in the *Fulton* litigation in the lower courts while she was working at Becket, has written that “it is difficult to see how forcing faith-based agencies to close their program does anything to provide more options to LGBTQ parents, as opposed to just providing fewer options to other families.”[[12]](#footnote-13) She tells the story of Cecelia Paul, a longtime foster parent and a plaintiff in the *Fulton* case, who was honored by Philadelphia as one of its foster parents of the year in 2015. Quoting a *Wall Street Journal* editorial, she reports that “because Mrs. Paul is certified through Catholic Social Services, her home has been vacant since April.”[[13]](#footnote-14) Another opinion article quoted by Barclay reports that on a typical day the religious agency in *Fulton* serves 120 foster children and speculates “that’s a lot of slack for other agencies to pick up” should the religious agency be forced to close.[[14]](#footnote-15) Some argued that religious agencies particularly excel in recruiting and retaining foster parents, based on surveys of foster parents that mentioned religious support as important in their process.

Barclay also describes the situation of Sharonell Fulton, the lead plaintiff in the *Fulton* case and a foster mother in Philadelphia, who wrote in another opinion piece “that her special needs children would be taken from her if Catholic Social Services had to close its program.”[[15]](#footnote-16) She also quotes lawyers for the Michigan Department of Health and Human Services as saying in a legal brief that “if faith-based agencies are not allowed to operate according to their religious principles, they will shut down, which can have the effect of reducing the number of available families. Such a result will do nothing to help a single child find a home.”[[16]](#footnote-17) These assertions were not backed with evidence and are, in any event, specific to certain localities.

On the other hand, supporters of equality requirements for child welfare agencies are claiming that such laws do not harm children. Eleven states filed a brief in the Fulton case explaining that “our experience shows that nondiscrimination requirements are fully consistent with maintaining sufficient private organizations with the expertise to provide foster care services.”[[17]](#footnote-18) The states claim that if religious agencies were allowed to be exempted from antidiscrimination rules, “[t]he result would be multifaceted harms to children.”[[18]](#footnote-19) They tell the story of how a religious agency in Michigan refused to place a child with a sibling because the sibling lived with a same-sex couple.[[19]](#footnote-20) Here too, the evidence cited is anecdotal and case specific.

Another amicus brief in *Fulton* reported evidence that same-sex couples serve as foster and adoptive parents at higher rates.[[20]](#footnote-21) Evidence for that proposition is supported by empirical surveys of foster parents, as data on sexual orientation of parents or children is not formally available today. The brief also notes that there is a shortage of foster care homes nationwide, so that more children are placed in group homes. Excluding families headed by same-sex couples therefore “risks” reducing the number of available foster families.[[21]](#footnote-22) However, the brief cites no study establishing that children would be harmed because they cannot be placed with those families by other agencies in any given locality.[[22]](#footnote-23)

In sum, to our knowledge, the literature contains many claims regarding the empirical effects of antidiscrimination rules on child welfare but no reliable, nationwide studies. What the literature does firmly establish is the need for a study like ours.

# Methodology

Our study consists of three stages. In the first stage, we compile a novel legal database of all developments in state regulation of antidiscrimination (AD) and religious exemptions (RE) in child welfare law.[[23]](#footnote-24) Second, we link the legal database with a national dataset on children in foster care to analyze the relationship between antidiscrimination regulations and outcomes for children. Our analytical strategy is counterfactual inference using machine learning methods, a powerful algorithmic method to infer the causal effect of the legislation on adoption and foster home outcomes. We build a prediction model, evaluate its performance, infer the causal effect of AD laws, and test and validate this effect in various ways.

## A database of AD laws in foster care and adoption

The landscape of AD laws in foster care and adoption has changed markedly over the years as more and more states introduced laws protecting prospective parents against discrimination on the basis of sexual orientation, followed by a counter-trend of religious exemptions. We collected data on the exact date by which each state introduced an AD or RE rule and whether that rule was applied to foster care, adoption, or both domains.

The dataset was constructed in several stages. First, we searched WestLaw and Lexis for rules mentioning sexual orientation, equality, fairness, discrimination, and foster care or adoption, and similar/variant terms. We aimed to find all state, county, and city AD rules that apply to prospective parents.[[24]](#footnote-25) We were cognizant of the fact that some of these rules would be public accommodation laws,[[25]](#footnote-26) or combinations of a general nondiscrimination policy and a licensing requirement.[[26]](#footnote-27) We have also recorded exemptions granted by state law to religious agencies or other entities. We verified the search with known cases where conflicts erupted due to the existence of such rules (e.g. Massachusetts, Illinois, New York, D.C.), confirming that the search returned the relevant laws of these jurisdictions.

Second, we searched for all legal disputes that involved foster care or adoption and sexual orientation, in all judicial instances. This search yielded both information on statutes and rules involved in litigation (thereby validating and at times expanding the first stage results) and information on court-led legal changes (e.g., when the prohibition on discrimination on the basis of sexual orientation was the result of a challenge of a state ban on gay adoptions). It was not rare for a state to have more than one relevant legal development in this area (for example, a court decision legalizing adoption by same-sex couples, followed by a legislative or administrative act that formalizes the decision). We recorded all such developments.

Finally, we consulted lists of AD laws compiled by the nonprofit organization Movement Advancement Project[[27]](#footnote-28) and resolved ambiguities through consultation and with advice from family law experts.

For each jurisdiction, we recorded the date and the effective date of the law, the source of law (legislation, regulation, court decision), the type of law (child welfare law, public accommodation law, licensing requirement), the citation, the verbatim rule, available exemptions (religious or other), and the verbatim exemption. For court decisions, we also recorded intermediary relief, legal outcome, date of outcome, appeals, and overall status of the case (pending/closed). If a jurisdiction experienced several legal changes, we tracked them and determined which of these changes effectively led to the imposition of AD obligations on foster care and adoption agencies, including citations confirming this analysis.

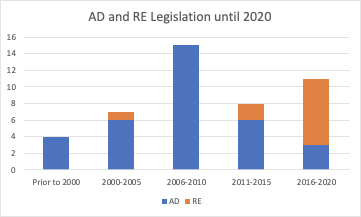
Our final dataset, available in the Online Appendix, contains for each state its AD status, effective date of the AD rule, RE status, effective date of RE rule, and citations. Notably, a few states distinguish between adoption and foster care in their laws, requiring separate coding for adoption laws (AD\_A) and foster care laws (AD\_F). The 2020 breakdown of the 50 states and the District of Columbia is detailed in Table 1:

**Table 1. Summary of State-Level AD and RE Adoption and Foster Care Laws**

|  |  |  |
| --- | --- | --- |
| **Type of Law** | **Adoption** | **Foster Care** |
| AD + RE | 3 | 3 |
| AD only | 31 | 32 |
| RE only | 8 | 8 |
| None | 9 | 8 |

Unfortunately, we have not been able to find reliable data on rules promulgated by county or municipal governments, as they are not tracked and recorded in legal databases or by nonprofit organizations, and may come in the form of a contract provision or licensing requirement. However, this limitation appears to be relevant for only a few states, because in most states the child welfare system is fully centralized and therefore all regulation, including equality provisions and religious exemptions, is enacted at the state level.[[28]](#footnote-29)

Figure 1 presents the yearly trends in the enactment of AD and RE laws regarding foster care and adoption.



We matched our state-level legislation data with child outcome data such that each child (See below) was assigned the applicable legal status during their time in foster care. ​​

## Data on Children in Foster Care

We linked the legal dataset with data on foster care and adoptions from the Adoption and Foster Care Analysis and Reporting System ([AFCARS](https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/afcars)). This is a national dataset, spearheaded by the federal government, and available freely for researchers, containing yearly information from 2000[[29]](#footnote-30) on children entering and existing the welfare system in each county.[[30]](#footnote-31) We obtained the data on all states from the earliest available year to the most recent, 2000-2019, consisting of 14,480,299 entries for 6,055,467 unique child cases.[[31]](#footnote-32)

The AFCARS dataset is a “big” dataset, with massive number of observations and high dimensionality due to the rich description on each child, including length of time in foster care, current placement (e.g., foster home, group home), reason for removal from home (e.g., neglect, drug abusing parent), reason for discharge from foster care (e.g., adoption, reunification with family, aging out), and quite expansive demographic information on the child, parents of origin, and adoptive parents (where relevant), including race, sex, age, disability status, and more.[[32]](#footnote-33) Each child may spend several years in foster care, resulting in several data years per child that add to the high dimensionality of the data. This format allows researchers to follow each case as it develops from entry to exit.

The dataset also has limitations. The submission of AFCARS data is required by all Title IV-E agencies and collected semi-annually,[[33]](#footnote-34) but states are not required to collect information on sexual orientation or religious affiliation (not for adults, nor for children) and do not report information on private agencies that may be in charge of foster care or adoption services. Notwithstanding these limitations, AFCARS is currently the most comprehensive dataset on children in foster care and it is used in many studies in this field (e.g. Sieger 2020).

Our focus in this project is on two types of positive outcomes: children’s success in finding an adoption and in finding a home placement while in foster care. We draw on AFCARS guides and documentation in the analysis of the data.

### Foster Care data

The foster care data consists of all ~6 million child cases reported in the data from 2000 to 2019 and documents the entire duration of each child in the system.

Our main outcome of interest is derived from the variable Current Placement Setting (CurPlSet) that records the type of setting in which the child currently lives, for each year of care. There are three home setting options (pre-adoptive; a relative home; a non-relative home) and three congregate care options (group home, typically from seven to twelve children; an institution, caring for more than 12 children; and supervised independent living for older children). Another type of setting is trial home visit, where the child returns to the principal caretaker under state supervision for a specified period of time. Finally, the child may be documented as a runaway if they independently depart from their foster care setting.

Our main analysis focuses on children’s success in finding a foster home versus being placed in congregate care while in care (Child’s BureauL 2014). Success is defined as cases that were placed in a home setting for their entire duration in care (69% of the cases) or experienced a stable positive transition from institute to home during this time (3%). Failure is defined as cases that spent their entire duration in care in congregate care (13%) or experienced a stable negative transition from home to congregate care (2%). We omit cases that were in a trial home visit status (~6%) for their entire duration in care, because in these cases the system did not intend to find an alternative home setting for the child. We also remove runaway cases (~3%), as they are hard to classify,[[34]](#footnote-35) and cases with missing (~1%) or erroneous placement information (~1% of the cases), and cases that experienced both positive and negative transitions (e.g. moved from home to institution to home, or vice versa; about 2%) as these cases defy simple categorization into success/failure in finding a home placement. Figure F1 plots the distribution of overall foster care placement.

In addition, we examine the time it took to place children in a home setting (computing the gap between the date when the child was last removed from home, LatRemDt, and the begin date for the first home placement, CurSetDt).

|  |
| --- |
|  |
| **Figure F2**: **The distribution of foster care placements**. A candidate is considered in “home” if they were placed in home settings for their entire duration in care (curplacment codes 1,2,3; an intermediate trial home visit period, 8, does not impact the classification), “inst” if they were placed in congregate care settings for their entire duration in care (curplacment codes 4,5,6; an intermediate trial home visit period does not impact the classification), “trial only” if they were only in 8 code, runaway (had a 7 code), “home\_to\_inst” if they had a transition from an institute code to home code (classification unaffected by trial home periods), “inst\_to\_home” if they had a transition from a home code to inst code (classification unaffected by trial home periods), missing (had a 99 code), err (other errors). |

*Sixty* percent of foster care cases were administered in the absence of an AD rule (AD=0 during all of their years in care), ~32.5% were administered under an AD rule (AD=1 during all of their years in care) and 7.5% of the cases transitioned from 0 to 1 during their time in care. *Ninety-four point five* percent of foster cases were administered in the absence of an RE rule (RE=0 during all of their years in care), 2% were administered under an RE rule (RE=1 during all of their years in care) and 3.5% of the cases transitioned from 0 to 1 during their time in care.

The Online Appendix details the list of original AFCARS fields, derived fields, and external merged data.

### Adoption data

To examine children’s adoption outcomes, we focus on children who are defined as waiting for adoption (hereinafter, “adoption cases”).[[35]](#footnote-36) Adoption cases comprise 22% of the children (1,322,500 cases) who were waiting for adoption in at least one of their years in the FC system. 97.5% of adoption cases had a single ‘adoption cycle,’ namely a single uninterrupted sequence of years during which they were candidates for adoption. For children with more than one cycle, we focused on the outcomes of the final cycle and created a dummy variable to account for the existence of previous cycles (mul\_waiting).[[36]](#footnote-37) 6.5% of the adoption cases were omitted because of missing information regarding their first year(s) as adoption candidates.[[37]](#footnote-38)

Our main outcome of interest is derived from the variable Discharge Reason (Disreasn) that documents the reason for which a child is released from foster care. This may include adoption (61.5%), permanent legal guardianship (3%), reunification with parents or primary caretakers (5%), living with other relatives (1%), emancipation (the child reached majority according to the law by virtue of age, marriage, etc.; 6% of the cases), transfer to another agency (where responsibility was transferred either in or outside of the state or Tribal service area), runaway, and death (less than 1% each), and open cases (22.5%).

|  |
| --- |
|  |
| **Figure 3A**: The prevalence of the discharge reason at the terminal adoption case year. |

We define success as finding a permanent home, i.e. being discharged due to adoption, guardianship, living with relatives, or reunification (70%).[[38]](#footnote-39) Failure is defined as emancipation, runaway, or death (6%). Transfers to another agency and open cases are omitted from the results, as in both types of cases there is no information on the ultimate adoption outcome (23% of cases). Overall, we have 946,642 cases which we can use for the analysis.

We also examine adoption cases’ time to permanent placement by examining their Total Days Stay in Foster Care in All Episodes (LifeLOS), as the literature strongly indicates that the less time children remain in care, the better their overall outcomes (Lockwood et al 2015; Testa 2004; Child’s Bureau 2014; Stott and Gustavsson 2010).

|  |
| --- |
|  |
| **Figure 3B**: The distribution of adoption cases with no adoption outcome (terminal disreasn =4,7,8), Adoption outcome (terminal disreasn = 1,2,3,4,5) and unknown outcome (terminal disreasn 99, typically open cases, or 6, transfer to another agency). |

Fifty-six percent of adoption cases were administered in the absence of an AD rule (AD=0 during all of their years in care), 37% were administered under an AD rule (AD=1 during all of their years in care) and 6.5 of the cases transitioned from 0 to 1 during their time in care.

Ninety-two percent of adoption cases were administered in the absence of an RE rule (RE=0 during all of their years in care), 3% were administered under an RE rule (RE=1 during all of their years in care) and 4.3% of the cases mixed transitioned from 0 to 1 during their time in care.

Due to our finding that religious exemptions laws apply to a tiny fraction of the data (2% of the foster care cases and 3% of adoption cases), the rest of our analyses will focus on the impact of AD laws, leaving the important task of studying RE law effects for the future, when more data accumulates on children in these jurisdictions.

We provide a detailed description of the adoption cases data in the Online Appendix.

## Strategy of analysis

### Choosing among analytical approaches

A large share of empirical work about policy questions relies on observational data. Drawing inference about the causal effect of a law from observational data is inherently challenging and the present case is no exception. To assess the effects of AD legislation on child outcomes, we need to be able to disentangle causation from correlation, despite the absence of a randomly controlled trial and the likely presence of any number of unobserved factors.

Researchers use a wide variety of strategies for attempting to draw causal inference from observational data in the presence of confounding factors, common among which are differences-in-differences methods and synthetic control methods. In this paper, we draw on recent advances in machine learning and the field of counterfactual causal inference in computer science, not before we explain why this method has multiple important advantages in the present case.

Difference-in-differences methods have been the go-to tool for studies of policy changes since Card’s classic study of the effects of low-skilled migration on the labor market (1990). These methods are typically used when some cities or states experience a treatment, such as a policy or legal change, while others do not. As legal change is not applied at random, and outcomes are not necessarily the same across cities/states in the absence of such change, the challenge for causal inference is to come up with a credible estimate of what the outcomes would have been for the treated group in the absence of the legal treatment (a counterfactual analysis). To do that, researchers need to assume, and then validate, that the pre-change trends were parallel between treated and control groups. But this assumption does not always hold, hindering the ability to conduct a counterfactual analysis.

Synthetic control methods provide an elegant improvement: instead of choosing a single or a set of control units, researchers use a weighted average of the set of controls, such that the synthetic average of the pre-trends is more similar to the treated group than any single control unit would be. The weights are then carried over to the post-change period to simulate the outcomes for the synthetic control (Abadie, Diamond, and Hainmueller 2010). Yet the fundamental challenge of both approaches remains settings with many covariates that interact with treatment outcomes. Such settings create heterogeneous treatment effects that differences-in-differences and synthetic control methods are not designed to estimate (Athey and Imbens 2017)

Machine learning (ML) methods can improve the performance of causal analysis, particularly in high-dimensional settings with a large number of covariates (Athey and Imbens 2017), as in the present case (see data description above). Originally, supervised machine learning methods focused on prediction problems--estimating a model on a subset of the data (the training set) and using it for predicting outcomes in the remaining data (the test set). The focus is typically on the accuracy of the model, without regard to the implications for causal inference (a classic example in law is The Supreme Court Forecasting Project, Ruger, Kim, Martin and Quinn 2004) and without focus on the relative role of any specific feature used in the model (Mullainathan and Spiess 2017; Ruger et al. 2004; Berk, Sorenson and Barnes 2016; Grogger et al. 2021).

This began to change as researchers started using machine learning methods in observational studies involving “big data,” to help control for the large number of covariates in a flexible manner (Wager and Athey 2015; Athey and Imbens 2016). In such settings, machine learning methods can help in matching treatment and control units across a broad range of features and without overtesting the data, an exceedingly difficult task using the classic regression-based methods (Grimmer 2014).

Causal random forests are a particularly promising advancement in this field, as random forests are known to perform very well in practice for prediction problems (Athey and Imbens 2017). Causal forests are able to estimate effects at the level of each individual and generate estimates of treatment effects that change more smoothly with covariates (Wager and Athey 2015; Athey and Imbens 2016; Kunzel 2019). Baiardi and Naghi (2021) revisit five influential papers that span a variety of topics in applied economics, and find differences both in the size of the treatment effect estimates and in statistical significance. They find that causal ML methods are more robust to potential nonlinear confounders than classic approaches. In addition, because ML methods can handle many covariates potentially responsible for effect heterogeneity in a systematic way, it is less likely that heterogeneous effects will be missed, compared to manual modeling. Furthermore, causal ML methods facilitate systematic model selection by comparing a wide range of alternative model specifications. The result is a data-driven model that may keep a smaller set of influential confounding variables from among a large set of potential controls, making the analysis both less discretionary and more powerful. Similarly, Wager and Athey (2016), Athey and Imbens (2019), Kunzel et al. (2019) and Louizos et al. (2017) compare ML methods to classic econometric approaches and find that the ML methods are substantially more powerful, especially in the presence of many covariates, and typically perform better than classic regressions, reducing errors in the estimation of both average and heterogeneous treatment effects.

The present case presents a massive number of foster care and adoption cases and a high-dimensional description of each case, including many child specific covariates in addition to potentially important external covariates at the state and country level (e.g. GDP, state funding, local characteristics of the social welfare system). This setting calls for using ML methods to select and control for covariates in a data-driven way, provide accurate estimates of average treatment effects, and expose and estimate potential heterogeneous effects of AD laws, to the extent that they exist. In the next section, we explain in detail how we use ML methods for causal inference.

### Causal Inference Using Machine Learning Methods

We estimate the effects of AD law per individual case by training a machine learning algorithm (“the learner” or “model”). As common in ML methods, we first divide the data at random into a training set (80%) and a test set (20%). In the training stage, the learner is trained on part of the dataset (“the training set”) to predict the outcomes of individual cases. In the inference stage, we use the learner to simulate the outcome of each case under AD and no-AD legislation (a counterfactual analysis). The difference in case outcomes between the two legal statuses is termed “lift,” or the “legal lift.” We validate the model and the legal lift on the hold-out data that was not previously accessible to the learner (“the test set”). In the final stage, we create a retrospective experiment to examine the predictive value of the learner on the test data, creating a comparison between cases with similar predicted legal lifts under different actual legal states. We explain each stage in the following sections.

#### Training a prediction model using gradient boosted decision trees

To build the model, we assume that each adoption candidate has a *home potential* (HP) -- the chance to find a foster home, or for an adoption candidate to find a permanent home -- a potential that can be viewed as a number between 0 (no potential) to 1 (full potential).

To predict the HP, the data should include three elements: (1) the outcome -- whether the case found a permanent home or not; (2) the treatment -- the legal status of the case (AD/no AD); (3) the full feature vector -- i.e., the variables describing each case. Including these variables in the model is important for two reasons: first, to produce an accurate adoption prediction, and second, to control for the effect of confounding variables, i.e., variables that are correlated with both the adoption outcome and the legal treatment (Louizos et al. 2017). The output of the model is a number between 0 and 1, indicating the HP of each case.

For training, we use an S-learner Catboost based classifier (Kunzel et al. 2019), a prediction model which is a gradient boosted ensemble of decision trees (a “forest”) with native categorical support. Boosted trees are a computational advancement in the random forest literature, which greatly improves the efficiency and accuracy of the prediction model (Kunzel et al. 2019; Kaufman, Craft, and Sen 2019).

Each tree is constructed using the available features (variables) on each individual case, including age, race, disability, time in the system, reason for removal from home, and so on. The algorithm builds each tree by selecting a feature and a threshold value that minimizes a loss function, e.g. that partitions the data best between adopted and not adopted cases. This means that for age, for example, the algorithm will choose the cut-off age under which most cases are adopted and over which most cases are not adopted). Each partition is partitioned again based on a newly-drawn case feature and threshold, until the tree reaches a partition number limit (“height”), or saturation. The final partitioned set is considered a “leaf” and receives a score based on the ratio of adopted cases in this set.

Additional trees are constructed to achieve a “boosted” adoption prediction, such that the algorithm trains on the residual error of the prediction of the previous trees, making each new tree focused on minimizing the errors of the previous tree. To avoid overfitting, the Catboost algorithm randomly selects a different data set for each iteration and calculates the residuals by using trees that are not trained on the current example. As each tree is constructed on a new, randomly sampled, set of the data, the trees differ in their set of examples and in their prediction goals (in the construction process, the residuals error values change from tree to tree).

Once the algorithm exhausts the creation of trees it calculates the HP score of each individual case across all trees. The mean score is returned as the prediction score of that case. Hence, each case receives an individual HP score. Once training is complete, the learner can be tested on the test data to examine whether the learner can accurately predict the actual outcomes of cases that it was not previously exposed to.

Catboost has a multitude of meta-parameters which we considered, including the number of trees, depth and learning rate. To maintain generality and not contaminate the test set, we selected the best values using a five-fold cross validation. In this method, the training set is split into five folds. Five models are then trained using 4-folds as the training set and the remaining fold for validation assessment. Using grid search we explored a multitude of values for the meta parameters. The top performing values, as indicated by the empirical results, were used to train a final model that was evaluated using the test set.

#### Inferring and validating the causal effect of legislation

In the inference stage, the model is used to understand what effect, if any, can be attributed to the AD legislation. We are interested in the difference in HPthat can be attributed to the legislation, defined as the *legal lift (****LL****),*

Whereis the HP probability function (the catboost model in our case). LL then indicates the increase or decrease in HP under legislation (AD = 1).

Inferring the LL builds on the fact that changes in case description can lead each case into a different leaf. We provide the S-learner with the case description and the Catboost classifier and simulate the case’s AP score under two different legal statuses: with and without AD law. The model returns the difference per case between the two statuses, namely the LL.

This inference is possible under three assumptions (Neal 2020):

(1) *Stable Unit Treatment Value Assumption* (eq1): that an individual’s treatment outcome is simply a function of the individual’s treatment. SUTVA is a common assumption in all causal inference exercises, which includes two stability concerns: no interference, i.e that the outcome for case *i* is unaffected by the treatment of other cases; and consistency, i.e. that the treatment is the same for all treated cases (Neal 2020; Hernán and Robins 2020).

(2) *Unconfoundedness* (Conditional exchangeability) (eq2): that the decision to apply the treatment (=legal status) is unconfounded by the characteristics of children, namely that the decision whether to legislate is not dependent on the characteristics of any individual child.

(3) *Overlap* (eq3) - that treated and untreated cases overlap in other characteristics (i.e. have non-legislation features in common), such that they can be compared.

|  |  |
| --- | --- |
| Equation 1: | [[39]](#footnote-40) |
| Equation 2: |  |
| Equation 3: |  |

Meeting the assumptions depends on relationships in the data and typically requires both adequate design and subsequent validation.

In terms of design, there is an inherent tradeoff between unconfoundedness and overlap: for unconfoundedness, one needs to control for all variables that may be confounded with the treatment and with the outcome. But the more variables we include in each case, the harder it is to find two cases that are exactly comparable (overlap). There is a similar tradeoff between model complexity and case overlap. In complex forest-based models, the trees enrich the case description with compound variables (e.g., the interaction effect of age below the threshold and drug abusing parents) on the expense of finding greater overlap between cases, by using simpler descriptions (Neal 2020).

Therefore, optimizing causal inference by design requires the broadest set of the data within which cases *overlap* in non-legislation characteristics (such that all treated cases in the dataset will have non-treated counterexamples that they can be compared with), while including as many *potentially* *confounding* variables (such that all variables associated with both legislation and case outcomes are controlled for). In other words, as we show in Section 4.1.1, model optimization requires to strike a balance between limiting the pool of cases and limiting the scope of case description.

In terms of data validation, we establish case overlap empirically using propensity scores (Online Appendix Section 2.2) and run various analyses to examine unconfoundedness (Sections 4.5 and 4.6 below). As for SUTVA and no interference, we first note that the model draws on cases from all states and it is highly unlikely that the adoption outcomes of child *i* in one state are influenced by the effect of legislation in *another* state on children in *that* state. We also find that cases very rarely migrate between states (Figure 3A), removing the possibility that cases travel after legal change. Therefore, the assumption of no interference appears reasonable for between-states case comparisons. To examine and account for potential interference within states, we included in the description of each case not only the case’s individual features, but also county-level means of all features (e.g. the mean of white children in the county, of drug abusing parents, and so on). We find that these features do not substantially advance the prediction of child outcomes (the contribution is close to zero, see Figures 4.4A,B), making it unlikely that the outcomes of child *i* are meaningfully influenced by the effect of legal change on other cases in the county (let alone the state).

Establishing consistency is somewhat trickier. The purpose of all treatments -- AD laws -- is the same, and they are often phrased similarly, but no law is exactly identical to all others, as rules are implemented in different states by different means and against different legal backgrounds. Therefore, we should not expect full consistency. Nonetheless, we design an analysis that equates the comparison field for 13 states that legislated their AD laws at about the same time and present the results of the model in each state. We find that the model produces similar results in most states, supporting the consistency assumption.

Finally, we validate the inferred LL by testing the learner on the test set -- the hold-out data that was not previously accessible to the learner -- and assessing the overall quality of the predicting model. A classifier with perfect prediction power can be seen as one that can account for each property of the adoption case and each modification of case properties (i.e. the counterfactual legal status of a case). The more precise the model in predicting case outcomes, the better we can simulate the outcomes under different legal treatments and assess the effect of legislation on case outcomes.

#### Testing Real-World Effects in a Retrospective Experiment

To provide further validation for the results and understand their realworld implications, in the final stage we use the test set as a retrospective experiment, to observe whether the predictions of the model are borne out by the data.

In this test, we compare the outcomes of cases for which the model predicts a large effect of the AD rule (above the 80th percentile of LL values) in *the presence and absence of such a rule*. This comparison is possible because the LL can be computed for all cases based on their description, regardless of whether they were, in fact, administered under an AD law or not.

Our prediction is that no-AD cases with a high predicted LL(i.e. cases for whom the effect of legislation could have been large, but were not administered under an AD rule) will display a lower *actual* success rate when compared to high LL cases administered under an AD law. In other words, if the model is accurate in its assessment of the effect of legislation, we should be able to observe a real difference in the outcomes of pertinent AD and no-AD cases. As we show below, the retrospective experiment finds such differences.

### Limitations

Several limitations qualify our analysis. First, recall that we are studying only “law on the books” in this Article. That means that we are limiting ourselves to looking at explicit law, not actual controversies. That said, we include here not only statutes, but also administrative regulations and judicial decisions that constitute the law of the state on antidiscrimination and religious exemptions. Our subsequent work on “law in action” will compensate in meaningful part for this first limitation, since it will examine jurisdictions that experienced actual conflicts.

Second, we look only at state-level rules in this Article, not county or municipal rules. This is a significant limitation because local governments can play a meaningful role in child welfare, and they may include nondiscrimination (or religious exemption) rules in local ordinances, administrative rules, or even in their contracts with agencies. Yet we found that the variability and unreported nature of these rules made them virtually impossible to study systematically.

Still, because state law overrides any local rules to the contrary, and because the intersection of religious liberty and child welfare has had high salience on the state level, we believe that this limitation may be less significant than it appears. In states with antidiscrimination rules, localities cannot craft religious exemptions from those rules. Conversely, in states with religious exemptions, localities cannot refuse to accommodate religious child welfare agencies despite any countervailing local equality policy. In sum, local rules have maximum efficacy only in states with no statewide antidiscrimination or religious exemption laws--which turns out to be only 9 states for adoption and only 8 for foster care. Even in those states, it is unclear that local governments can enact antidiscrimination or religious exemption rules, as the foster care system is often centralized at the state level.[[40]](#footnote-41)

A third limitation regards the overlap requirement. For the s-learner model to be reliable, all subgroups of the data with different covariates should have some probability of receiving any value of treatment. Therefore, the training set for the model should include all cases that have a similar counter-legal example in the data set, and exclude cases with no counter-examples (Neal 2020). This means that we have no ability to estimate a causal effect of AD laws on the subgroup of excluded cases.[[41]](#footnote-42) However, because we divide our data to a training set and test set randomly and exclude cases only from the training set, validating the model in the test set should indicate whether the model is robust to the inclusion of cases without counterexamples.

# 4. Results

For the sake of brevity, we present our results from the adoption and foster home analyses together, as the procedure applied to the analysis of both questions was the same and the results are mostly consistent. We note differences relating to the structure of the data or the dependent variables along the way.

## 4.1 Optimizing Model Training and Model Selection

We began by weighing the training data, as our exploratory analyses of both adoption and foster home outcome variables indicated that the dataset was heavily imbalanced: Only ~8% of adoption cases do not find a permanent home placement and only about 15% of all cases do not find a foster home placement (hereinafter ‘negative cases’). Age worsens the imbalance (*see* Figure: 4.1). This poses a challenge when using counterfactual methods, as the data set is skewed for young ages; to overcome this challenge, we introduced weighting and assigned higher weights to advanced ages. We weighted each class by its counter frequency (negative adoption cases were set to 0.92 points whereas positive cases were set to 0.08 points; the same procedure, with the necessary changes, was applied to foster home cases). In addition, each case was weighed by the counter frequency in the corresponding age group (such that young negative cases will be valued the same as young positive cases). The distribution of ages was the same within each legal status (Figure OA2).

|  |
| --- |
|  |
| **Figure 4.1. Adoption rate and age**. |
|  |
| **Figure 4.1. Home foster care rate and age**. |

To optimize model selection, we fitted 36 different models to the weighted data by creating different combinations of model complexity (set of meta-parameters) and case features vector.

First, we used three levels of model complexity -- low, medium, and high (Table 1 defines each level by the number and depth of trees). Second, each level of model complexity was implemented on several combinations of features (see Table 2). In terms of AFCARS data, the two feature vector options were (a) a subset of the AFCARS features selected to avoid correlation with AD\_A[[42]](#footnote-43) (left column) or (b) the full AFCARS features vector (the additional variables appear in the middle column). The models could use, in addition to the AFCARS data, a large set of socio-economic features at the *state* or at the *region* level, including employment rate, GDP, welfare support per capita, government support for nonprofit organizations, and more (right column). Third, models could use county-level means of AFCARS features or not.

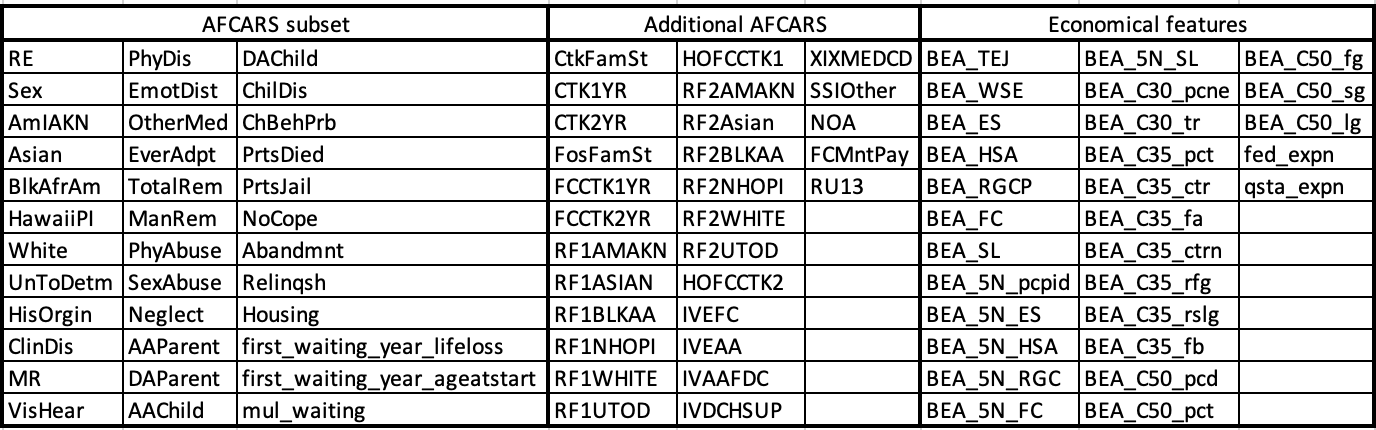
Table 3 provides the list of all 36 models (3 levels of complexity \* 2 AFCARS options \* 2 county means options (yes/no) \* 3 socio-economic options (state/regional/none) = 36 models).

Each dataset (adoption and foster home) was divided into a 20% test set which is on hold-out until the end of our analysis, and an 80% train set, which was further divided into 5-folds (numbered 0-to-4). Each model was trained on one of the four training folds within the training set and was tested on the fifth fold -- the test fold within the training set. Parameter search was performed using grid-search and 5-fold cross-validation. This process simulates additional test-sets and allows the algorithm designer to adapt the model to the problem without contaminating the test-set assessment of performance, guaranteeing that the model would not overfit the data.

**Table 1**: The different options for level of algorithmic complexity

|  |  |  |
| --- | --- | --- |
| Model complexity | Number of trees | Tree depth |
| Low | 5 | 2 |
| Medium | 5 | 4 |
| High | 20 | 8 |

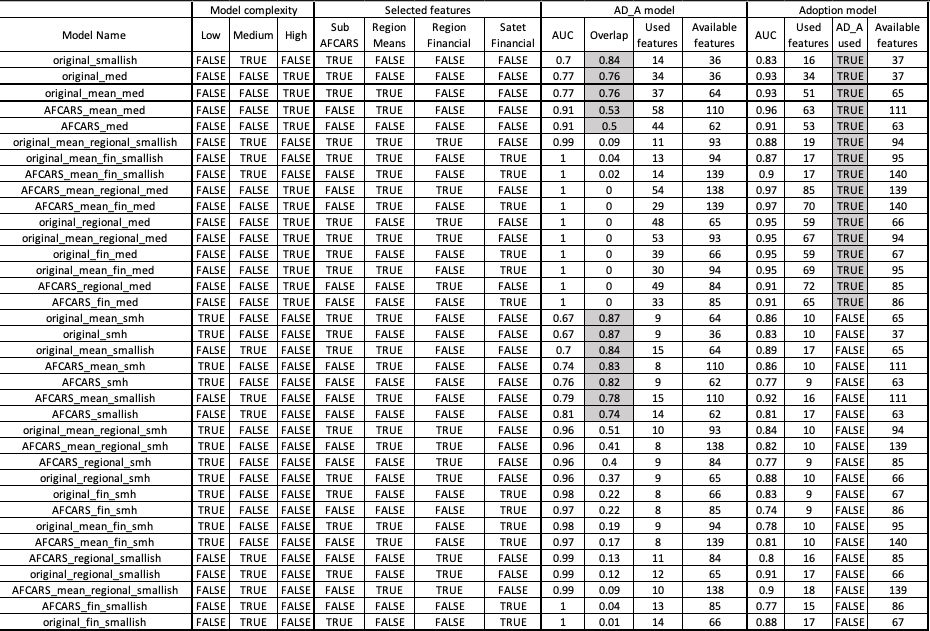
**Table 2**: The different options for case features vector



*Table note*: Models that received the “Additional AFCARS” features received both the AFCARS subset and the additional AFCARS features; models that received the “Economical features” features also received the latter two sets of features.

For the adoption prediction, only five of the 36 models satisfied (1) an overlap of at least 50% of the cases (2) actively using the AD\_A feature (an applicable antidiscrimination law pertaining to adoptive parents).[[43]](#footnote-44) All of these models are complex models. All of these models have strong predictive ability of the adoption potential (high AUC). None of these models use economic features, primarily because these features are highly predictive of AD laws and therefore including them in the model would have eliminated the overlap of cases (see Table 3).

**Table 3**: A summary of the parameters and performance of all 36 models



Among the five models, we chose the model that maximized the tradeoff between overlap (76%)[[44]](#footnote-45) and unconfoundedness (the number of available features for the Adoption model, right column, N=65) while also providing high predictability of adoption outcomes (AUC=0.93). This model includes a subset of the AFCARS variables in addition to the county level means of these variables. It is a medium complexity model.

For the foster home prediction, three models provided sufficient overlap and used the AD\_F feature, hence we selected to proceed with the “original\_mean\_med” model as in the adoption model. The selected model provided a legal overlap of 75%, had access to the same set of features, N=64, and although its AUC was lower than the adoption model AUC (0.85), it was still high.

It is important to note that both the adoption and foster home models exclude the state feature from the case description, such that information on the state where a case resides is not available to the model. This was crucial to maintain the integrity of the counterfactual simulation, as keeping state in the model would have substantially increased the number of cases excluded from analysis. This is because cases from states that did not experience transition during the studied timeframe (e.g., Massachusetts that enacted its AD law before 2000, or Kentucky that has never enacted an AD law) would not have counter-legal matches, even if such matches exist outside the state. For example, if a child in Massachusetts is similar in all respects to a child in Connecticut or Rhode Island, except for their legal status, the inclusion of the state feature would have deprived the two cases of being compared. Importantly, excluding the state feature was possible because it did not contribute to predicting case outcomes (.94 AUC versus .93). In other words, with all other features in the model, the identity of the state in which the child resides does not provide additional important information about the chances of finding home placement. As such, including this feature introduced costs -- removing relevant cases from the analysis -- without benefits -- without adding predictive value. We return to evaluate the effects of AD laws at the state level in our robustness checks section.

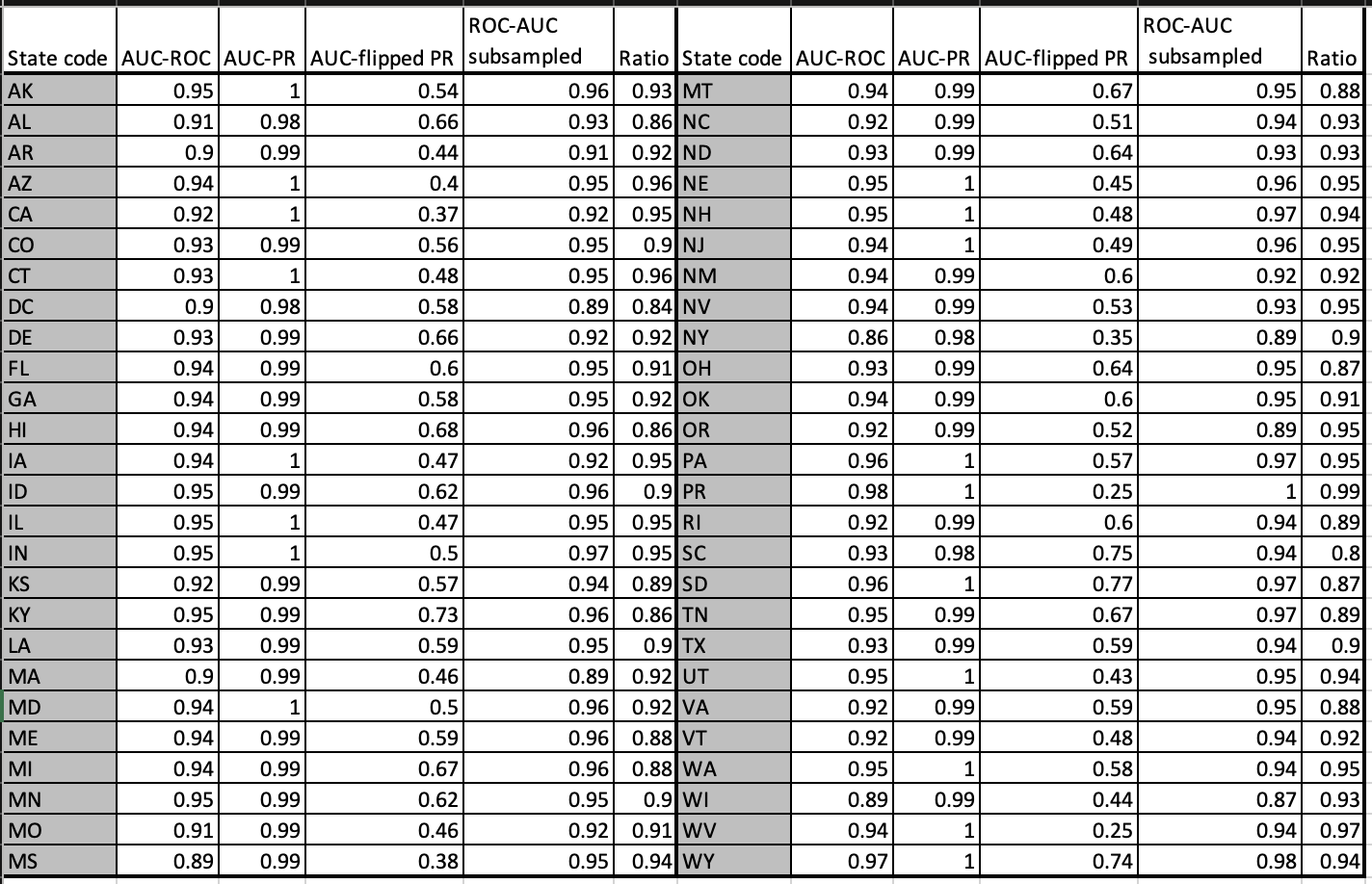
## 4.2 Model Performance

In this section we evaluate the performance of our models -- how well do they predict the chances of adoption cases and foster care cases to find a permanent/foster home placement?

Figure 4.3A presents the results of the adoption model after it was trained on the entire training set and evaluated on the entire test set. The prediction quality of the model was consistently very high in both the train set and the test set, in all five training folds, for different states (Average AUC-ROC = 0.932, ranging from 0.86 in NY to 0.97 in WY), and for different data years (Average AUC-ROC = 0.946).[[45]](#footnote-46) Tables 4 and 5 provide detailed breakdown of model performance by state and year, including additional classification matrices.

|  |
| --- |
|  |
| **Figure 4.3A:** **The Performance of the Adoption Potential Classifier.** AUC-ROC are presented for the test and the train set (top left), for the five folds (top right), for all states, (bottom left) and for each first waiting year, 2004-2016 (bottom right). |

**Table 4. Model Performance and Adoption Rate (True/Total ratio) by State**



**Table 5. Model Performance and Adoption Rate (True/Total ratio) by Year, 2000-2018**

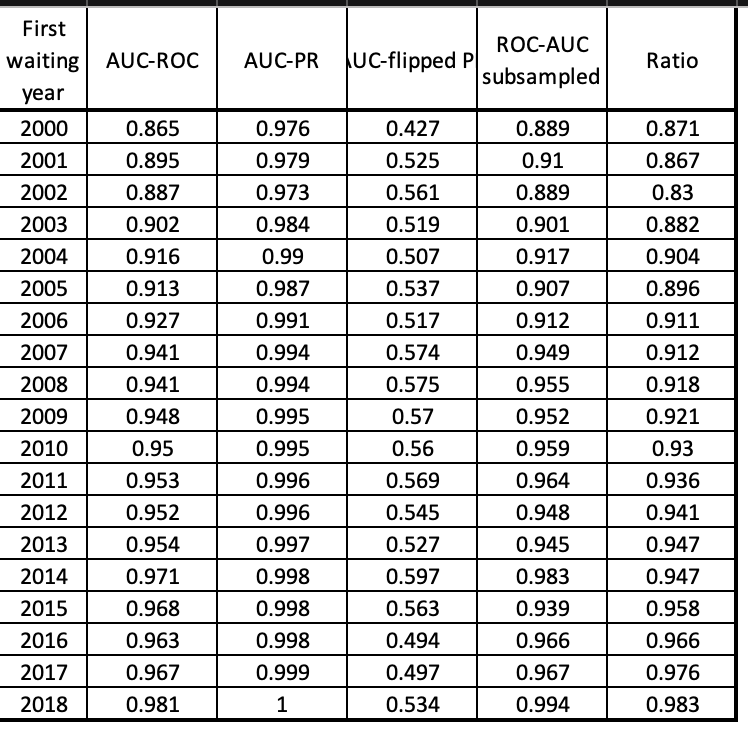


Figure 4.3B presents the performance of the foster home model that is somewhat more noisy, but still high by comparative standards (overall AUC = 0.87; mean state AUC = 0.86;[[46]](#footnote-47) mean tear AUC = 0.874; compare, e.g., with Ruger et al. 2004). Tables OA2 and OA3 present data by state and year on model performance and Foster Home Rate (True/Total ratio).

|  |
| --- |
|  |

In terms of feature importance, we find that the most important feature for predicting adoption outcomes is the child’s age at the first waiting year. Age is similarly important for predicting success in finding a foster home. For adoption outcomes, additional important figures are time in foster care, being diagnosed with emotional distress, drug abusing parents, and being black. For foster home placement, additional important features are being removed from home due to a behavioral problem, being previously adopted, and a clinical diagnosis of disability.

It is important to note that feature importance does not mean that the feature is indispensable for prediction. It is a common finding in machine learning studies that features are interchangeable with other features due to high correlations (Spiess and Mullainathan 2019).

|  |
| --- |
|  |
|  |
| **Figure 4.4A. Adoption Prediction Feature importance:** Top figure presents all features that are selected by the prediction model and their feature importance score. The bottom figure allows for a better comparison of feature importance by excluding the age of the child at the first waiting year, as it dwarfs all other features. |

|  |
| --- |
|  |
|  |
| **Figure 4.4B: Foster Home Prediction Feature importance**. Top figure presents all features that are selected by the prediction model and their feature importance score. Bottom figure excludes age at entrance to foster care, as it dwarfs all other features. |

## 4.3 Inferring the Effect of AD Laws from the Model

The high performance of the models enables the calculation of the legal lift (LL). We do so by simulating the home potential of the cases *as if* they are after the legal change and subtracting the result from the home potential of the cases *as if* they were under no legislation (while keeping all other case features constant). As each case has a factual legislation value, one simulation corresponds to the factual state and the other is the counterfactual simulation. Positive values indicate that legislation improved the home potential, and vice versa.

Notably, the legislation lift is an *individual* effect, as it accounts for the difference in the home potential of each case given all of the case’s specific characteristics. This strength allows us to derive both the mean overall LL and the LL per certain case characteristics (age, race, etc.).

We calculated the LL function based on the train set and validated it on the test set. Figure 4.7 provides the quantile plot of the legal lift values (results from the train set and test set overlap).

We find that for the vast majority of adoption cases (~65%), the enactment of an AD law has zero impact on adoption outcomes. About 6% of adoption cases experience a very small negative lift and about 29% of the adoption cases experience a modest positive lift.

For foster home placements, we find that for about 83% of the cases legislation makes a positive difference. For 13% of the cases legislation makes no difference (around zero lift) and for 4% of the cases the LL is negative.

With respect to both outcomes, then, the average LL is positive: on average, AD legislation improves adoptive children’s chances to find both a permanent and foster home placement.

|  |
| --- |
|  |
|  |
| **Figure 4.5**: A quantile plot of the AD legal lift for adoption cases (Top) and foster home cases (bottom). |

## 4.4 Testing Real-World Effects in Retrospective Experiments

Next, we elucidate the value of the inferred LL using the test set as a retrospective experiment and examining whether the predictions of the model are born out by the data. To this end, we compare the outcomes of cases for which the model predicts a large effect of the AD rule (above the 80th percentile of LL values) in the presence (AD\_A = 1) and absence (AD\_A = 0) of such a rule. If the model is accurate in its assessment of the effect of legislation, we should be able to observe a real difference in the outcomes of pertinent AD and no-AD cases.

Figure 4.6A plots the results for adoption cases, showing a significant 10% difference in real adoption outcomes (p < .001). As can be shown, cases that are above the 80th percentile of LL values appear to be more challenging cases that have lower success rates than cases below the threshold. Yet these cases perform modestly better under an AD rule. Cases below the LL threshold have very high success rates (close to 100%) which are not influenced by AD law. In other words, the LL appears to improve difficult-to-adopt cases. We further explore this finding in the next section.

Second, we examine how long it took for adoption cases to be released from foster care (achieve permanency), using the same method. Figure 4.6B shows that cases above the 80th percentile of LL values take less time to place under an AD rule, a reduction of 6.5% in their overall time in care, about 3.1 months less (p < .001).

|  |  |
| --- | --- |
|  |  |
| **Figure 4.6A. The retrospective experiment: Success in finding a permanent home.** Cases with a legal lift above the 80th percentile threshold present higher actual adoption rates under AD law, a 10% increase or a ratio of 1.1. (P value < 0.001). | **Figure 4.6B.** **The retrospective experiment: Overall length of stay in foster care.** Cases with legal lift above the 80th quantile show a shorter actual length of stay in foster care under AD law, a 6.5% reduction, or 95 days less (P value < 0.001). |

Figure 4.7 plots the results for foster home placements, showing a statistically significant 7% average increase in finding a foster home (p < .001). The effect is larger for cases that are above the 80th percentile of LL values, and is also evident for cases below this threshold (a 3% increase, p < .001). The cases experiencing the larger LL again appear to be the harder-to-place cases, whose foster home placement rates are lower on average, yet improve under AD laws.

|  |  |
| --- | --- |
|  | **Figure 4.7A. The retrospective experiment: Success in finding a foster home.** Cases with a legislation lift above the 80th percentile threshold present higher actual placements in foster homes under the AD legislation, a 12% increase or a ratio of 1.12 (p < 0.001). Cases below the threshold also show an average increase, albeit a smaller one of 5% or a 1.05 ratio(p < 0.001). |

Next, we examine how long it took to place children in a foster home. This analysis naturally focuses on children that were placed in a foster home, and Figure 4.7B presents the results for “home positive” cases, namely cases that were placed in a foster home for their entire duration of care or experienced a stable positive transition from institution to home. An analysis including all cases that were placed in a foster home, even if temporarily, yielded the same results.

We provide two calculations of the length of time it took to find a foster home: from the date the child was *first* removed from home (Figure 4.7B top) and from the date the child was *last* removed from home (Figure 4.7B bottom). We see the same results: cases both above and below the 80th percentile of LL values take less time to find a foster home placement under an AD law and this reduction is larger in magnitude for cases above the 80th percentile of LL values.

|  |  |
| --- | --- |
|  |  |
| **Figure 4.7B.** **The retrospective experiment: Time to place in a foster home.**  On the left/right, After legislation, cases above the 80th quantile of LL values take 86/72 days less to place in a foster home from the day of first/last removal (p<0.001). Cases below the threshold present a 46/28 days improvement (p<0.001). | |

## 4.5 Heterogeneous Effects of AD Laws

Thus far we validated the AD LL empirically and estimated average treatment effects, namely effects at the level of the entire children population. In this section we move to explore potential heterogeneity in treatment effects, namely whether the effect of AD laws is uniform or varies between different subgroups of children in care.

The first and most important variable to explore is children’s age at entry to foster care and the adoption process. As we already noted several times, the data show that age is a major factor shaping outcomes for children in care, with older children being less successful in placing into foster homes and permanent homes.

Interestingly, we find that the effect of AD law varies between age groups in terms of permanent home outcomes, but less so in terms of foster home outcomes. As Figure 4.8A shows, AD laws substantially increase the chances of older adoption candidates to find a permanent home. This effect is found even though the data was weighted by age and that the lift was calculated separately within each age group. Yet Figure 4.8B shows that when it comes to finding a foster home, most age groups benefit uniformly from AD laws, with only ages 12-14 benefiting more than their counterparts.

|  |
| --- |
|  |
| **Figure 4.8A. The effect of AD Law on permanent home placements in different ages [adoption cases]** |

|  |
| --- |
|  |
| **Figure 4.8. The effect of AD Law on foster home placements in different ages [foster cases]** |

Next, we explore heterogeneity treatment effects in other features that the models identify as important for case outcomes. For adoption cases, we find heterogeneous treatment effects in several respects: the AD LL effect is significantly larger for black children (blkafram), mentally retarded children (mr), sexually abused children (sexabuse), children with visual or hearing impairment (vishear), children with clinical disability (clinDis), emotional distress (emotdist), and for children who were previously adopted (everadpt), in addition to several other features. Notably, the LL is positive in all features and all feature levels. Not all covariates show heterogeneous treatment effects.

|  |
| --- |
|  |
| **Figure 4.9A. Heterogeneous AD LL for the most important features for predicting adoption outcomes.**  Covariate labels: **Asian**, **mr** - mentally retarded, **sexabuse** - Removal reason - sexual abuse; **vishear**; **AD\_A** - AD legal status; **hawaiipi**; **prtsail -** removal reason - parents in jail; **aachild** - removal reason - alcohol abuse of the child; **dachild -** removal reason - drug abuse of the child; **othermed** - Other Medically Diagnosed Condition Requiring Special Care, including autism, AIDS, cancer, diabetes, epilepsy, and other conditions; relinqsh; **ClinDis** - removal reason - child disability; **prsdied - r**emoval reason - parent died; **mul\_waiting** - the child has multiple adoption cycles (see Online Appendix ); **phydis** - the child is physically disabled; **amiakn -** child American Indian or Alaska Native; **emotdist** - Emotionally Disturbed; **manrem** - Removal manner (1= voluntary; 2 = court ordered; 3 = not yet determined, mostly short-term cases); **housing**  - removal reason inadequate housing; **ChBehPrb** - removal reason - child behavior problem; **RE** - an applicable RE law (0 = none, 1 = yes; -1 = the law changed from 0 to 1 while in care); **everadpt** (1 = child has been previously adopted; 2 = no; 3 = the child has been abandoned or the child’s parents are otherwise not available to provide such information; unclear what zero is); **hisorgin** - Child Hispanic Origin (1 = yes, 2 = no, 3 = unable to determine; unclear what zero is); **Childis** - removal reason - child disability; **apparent** - alcohol abusing parent; **untodetm** - unable to determine race; **white** - child is white; **neglect** - removal reason is neglect; **nocope** - removal reason - caretaker inability to care for the child; **phyabuse** - removal reason physical abuse; **daparent** - Removal Reason - Drug Abuse Parent; **Blkafram** - Child Black/African American; **abandmnt** - removal reason - abandonment; **sex** - 1 is male. |

For foster home placements, we find heterogeneous treatment effects in several features, some showing the same heterogeneity and in the same direction as in adoption outcomes (sexual abuse, emotional distress, disability, visual/hearing impairment) and some are new (alcohol and drug abuse of child, behavioral problems, and more - children with these conditions show greater improvement than children without these conditions).

As in adoption, the AD lift is positive for all covariates and for all covariate levels, but some covariates appear to be influenced differently by AD in adoption as compared with foster home outcomes. For example, black children experience a higher LL than non-blacks in adoption outcomes, but a smaller LL than non-blacks in foster home outcomes (in both cases, the effect of AD laws is positive). In contrast, drug abusing children and children with behavioral problems experience higher LL in foster home outcomes, and lower in adoption outcomes, as compared with children without these conditions. And yet per other covariates (e.g. abandonment) there is heterogeneity in the AD effect in one set of outcomes and uniformity of the effect in the other set of outcomes. Again, there is no group that shows a negative impact on one of its levels.

|  |
| --- |
|  |
| **Figure 4.9B. Heterogeneous AD LL for the most important features for predicting foster home outcomes.** |

## 4.6 Robustness Checks

In this section we describe several supplementary analyses. First, we consider several features of the data that introduced limitations on causal inference in our states’ analysis. Second, we examine the average AD effect throughout the years and demonstrate certain time trends in children covariates. Third, we explore the relationship between the AD lift and socioeconomic features we collected, to identify potential mechanisms or confounders of the effect.

### Examining the AD effect in the states

As we discussed in Section 4.1, our prediction models exclude the state covariate from the analysis as inclusion of that variable would have drastically reduced case overlap and exclusion did not detract from the ability to predict child outcomes and. In section 4.2 we showed that both the adoption and the foster home models reach high levels of accuracy in predicting outcome predictions in all states.

Alongside this reassurance regarding the accuracy of the prediction, one might worry about the implications of the known variation between states for the estimation of the average treatment effects. Yet comparing the average effect of AD laws between states is quite challenging for two reasons: First, the dataset has a limited year range, 2000-2019. Second, states implemented legal change at different points in time. Some states implemented the change before the first data year and therefore have no pre-change data. Other states implemented the change early or late in the data, creating uneven terms for examining the effect on child outcomes. As placing children in both temporary and permanent homes may take several years, states that implemented the change early or late may introduce bias to the comparison.

Consider a state that legislated early, relative to the first year of data, say in 2004. Successful cases in the pre-change period are likely to be relatively young cases that take less time to find placement. In contrast, successful cases in the post-change period will cover the entire range of ages. Similarly, challenging-to-place cases have less time to place in the pre-period, and thus may show higher failure rates than comparable cases in the post-period, that have a long period of time to find placement simply because the dataset creates asymmetry in the number of data years before and after the legal change. The same is true, of course in the opposite direction, for states that legislated *late* in the data, for example in 2015. A more general issue with the limited year range is that data in the first and last years have more missing information (either because of more incomplete or erroneous records in the first years, or because children’s outcomes are still unknown in the late years).

While this limitation is a function of both the dataset and the pace of legal events, this section seeks to mitigate it to the extent possible by artificially leveling the comparison fields between states. To this end, we test the model in a new retrospective experiment that focuses on thirteen states that implemented the legal change between 2008-2012, the middle years of the dataset. For adoption cases, we force equal conditions for comparison by considering a case successful if it found a permanent home within 4.2 years, the 80th quantile of time to adoption (Figure OA5). Otherwise, the case is defined as failure. We exclude cases that began the adoption process in a window of 4.2 years around the year of change, as these cases did not have the same chances of finding a home as the others. This buffer zone also addresses potential anticipation effects. Changes are often announced before they are implemented, which may lead to changes on the ground even before the formal date of change. Under these restrictions, we examine the permanent home outcomes and time to permanency of adoption cases above and below the 0.8 quantile of AD LL values.

In line with our general results, we find that cases above the threshold succeed more on average in finding permanent homes in the post-period, and take less time to place, outperforming equivalent counter-cases in the pre-period. In some states, cases below the threshold *also* succeed more in the post-change period (this is more true in terms of adoption rates and less in terms of time to permanency). The effects are mostly smaller than those observed for the top LL cases, in accordance with expectations.

Figures 5A and 5B plot the results and the accompanying tables provide the results of the Mann-Whitney tests of significance. In terms of permanent home rates, the differences are highly statistically significant except for Illinois and Vermont, where the change in adoption rates is in the same direction but not statistically significant. Notably, cases below the threshold in Illinois do show significant increase in adoption rate in the post-period. In terms of time to permanency, we find significant reductions in time to permanency among above-the-LL-threshold cases, except for two states--Oregon and Maryland--where time to permanency increased significantly. These two states experienced both an improvement -- higher rates of permanent home placements -- and a deterioration -- longer time to place in permanent homes. The results for the other states were consistent along the two outcomes.

|  |
| --- |
|  |
| **Figure 5A: Permanent home placement rates in mid-range AD states under comparable comparison conditions [adoption cases]** |
|  |

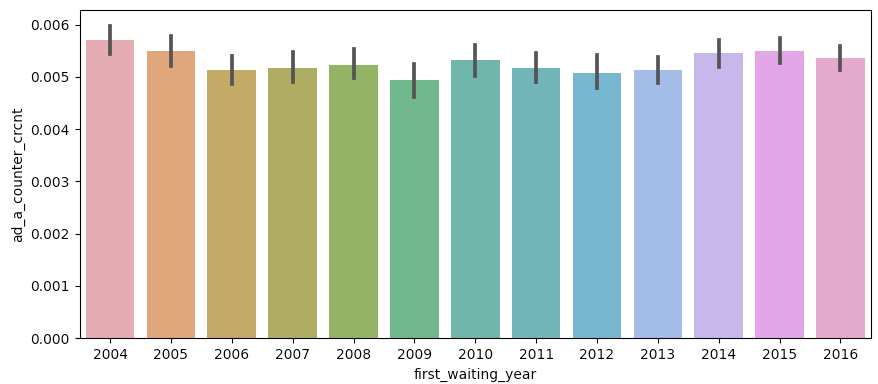
|  |
| --- |
|  |
| **Figure 5B: Time to permanency in mid-range AD states under comparable comparison conditions [adoption cases]** |
|  |

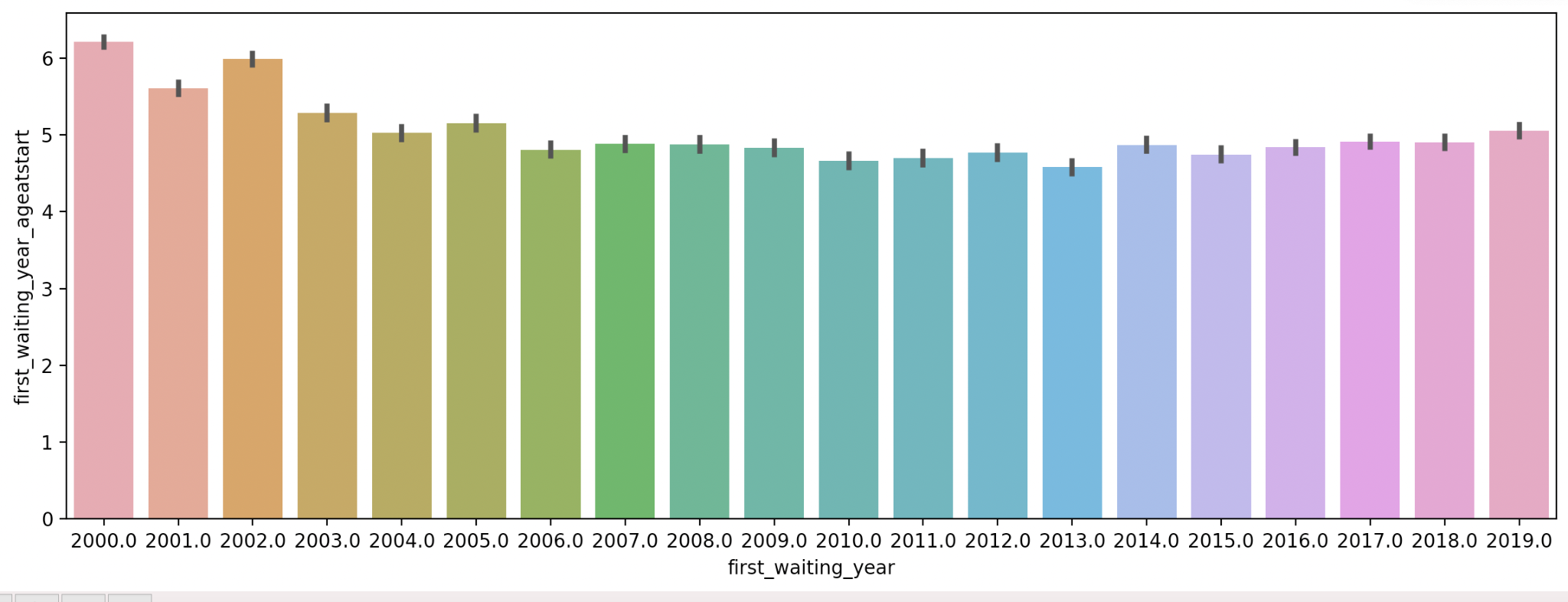
We conduct the same analysis for foster home placements, under the same principled restrictions, considering a case successful if it found a foster home within 3.18 years, the 80th quantile of time in foster care (Figure OA5) and creating an appropriate window around the date of legal change. We find higher rates of foster home placements in cases both above and below the LL threshold after the legal change (all p’s < .0001).

|  |
| --- |
|  |
|  |
| **Figure 5C: Foster home placement rates in mid-range AD states under comparable comparison conditions [foster cases]**. On the y-axis, the percent of cases in category who were able to find a foster home in 3.18 years. We compare cases above and below the 80th quantile of LL values pre- and post-legal change. |

### Examining the AD effect in different years

As Figure 6A shows, the AD lift fluctuates to some degree between data years, remaining positive for all years (we focus on the 2004-2016 range due to the higher rates of missing data in the early and late data years, discussed above). However, Figures 6B and 6C show yearly comparisons are vulnerable to the fact that the children's population changes over the years. Children in the first data year, 2000, were more than a year older than children from 2006 to 2016, and it is unclear whether there is a slight increase in ages towards 2019. These findings lend further support to our decision to weigh the data according to age and to the series of limited-year, limited-states retrospective experiments that we reported in the previous section. It is possible that additional factors also change over the years: for example, the rate of children with emotional distress fluctuates over the years and that the rate of children removed for drug abusing parents is steadily on the rise (Figures OA7,8).





### Socio-Economic Factors

Socio-economic factors may be confounded with legislation, yet the models that had access to socio-economic factors as part of the feature vector eliminated the overlap between AD and no-AD cases, and therefore were not analyzed further (See Table 3). We therefore explore directly the relationship between the AD lift and 90 socio-economic features we collected from the U.S. Bureau of Economic Analysis (BEA), the Census, and several other sources of publicly available data described in the online appendix. These features include yearly figures on state/region employment, per capita earnings, personal transfer receipts, family assistance, transfer receipts of nonprofit institutions, number of jobs and funding for education, health care, and social assistance, number of jobs and funding for Religious, grantmaking, civic, professional, and similar organizations, number of jobs and budget of state and local governments, the rate of same-sex couple households, and more.

We examined the Pearson correlations between the AD lift and each of these 90 variables and none of these relationships was statistically significant. In other words, we did not find evidence of socio-economic factors confounding the effect of AD laws on child outcomes. We present a few joint histograms of factors that we were particularly interested in in the Online Appendix.

# Discussion and conclusions

Against a heated legal debate on the implications of equality rules for child outcomes, we set out to research the question of whether children are negatively affected by antidiscrimination rules for child welfare agencies. We estimated the effects of AD laws on children in foster care by training a machine learning algorithm to predict the permanent and temporary home outcomes of individual cases, simulating the causal effect of legislation using counterfactual examples from the data, and validating the inferred effect in a set of retrospective experiments on the hold-out data that was not previously accessible to the algorithm. Notably, we find that it is too early at this point to estimate the effect of religious exemption laws, as most of them were enacted very recently and they overall apply to fractions of the data. We plan to pursue this work in the future.

Our main finding is that of consistent and positive effects of state-level antidiscrimination laws on child outcomes: children waiting for adoption increase on average the chances of finding a permanent home and wait less time to be discharged from foster care. In addition, children in foster care increase their chances of finding a foster home while in care and wait less time to be placed in a foster home.

Furthermore, we find that these outcomes do not fall evenly across cases, but are particularly concentrated among harder-to-place children. In particular, we find that the positive effects of antidiscrimination laws increase for older children, who are often most challenging to place, as well as for additional vulnerable subgroups such as sexually abused children and children with several disabilities.

Alongside these positive effects, we find that antidiscrimination laws at the state level have mixed effects in some states, for example increasing permanent home placements but also increasing the time to achieve such placements. While these mixed effects pertain to a minority of the compared states and most effects are consistently positive, treatment heterogeneity provides important grounds for additional research in this area.

Our study is important for at least two reasons. First and most obviously, answering the empirical question correctly is necessary for resolving the legal question in a manner that is most just. However the balance between religious freedom and equality law is struck by courts and legislatures, that should be done on the basis of real evidence about the effects on children, whose welfare should be at the center of the debate. We find that the enactment of antidiscrimination rules at the state level does not harm children and in most cases and with respect to most outcomes it actually benefits them. We recommend that further research is carried out to examine heterogeneity in treatment effects in this area, as well as additional legal effects: the effect of religious exemptions (RE) laws, and the combination of AD and RE laws, conditioning on the availability of sufficient and reliable data; our “law in action” analysis of the effects of conflicts on child outcomes; and analyses of the effects of antidiscrimination rules at the local level.

A second contribution of our work is for establishing a reliable empirical foundation for the heated constitutional and political battles within the culture wars (Barak-Corren 2020, 2021). Values are central to those debates, of course, and so are legal rules. But often the debates proceed on the basis of contrasting and untested assumptions about the facts, as each side relies on anecdotal evidence that serves its argument. We hope that facts can help to bring the parties closer together by establishing a common basis of evidence against which those questions can be debated, refined, and nuanced (Barak-Corren 2021). Assuming that both parties agree that children’s welfare should be at the center of the debate, gravitating the debate towards examining child outcomes can generate new points of agreement and more nuanced avenues for the future of culture war conflicts. Not every aspect of these difficult questions can be resolved empirically, but some can -- or so we hope.

# References

Alberto Abadie, Alexis Diamond & Jens Hainmueller (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program, Journal of the American Statistical Association, 105:490, 493-505, DOI: [10.1198/jasa.2009.ap08746](https://doi.org/10.1198/jasa.2009.ap08746)

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31 (2): 3-32.

Athey, Susan, and Guido W. Imbens. 2019. [Machine Learning Methods That Economists Should Know About](https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-080217-053433). Annual Review of Economics, 11:1, 685-725

Baiardi, Anna, and Naghi Andrea A. The Value Added of Machine Learning to Causal Inference: Evidence from Revisited Studies. [arXiv:2101.0087](https://arxiv.org/abs/2101.00878)

Berk, R.A., Sorenson, S.B. and Barnes, G. (2016), Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. Journal of Empirical Legal Studies, 13: 94-115. <https://doi.org/10.1111/jels.12098>

Children's Bureau of the U.S. Department of Health and Human Services (2000), Administration On Children, Youth And Families, Administration For Children And Families, U.S. Department Of Health And Human Services. *National Child Abuse and Neglect Data System, Detailed Case Data Component (NCANDS DCDC), 1995-1997*[Dataset]. National Data Archive on Child Abuse and Neglect. <https://doi.org/10.34681/JD5D-Y494>

Children׳s Bureau of the U.S. Department of Health and Human Services. (2014). Child Welfare Outcomes 2009–2012: Report to Congress. Washington, DC. <http://www.acf.hhs.gov/programs/cb/resource/cwo-09-12>

Kaufman, A., Kraft, P., & Sen, M. (2019). Improving Supreme Court Forecasting Using Boosted Decision Trees. *Political Analysis,* *27*(3), 381-387. doi:10.1017/pan.2018.59

Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, *116*(10), 4156-4165.

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. *PS: Political Science & Politics,* *48*(1), 80-83. doi:10.1017/S1049096514001784.

Katie K. Lockwood, Susan Friedman, Cindy W. Christian, Permanency and the Foster Care System, Current Problems in Pediatric and Adolescent Health Care, 45(10), 2015, 306-315,

https://doi.org/10.1016/j.cppeds.2015.08.005.

Louizos, C., Shalit, U., Mooij, J., Sontag, D., Zemel, R., & Welling, M. (2017). Causal effect inference with deep latent-variable models. *arXiv preprint arXiv:1705.08821*.

Neal, Brady (2020). Introduction to Causal Inference from a Machine Learning Perspective. <https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf>

Ruger, T. W., Kim, P. T., Martin, A. D., & Quinn, K. M. (2004). The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review*, *104*(4), 1150–1210. <https://doi.org/10.2307/4099370>

Sieger Lloyd MH. Foster Care Factors and Permanency for Children With Substance-Related Removals. *Families in Society*. 2021;102(1):91-103. doi:[10.1177/1044389420947225](https://doi.org/10.1177/1044389420947225)

Stott, Tonia and Nora Gustavsson, 2010. Balancing permanency and stability for youth in foster care, Children and Youth Services Review 32(4): 619-625, https://doi.org/10.1016/j.childyouth.2009.12.009.

Testa, M. F. (2004). When Children Cannot Return Home: Adoption and Guardianship. *The Future of Children*, *14*(1), 115–129. <https://doi.org/10.2307/1602757>

Wager, Stefan & Susan Athey (2018) Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, Journal of the American Statistical Association, 113:523, 1228-1242, DOI: [10.1080/01621459.2017.1319839](https://doi.org/10.1080/01621459.2017.1319839)

1. \* Netta Barak-Corren is an Associate Professor of Law at the Hebrew University of Jerusalem and a member of the Federmann Center for the Study of Rationality. Yoav Kan-Tor is an affiliate of the Hebrew University’s Center for Interdisciplinary Data Science Research and the computer science department. Nelson Tebbe is the Jane M.G. Foster Professor of Law at Cornell Law School. [↑](#footnote-ref-2)
2. Burwell v. Hobby Lobby Stores, Inc., 134 S. Ct. 2751 (2014) (contraception); Masterpiece Cakeshop, Ltd. v. Colo. Civ. Rts. Comm’n, 138 S. Ct. 1719 (2018) (marriage). [↑](#footnote-ref-3)
3. *Masterpiece*, 138 S. Ct. 1719. [↑](#footnote-ref-4)
4. EEOC v. R.G. & G.R. Harris Funeral Homes, Inc., 201 F.Supp.3d 837 (E.D. Mich. 2016). [↑](#footnote-ref-5)
5. Miller v. Davis, 2015 WL 10692640 (6th Cir. 2015). [↑](#footnote-ref-6)
6. 141 S. Ct. 1868 (2021). [↑](#footnote-ref-7)
7. Draft legislation includes the Child Welfare Provider Inclusion Act, S. 656, which would shield religious FCAs from civil rights laws. At least 11 states have enacted similar legislation, and several others are considering such measures. See, e.g., 2020 Tennessee Laws Pub. Ch. 514 (H.B. 836) (enacted Jan. 24, 2020); Arkansas S.B. 352 (2020) (pending); Kelly (2020) (surveying state laws). [↑](#footnote-ref-8)
8. Brief for City Respondents, Fulton v. City of Philadelphia, No. 19-123 (Aug. 2020), at \*3 (arguing that allowing CSS to exclude same-sex couples “would be harmful to its residents and the thousands of children it has a duty to protect”). [↑](#footnote-ref-9)
9. See *infra* FN 9. [↑](#footnote-ref-10)
10. Barclay, S. H. (2018). Book Review: A Dialogue about Religious Beliefs and Third-Party Harms in Family Law, 52(2) Family Law Quarterly 413-434, 429-432; Editorial Board, Suffer the Little Children, Wall St. J. May 22, 2018. [↑](#footnote-ref-11)
11. See, e.g., Brief of [Ten States] as Amici Curiae in Support of the CIty of Philadelphia, at 3, Fulton v. City of Philadelphia, 2018 WL 4862577 (3rd Cir. 2018) (sharing the states’ “experience” that “enforcing antidiscrimination requirements has resulted in no shortage of private agencies”). [↑](#footnote-ref-12)
12. Barclay, at 431. [↑](#footnote-ref-13)
13. Id. at 430 (quoting Editorial Board, Suffer the Little Children, Wall St. J. May 22, 2018). [↑](#footnote-ref-14)
14. Id. (quoting Kathleen Parker, Philadelphia’s Unnecessary War on Catholics, Wash. Post (May 22,

    2018)). [↑](#footnote-ref-15)
15. Id. at 431 (discussing Shatonell Fulton. Opinion, My Faith Led Me To Foster More Than 40 Kids; Philadelphia Is Wrong To Cut Ties With Catholic Foster Agencies, Philadelphia Inquirer (May 24, 2018)). [↑](#footnote-ref-16)
16. Id. at 432. Barclay is quoting a newspaper article that in turn purports to quote a legal brief, but none of the briefs linked in the article appears to contain the quote. [↑](#footnote-ref-17)
17. Brief of Massachusetts et al. As Amici Curiae In Support of Respondents, Fulton v. City of Philadelphia, 2020 WL 5074341 (Aug. 20, 2020), at 4. [↑](#footnote-ref-18)
18. Id. at 25. [↑](#footnote-ref-19)
19. Id. [↑](#footnote-ref-20)
20. Brief of Amici Curiae Scholars Who Study the LGB Population in Support of Respondents at 5-10, Fulton v. City of Philadelphia, 2020 WL 5020359 (August 20, 2020). [↑](#footnote-ref-21)
21. Id. at 16-17 (“When foster care agencies discriminate against same-sex parents, that reduces the pool of eligible parents—which means fewer available homes for placement and more children stuck in group care settings.”). [↑](#footnote-ref-22)
22. But see id. at 18-19 (arguing that LGBT parents who know that religious agencies are permitted to discriminate may avoid the system altogether, and pointing out that some localities may not have sufficient numbers of agencies to serve same-sex parents who are excluded from religious agencies). [↑](#footnote-ref-23)
23. In a separate work-in-progress, we track and record all conflicts brought about by such laws and analyze their specific effects using the AFCARS data. We then integrate the quantitative analyses with in-depth interviews with child welfare professionals in all conflict jurisdictions. In these interviews, we gain knowledge on the pertinent welfare systems, trace how conflicts developed and were resolved from the local perspective, and learn about unrelated developments that may have occurred during the relevant years and should be factored into the analysis (E.g., an unrelated policy that entered into effect and could have had an independent influence on child outcomes). Finally, agency dominance can moderate the effect of its closure on children. Therefore, we use multiple methods—lists of agencies, interviews, media searches, and web archives—to estimate how dominant the closing agency was in all conflict jurisdictions. [↑](#footnote-ref-24)
24. States that prohibit discrimination on the basis of sexual orientation either apply this prohibition to all individuals (adults and youth), only youth, or only adults. We focused on prohibitions of discrimination against adults as prospective parents. [↑](#footnote-ref-25)
25. Phila. Code § 9-1100, et seq. [↑](#footnote-ref-26)
26. 110 MASS. CODE REGS. 1.09(3) (2014) (general nondiscrimination policy); 102 MASS. CODE REGS. 1.03(2) (2014) (license requirement), 1.03(1) (nondiscrimination requirement) [↑](#footnote-ref-27)
27. <https://www.lgbtmap.org/equality-maps/foster_and_adoption_laws>. [↑](#footnote-ref-28)
28. In nine states the system is decentralized and administered by county (California, Colorado, Minnesota, New York, North Carolina, North Dakota, Ohio, Pennsylvania, Virginia). In two additional states (Nevada and Wisconsin) the system is partially administered by counties. However, seven of these 11 states have state-level antidiscrimination laws that protect same-sex couples, trumping any local rule. Out of the four remaining county-administered states, North Dakota and Virginia have a state-level religious exemption law and North Carolina has neither law at the state level. [↑](#footnote-ref-29)
29. While reporting started in 1995, the standards of reporting were not uniform and the collection of data was partial, inconsistent, and involved only 15 states, Children's Bureau (2000). Starting from 2000, data collection was unified and reported in a consistent format in all states. It seems that issues continued in the early years of the new dataset, as we found missing data and irregularities during 2000-2002, particularly in some states, making those early years less reliable for direct focus. For a detailed description of these issues, see Online Appendix. [↑](#footnote-ref-30)
30. To preserve the anonymity of children in the system, counties with less than 1000 children in a given year are grouped under code 8 in lieu of their federal FIPS code. Cases are still identified by their state. The proportion of code 8 data varies from state to state. [↑](#footnote-ref-31)
31. Identified by state and record number. Data on Puerto Rico was kept in the training set of the model but is not part of further analyses. [↑](#footnote-ref-32)
32. For AFCARS codebook that describes all of the variables, see: https://www.ndacan.acf.hhs.gov/datasets/pdfs\_user\_guides/afcars-foster-care-file-codebook.pdf [↑](#footnote-ref-33)
33. Although states report semi-annually, the database curators at [The National Data Archive on Child Abuse and Neglect](https://bctr.cornell.edu/projects/national-data-archive-on-child-abuse-and-neglect/), or NCACAN, recommend using the yearly files, that undergo close examination and refinement. [↑](#footnote-ref-34)
34. Children can run away from both foster homes and congregate care placements and it is unclear whether the reasons are related or unrelated to the quality or type of placement. [↑](#footnote-ref-35)
35. Not all children in foster care are waiting for adoption. Many are removed temporarily from their home and are waiting to be reunified with their birth parents. Others may have entered the system at an age that was too advanced for them to be candidates for adoption. The AFCARS codebook explains that a child is waiting for adoption if they are between 0 and 17 years old and their parents have lost parental rights, or if the child's case goal is adoption and they are in the foster care system at the end of the fiscal year, excepting kids 16 or 17 years old whose parents have lost parental rights and their case goal is emancipation. This definition encompasses all children for which welfare authorities seek to find an alternative permanent home, whether because social workers determined the child should be adopted, or because the parents already lost parental rights. [↑](#footnote-ref-36)
36. In addition, twelve percent of the cases were previously adopted. We used this feature in our prediction model. [↑](#footnote-ref-37)
37. These were cases where children were removed from home more than one year before their first data year (this might have happened if a child entered foster care before 2000, the first year of the data, or because of administrative failure to record the case on time) and were missing information about when they first began being candidates for adoption. Without information about the start of the adoption candidacy, we could not derive variables important for prediction, such as age at the beginning of the adoption process. To be as inclusive as possible, we kept cases with missing early entry years as long as they had at least one non-waiting year before their adoption process began, and labeled such cases as “missing\_data”. [↑](#footnote-ref-38)
38. We grouped all positive outcomes together because (a) they appear to be meaningfully the same in terms of achieving permanency, in contrast to emancipation, runaway, or death, see Children׳s Bureau of the U.S. Department of Health and Human Services (2014); (b) although we would have ideally wanted to be able to distinguish between permanent placements with relatives and non-relatives, as conflicts with religious agencies typically relate to non-relative adoptive parents, the structure of the variable does not permit such distinctions. The issue here is that although there is a separate category for living with relatives, both adoption and legal guardianship can be awarded to relatives and non-relatives (AFCARS codebook explicitly specifies this with respect to guardianship, but most states permit and even encourage adoption by relatives (Testa 2004). NDACAN statisticians confirmed in our consultations that adoption may include both types of adoptive parents. Hence, it is impossible to distinguish between relatives’ and non-relatives’ homes. [↑](#footnote-ref-39)
39. In practice the value of EX[E[Y|X,T=1]] and EX[E[Y|X,T=0]] are approximated using a machine learning model theta. [↑](#footnote-ref-40)
40. See supra pp. 7-8. [↑](#footnote-ref-41)
41. This issue is not unique to our method. Regression models have been shown to perform poorly in situations where there is insufficient overlap, but their standard diagnostics do not involve checking this overlap ([Dehejia and Wahba, 1999](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/" \l "R21), [2002](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/#R22); [Glazerman *et al.*, 2003](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/#R26)). [↑](#footnote-ref-42)
42. AD\_A is the variable in the data capturing whether a case is administered under an AD rule that applies to adoptions, as distinct from AD\_F, that captures the same for foster care. As noted in the section describing the AD laws database, less than a handful of states have AD rules that apply only to adoptions or only to foster care. We therefore have two separate indicators of legislative status, one for each. [↑](#footnote-ref-43)
43. 44% of the models actively use the AD feature, indicating that being under an AD law is not an indispensable factor in the prediction of adoption outcomes. [↑](#footnote-ref-44)
44. In the Online Appendix we explain the empirical establishment of case overlap using propensity scores. Overall, 24% of the cases that had no control cases were excluded from the analysis. [↑](#footnote-ref-45)
45. We focus on the 2004-2016 year range because first-data and end-of-data years have more missing information -- e.g. cases without full history in the early data years and many open cases with unknown outcomes in end-of-data years. Table 5 shows the results 2000-2018. [↑](#footnote-ref-46)
46. DC presents poorer model performance compared to the other states (AUC=0.68). This appears to be explained by the weaker relationship between age and home potential in DC: the ages of cases that find a foster home and cases that do not find a home are exceptionally similar, diminishing the usefulness of the main prediction feature in the model. See Figure OA2 for more details. [↑](#footnote-ref-47)