

Predictors of Placement Instability in Foster Care: A Machine Learning Analysis of AFCARS Data

William Kang
willkang07@gmail.com

ABSTRACT

Placement instability, when a child in foster care experiences multiple placement changes, is associated with many things, including disrupted schooling, damaged relationships, and compounded trauma. This study makes use of restricted data from the federal Adoption and Foster Care Analysis and Reporting System (AFCARS) and works to identify factors associated with placement instability among children who exited foster care during federal fiscal year 2024, drawing on an analytic sample of $N = 131,315$ children with complete data on the selected features, out of 179,420 children who exited that year. Placement instability was defined as three or more placement episodes within a child's current removal episode, a threshold consistent with prior foster care research. Three models were compared in this project: logistic regression, XGBoost, and a class-weighted XGBoost model. The class-weighted model achieved the strongest recall on the unstable class (0.60). It correctly identified 60 percent of children who experienced instability, compared to 43 percent for the unweighted model. Using SHAP analysis, age at entry and most recent case plan goal emerged as the two strongest predictors. Older age at entry was associated with higher instability, and children with a case plan goal of long-term foster care (72.3 percent instability) or emancipation (70.6 percent) had substantially higher instability rates than those with a goal of reunification (30.2 percent). These findings are correlational; because case plan goal is recorded as of a child's most recent AFCARS record rather than fixed at intake, there is likely a bidirectional relationship between case plan goal and instability rather than a one-directional predictive one. This analysis is best understood as a retrospective, cross-sectional look at which case characteristics are associated with instability among children who have already exited care, not as a prospective screening tool. Therefore, it is intended to complement direct casework and professional judgment.

INTRODUCTION

Hundreds of thousands of children in the United States spend time in foster care every year, and for many of them, that time is marked by more than one placement, or location lived at. A child might be moved from a foster home to a group home, then to a relative's house, then somewhere else entirely, each move disrupting their life, including relationships, schooling, and whatever stability they'd started to rebuild. Placement instability shapes how a child experiences an already difficult chapter of their life, and it can leave lasting effects that remain even years after the system is left.

I wanted to understand this problem more rigorously than intuition allows. What actually predicts whether

Oct 2026
Vol 12, No 5.

a child will experience a stable placement versus repeated disruption? Is it their age when they entered care? The type of placement they were in? The permanency goal their caseworker set for them? Lacking that kind of systemic, population-level data, I turned to the Adoption and Foster Care Analysis and Reporting System, known as AFCARS, a federal dataset that tracks case-level information on children in foster care nationwide.

I use one year of restricted AFCARS data in this paper to build and compare several machine learning models that predict placement instability, then examine which factors the models identify as most important. I also try to be honest about what these models can and cannot tell us. A model that predicts a pattern in the aggregate is not the same as a tool that can reliably flag which individual child is at risk, and that distinction matters for how findings like these should be used and interpreted. It is also worth stating upfront that, because this analysis is built from a single fiscal year of case records rather than data collected prospectively at the point a child enters care, its findings describe retrospective associations among characteristics recorded at or near the time a child exited foster care, not predictions made from a fixed, early point in a child's case. Before turning to the data, it is helpful to examine existing research on placement instability and predictive modeling that has already been done.

RELATED WORK

Research on placement instability predates the availability of large administrative datasets like AFCARS. James (2004) analyzed case records from a sample of children in California foster care and found that roughly one in five placement changes were behavior related, and that older age at entry, externalizing behavior, and a history of emotional abuse each increased the hazard of a behavior-related placement change. Risk was highest in the first hundred days after a child entered care. Another study, James, Landsverk, and Slymen (2004), used survival analysis on the same population to describe how placement movement unfolds over time, finding that the pace of movement is not constant and that certain child and case characteristics predict both the timing and frequency of moves. Newton, Litrownik, and Landsverk (2000) reached a related conclusion using a different lens, showing that child behavior problems and placement instability are entangled in both directions: children with more behavior problems tend to experience more placements, and more placements tend to be followed by worsening behavior, which complicates any claim that behavior problems simply cause instability.

The threshold used in this study, three or more placement episodes as the marker of instability, follows a convention with some history behind it. Cross, Koh, Rolock, and Eblen-Manning (2013) set instability at three or more placements in their study of Illinois foster care cases. It is also consistent with earlier work by Hartnett, Leathers, Falconnier, and Testa (1999, as cited in Cross et al., 2013). This study adopts the same threshold for comparability with that literature, though, as discussed further in the Data and Methods and Limitations sections, no single definition of instability is universal across studies and is set to a perfect number.

Much less of the literature on placement instability uses interpretable machine learning at a national scale. Most closely related to this project is the work on predictive risk modeling that has grown out of child protection screening. This began with Vaithianathan, Maloney, Putnam-Hornstein, and Jiang (2013), who built a predictive risk model from linked public benefit and child protection records in New Zealand to identify children at risk of substantiated maltreatment. That line of work established both the promise and the controversy of using administrative data and machine learning to flag risk in child welfare settings, and it raised many of the same interpretability and fairness questions that motivated the SHAP-based approach used here. A bit closer to the specific outcome studied in this paper, Chor, Luo, Rodolfa, and Ghani (2025) compared machine learning models with conventional regression models to support placement stabilization decisions. They used state administrative data. Trudeau, Yang, Di, Lu, and Kraus (2023) built machine learning models to predict successful placements for youth in the child welfare system. Both studies, like this one, favor models that caseworkers and administrators can act on rather than models optimized purely for accuracy.

Finally, this study's finding of a moderate association between race and ethnicity and placement instability sits alongside a larger literature on racial disparity in foster care outcomes. Heaton, Sabol, Baumann, Harison, and Goodell (2025) used AFCARS data from 2012 through 2020 to show that Black children experience greater placement instability than White children. It can also be seen that this gap is shaped by county- and state-level socioeconomic conditions, including poverty, public welfare expenditure, and residential mobility, rather than by race in isolation. That framing, that racial differences in an outcome like instability are more plausibly explained by the structural conditions correlated with race than by race itself, is the same interpretive stance taken later in this paper's discussion of the race and ethnicity feature. Taken together, this literature leaves a gap. This study tries to address that gap: a national, single-fiscal-year AFCARS sample, analyzed with an interpretable gradient-boosted model and SHAP.

DATA AND METHODS

I used the AFCARS Foster Care AB File for federal fiscal year 2024, obtained through Cornell University's National Data Archive on Child Abuse and Neglect. The full file contains 522,475 records, each representing a child who was in foster care at some point during the year. There are also 107 variables describing their demographics, case history, placement details, and much more.

Before building any model, I made a decision that shaped everything downstream: I restricted the dataset to children who exited foster care during the fiscal year, rather than including everyone who appeared in the file. Just about a third of the children in the full file were still in care when the reporting period ended, which meant their eventual number of placements was not yet known. Including them would have meant treating an unfinished outcome as though it were final, understating instability for children whose cases simply had not concluded yet. Restricting the children to those that had left that year brought me to 179,420 cases where the placement count was complete and reliable.

I defined placement instability as having three or more placement episodes within a child's current

Oct 2026

Vol 12. No 5.

removal episode, based on the variable NumPlep. It is worth being precise about what this measures. NumPlep counts placements within a child's most recent stay in care, not across their entire lifetime history with the foster care system, so a child who had a stable current episode but an unstable one years earlier would not be captured as unstable by this definition. The choice of three or more placements as the instability threshold follows Cross, Koh, Rolock, And Eblen-Manning (2013) and the earlier work they cite (Hartnett, Leathers, Falconnier, and Testa, 1999), which use the same threshold to account for the emergency or short-term placements many children pass through before settling into a more stable arrangement. As noted in the Related Work section, this is a common but not universal definition; studies that use a lower or higher threshold, or that measure instability as a rate of moves per unit time rather than a raw count, sometimes reach different conclusions about which children look unstable.

Additionally, I cross-referenced the AFCARS codebook to identify how missing or unknown values were coded, since AFCARS uses sentinel values such as -1 and 99 to represent missing data rather than leaving fields blank. I replaced these with true missing values, and also removed a handful of clearly erroneous entries, such as an age at entry of 76, which is not a super likely value for this population and likely a data entry error. Concretely, entries coded -1 for age at entry were recoded to missing, as were entries coded 99 for current placement setting, race and ethnicity, and most recent case plan goal. Age at entry values above 21 were treated as implausible and also recoded to missing, and total prior removals was capped at 10, with values above that treated as missing rather than as genuine outliers, since a small number of extremely high values would otherwise have had an outsized effect on a variable most children have at low values.

I selected ten features for the final model: age at entry, sex, current placement setting, total prior removals, physical abuse, sexual abuse, and neglect as reasons for removal, race, and ethnicity, most recent case plan goal, and clinical diagnosis status. I also tested five additional variables, including parental substance abuse and child behavior problems, but these did not meaningfully improve model performance while noticeably reducing the sample size due to missing data, so I excluded them from the final model. After removing rows with missing values in the selected features, the modeling sample consisted of 131,315 children. Out of all those children, about 36.5% percent met the definition of unstable.

Understanding timing is important here, as it affects how the results should be read. Age at entry, sex, race and ethnicity, and the reasons for removal are recorded at or near the point a child enters care, so they really precede the placement history captured by NumPlep. Current placement setting, most recent case plan goal, total prior removals, and clinical diagnosis status, however, are recorded as of the child's most recent AFCARS reporting period, which for most children in this sample is at or close to the point they exited care. To say that all more simply, several of the features used to predict instability are themselves measured when the instability has already occurred, most importantly case plan goal, which a caseworker may well have set as a response to a child's placement history rather than prior to everything. This means the models in this paper should be read as describing which case characteristics, as recorded at or near exit, are associated with having experienced instability during that episode, rather than as a model that

could be run at the point a child enters care to flag risk before it happens. This is talked about and elaborated on in the Discussion and Limitations sections.

I trained three models on this data. The first was a logistic regression, chosen as a baseline and simple model. The second was a gradient boosted tree model using XGBoost, which can capture more complex, nonlinear relationships between features. The third was the same XGBoost model with class weighting applied, which primarily adjusts the model to pay more attention to a specific class or group, in this case, children who experienced instability. I made this adjustment deliberately, because in a child welfare context, failing to identify a child at risk of instability is a more costly error than incorrectly flagging a child who turns out to be stable. All models were evaluated on a held-out test set representing twenty percent of the data. The models never saw this data during their training.

To make this reproducible, a few specifics are worth stating plainly. The eighty-twenty train-test split was stratified on the outcome variable, so the proportion of unstable children was held constant across both sets, and a fixed random seed (42) was used throughout so the split and model training are exactly reproducible. Categorical features (sex, current placement setting, race and ethnicity, most recent case plan goal) were passed to the models as integer-coded categories rather than one-hot encoded, consistent with how scikit-learn's logistic regression implementation and XGBoost handle categorical-style inputs by default. The logistic regression model used scikit-learn's default L2 regularization with a maximum of 1,000 iterations to ensure convergence. The XGBoost models used 200 trees, a maximum depth of 4, and a learning rate of 0.1; these values were set based on commonly used starting points for tabular classification problems of this size rather than tuned through a formal grid search or cross-validation procedure, which is a limitation I return to below. For the class-weighted model, the `scale_pos_weight` parameter was set to the ratio of stable to unstable children in the training set, which up-weights the minority (unstable) class in proportion to its underrepresentation. All three models used the default 0.5 probability threshold for classifying a child as unstable. In other words, no separate threshold tuning was performed, so the recall and precision figures reported below reflect that default threshold rather than a threshold chosen to optimize a tradeoff.

To interpret the models beyond simple accuracy, I used SHAP, a method that allows us to quantify how much each feature contributes to a model's individual predictions. This allowed me to identify not just how well the model performed overall, but also the factors driving its predictions.

RESULTS

The first model, which was the logistic regression model, achieved 63.8 percent accuracy, only marginally better than the 63.5 percent a model would achieve by simply guessing the majority class for every child. Its recall on the unstable class, meaning the proportion of actually unstable children it correctly identified, was 0.32.

The XGBoost model performed considerably better, reaching 68.0 percent accuracy and a recall of 0.43 on the unstable class. This improvement suggests that the relationships between these features and placement instability are not purely linear, and that a more flexible model could also be used to try to solve this problem.

Applying class weighting to the XGBoost model produced a different tradeoff. Accuracy dropped slightly to 66.1 percent, but recall on the unstable class rose substantially to 0.60, meaning the model now correctly identified 60 percent of children who experienced instability, compared to 43 percent without weighting. Since an at-risk child carries a higher real-world cost than a false alarm, I treat this weighted model as the primary result of this analysis, even though the raw accuracy was lower.

Table 1. Model performance comparison.

Model	Accuracy	Recall (Unstable)	Precision (Unstable)
Logistic Regression	63.8%	0.32	0.58
XGBoost	68.0%	0.43	0.64
XGBoost (class-weighted)	66.1%	0.60	0.57

After running the SHAP analysis of the data, age at entry and case plan goal were found to be the two most influential features, with substantially larger effects than any other variable. Other factors that I looked into, such as placement type, race and ethnicity, and clinical diagnosis status, had moderate effects, while sex, prior removals, and sexual abuse as a removal reason contributed very little.

Figure 1. SHAP summary plot of feature contributions to model predictions.

When looking at the age at entry, the direction of the effect was consistent. As it can be seen, older age at entry was associated with a higher likelihood of instability, while younger age at entry was associated with a much greater stability.

The CaseGoal factor required a different kind of analysis, since it is a categorical variable rather than a continuous one. Examining instability rates directly within each case goal category revealed a striking pattern, shown in Table 2.

Table 2. Instability rate by most recent case plan goal.

Case Plan Goal	Instability Rate	N
Long-Term Foster Care	72.3%	9,352
Emancipation	70.6%	201
Adoption	47.7%	41,228
Live with Other Relatives	42.4%	4,222

Guardianship	38.1%	10,147
Reunify with Parents	30.2%	65,585
Goal Not Yet Established	1.9%	580

Children whose most recent case plan goal was long-term foster care had an instability rate of 72.3 percent. Those with a goal of emancipation had a rate of 70.6 percent. By contrast, children with a goal of reunification with parents had an instability rate of 30.2 percent. Those with a goal of guardianship had a rate of 38.1 percent.

DISCUSSION

From this analysis, the most important finding is that age at entry matters a great deal, and in an intuitive direction. Older children entering foster care are more likely to experience multiple placements. This aligns with what is generally understood about the foster care system: younger children are often easier to place in traditional foster homes, while older children are typically more likely to enter group settings or experience placement breakdowns due to other factors that make finding a stable home more difficult.

The case plan goal finding is more complex, and it is important that we do not overstate the results. It is tempting to read the sharp difference between long-term foster care and reunification goals as evidence that certain goals cause instability. However, it would be more accurate to conclude that this relationship likely runs in both directions. A child who has already experienced significant placement disruption may be more likely to be assigned a goal of long-term foster care or emancipation because repeated instability made reunification or adoption less feasible. In other words, case plan goal may be as much a reflection of a child's instability as a predictor of it.

This is not simply a caution about interpretation; it follows directly from how the variable is measured. Because most recent case plan goal is recorded as of a child's latest AFCARS record rather than fixed at intake, in some fraction of cases the goal was likely set, or revised, after some of the instability captured by NumPlep had already happened. That means this finding is better described as a retrospective association observed among children as of their most recent record, not as a forward-looking prediction from a fixed starting point in a child's case.

Untangling this relationship better would require longitudinal data that tracks how case goals change over time relative to placement changes. This single fiscal year of data cannot resolve that question.

I also want to raise a similar caution about race and ethnicity (a moderate effect in the model). I do not interpret this as evidence that race itself drives instability. It more likely reflects structural and systemic factors that are correlated with race but not captured by the dataset that was used. This reading is consistent with Heaton, Sabol, Baumann, Harison, and Goodell (2025), who found that county- and

state-level socioeconomic conditions, rather than race itself, explain much of the racial gap in placement stability in a national AFCARS sample. This is an important area for further research.

LIMITATIONS

This analysis has several limitations that should inform how its findings are interpreted. First, it relies on a single fiscal year of cross-sectional data, so it cannot capture how instability unfolds over a child's entire time in the foster care system, only within their most recent episode.

Second, the relationship between case plan goal and instability is very likely to be bidirectional, and this dataset cannot separate cause from consequence; as discussed above, this is partly a function of when case plan goal is recorded relative to a child's placement history, and not just an inherent feature of the variable itself.

Third, AFCARS is compiled from state-level administrative data, and reporting practices are known to vary by state, which may introduce inconsistencies that this analysis cannot fully account for.

Fourth, of the 179,420 children who exited foster care during FY2024, 48,105 (about 26.8 percent) were excluded from the final modeling sample due to missing values across the ten selected features. A more than one out of four exclusion rate is large enough that it deserves direct attention rather than a passing mention. Most of the missingness traces to a smaller number of features, particularly most recent case plan goal and race and ethnicity, and if children with missing values on these features differ systematically from those without them, for instance if incomplete case records are themselves associated with less stable or less closely monitored cases, then the modeling sample used here may understate the true prevalence or severity of instability in the full exited-children population. Future work should characterize how children excluded for missing data differ from those retained. This analysis does not attempt to do just that.

Fifth, model hyperparameters were set using common defaults for tabular classification problems of this size rather than tuned through cross-validation or a formal grid search, and the classification threshold used throughout is the default 0.5 rather than one selected to optimize a specific precision-recall tradeoff. A tuned model, or a different threshold, could perform differently in either direction than what is reported here.

Finally and most importantly, these findings are correlational and retrospective rather than predictive. They describe factors associated with elevated instability among children as recorded at or near exit, not factors that cause it or that could be used to flag risk prospectively. And they should not be interpreted as a basis for individual-level clinical or casework decisions.

ACKNOWLEDGEMENT

The data used in this publication, Dataset #310, AFCARS Foster Care AB File, FY2024, were obtained from the National Data Archive on Child Abuse and Neglect. All was used in accordance with its Terms of Use Agreement license. The Administration on Children, Youth and Families, the Children's Bureau, the original dataset collection personnel or funding source, NDACAN, Duke University, Cornell University, and their agents or employees bear no responsibility for the analyses or interpretations presented here.

WORKS CITED

- Administration for Children and Families, Children's Bureau. Adoption and Foster Care Analysis and Reporting System (AFCARS) Foster Care File, FFY 2024 [Dataset #310]. National Data Archive on Child Abuse and Neglect, Cornell University, 2025.
- Chen, Tianqi, and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.
- Chor, Ka Ho Brian, Zhidi Luo, Kit T. Rodolfa, and Rayid Ghani. "Exploring Machine Learning to Support Decision-Making for Placement Stabilization and Preservation in Child Welfare." *Journal of Child and Family Studies*, vol. 34, no. 1, 2025, pp. 282-297.
- Cross, Theodore P., Eun Koh, Nancy Rolock, and Jennifer Eblen-Manning. "Why Do Children Experience Multiple Placement Changes in Foster Care? Content Analysis on Reasons for Instability." *Journal of Public Child Welfare*, vol. 7, no. 1, 2013, pp. 39-58.
- Heaton, Leanne, William Sabol, Miranda Baumann, Arya Harison, and Charlotte Goodell. "The Effects of Socioeconomic Contextual Factors on Racial Differences in Foster Care Placement Stability." *International Journal of Environmental Research and Public Health*, vol. 22, no. 8, 2025, article 1274.
- James, Sigrid. "Why Do Foster Care Placements Disrupt? An Investigation of Reasons for Placement Change in Foster Care." *Social Service Review*, vol. 78, no. 4, 2004, pp. 601-627.
- James, Sigrid, John Landsverk, and Donald J. Slymen. "Placement Movement in Out-of-Home Care: Patterns and Predictors." *Children and Youth Services Review*, vol. 26, no. 2, 2004, pp. 185-206.
- Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- Newton, Rae R., Alan J. Litrownik, and John A. Landsverk. "Children and Youth in Foster Care: Disentangling the Relationship Between Problem Behaviors and Number of Placements." *Child Abuse & Neglect*, vol. 24, no. 10, 2000, pp. 1363-1374.
- Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825-2830.
- Trudeau, Kimberlee J., Jichen Yang, Jiaming Di, Yi Lu, and David R. Kraus. "Predicting Successful Placements for Youth in Child Welfare with Machine Learning." *Children and Youth Services Review*, vol. 153, 2023, article 107117.

Oct 2026

Vol 12. No 5.

Vaithianathan, Rhema, Tim Maloney, Emily Putnam-Hornstein, and Nan Jiang. "Children in the Public Benefit System at Risk of Maltreatment: Identification via Predictive Modeling." *American Journal of Preventive Medicine*, vol. 45, no. 3, 2013, pp. 354-359.