

Legally Free Predictive Analytics: Identification of Alternative Subpopulations with Elevated Emancipation Risk Scores

June 8, 2017



Submitted to:

Korey Elger, M.S.W.
Derek Dzien
Jason Lester, M.S.

Colorado Department of Human Services
Office of Children, Youth & Families
Division of Child Welfare
1575 Sherman St., 2nd Floor
Denver, CO 80203

Submitted by:

Dallas J. Elgin, Ph.D.
Jessica Smith, M.P.P.

IMPAQ International, LLC
10420 Little Patuxent Parkway
Suite 300
Columbia, MD 21044

Contents

Introduction 3

Methodological Approach for Reviewing the ERS Algorithms 4

 Methodological Approach 4

Exploration of alternative options for identifying subpopulations with elevated risk Scores 6

Updating the ERS Processes to Focus on Subpopulations of Interest 8

Discussion of the Associated Implications 9

References 11

Appendix A: Detailed Overview of the ERS Algorithms 12

INTRODUCTION

Legally free children and youth whose parents have had their parental rights terminated have been shown to be more likely to emancipate from the child welfare system without a permanent, legal connection (i.e., “permanency”). The children and youth comprising the legally free population face complex factors that increase the risk of emancipation (Elgin, Sushinsky, Johnson, Russo & Sewell, 2015) and the subsequent likelihood of adverse outcomes including homelessness, victimization, and incarceration, among others (Keller, Cusick & Courtney, 2007). As part of its C-Stat initiative, the Colorado Department of Human Services (CDHS) executive management team and staff from the Division of Child Welfare (DCW) regularly review the population of legally free children that are approaching the age of emancipation and discuss options for removing barriers to permanency. To assist with this process, the CDHS executive management team commissioned a study of the factors affecting permanency for the state’s population of legally free children and youth, which identified a collection of distinct factors within different age groups (Elgin et al., 2015). The study’s findings were utilized to develop a collection of Emancipation Risk Score (ERS) algorithms to identify children with elevated risk levels for emancipating from the state’s child welfare system.

Over the past two years, the ERS algorithms have been used to identify children and youth with the highest risk of emancipation so that CDHS and county stakeholders could identify pertinent policies, practices, and resources that could be leveraged to help obtain permanency. The algorithms have proven effective in CDHS’s efforts to identify earlier those children with higher risk levels and to leverage resources for achieving permanency. For example, CDHS has collaborated with The Adoption Exchange and Wendy’s Wonderful Kids recruiters, who have implemented proactive and proven child-focused recruitment programs designed to find permanent homes for children and youth with the highest risk scores. While these efforts have proven effective in increasing permanency for the legally free population, a need emerged for a review and update of the subpopulations identified by the ERS algorithms. Feedback from CDHS and county staff shows that the ERS algorithms consistently identify subpopulations with the highest risk scores that are close to the age of emancipation or that possess a collection of factors that make an outcome other than emancipation highly unlikely. CDHS thus inquired as to whether the algorithms could identify additional subpopulations with risk scores that would allow CDHS and its county partners to leverage pertinent policies, practices, and resources at an earlier point in time. To achieve this goal, CDHS contracted with IMPAQ International (IMPAQ) to conduct a comprehensive review of the ERS algorithms and identify additional subpopulations with elevated risk scores.

This final report details the process used to review and update the way ERS algorithms identify populations with elevated and high-risk scores that would benefit from a coordinated effort by CDHS and county stakeholders to identify pertinent policies, practices, and resources that promote permanency. The report consists of four sections:

- An overview of the methodological approach used to review the algorithms and develop recommendations for examining different subpopulations
- A description of the approach used to examine alternative options for identifying subpopulations with elevated risk scores
- Details of the changes made to CDHS' processes for identifying children and youth with elevated risk scores
- A discussion of the associated implications

METHODOLOGICAL APPROACH FOR REVIEWING THE ERS ALGORITHMS

CDHS currently utilizes a multi-step process to identify legally free children and youth with the highest risk of emancipating from the child welfare system. In the first step, the results from the “Legally Free Predictive Analytics Project” are utilized to construct a risk score calculating the likelihood of emancipation based on a collection of statistically significant factors. The significant factors are identified via a collection of survival analysis models, in the form of stratified Cox regression models, that modeled the likelihood of achieving permanency across three different age groups – “Birth through the Age of Five”, “Six through the Age of 12”, and “13 through the Age of 17” (Elgin et al., 2015). The statistically significant variables (including factors such as race, gender, previous involvement, permanency goals, number of placements, and current placement setting) from each of the three models are used to develop independent ERS algorithms for each age group. Appendix A includes a detailed overview of the ERS algorithms.

In the next step, software is used to query the population of legally free children and youth quarterly. The resulting query consists of administrative data, including the variables utilized to construct the risk scores, for all legally free children and youth served by the state's child welfare system. In the third step, the administrative data is imported via an Excel spreadsheet into SPSS statistical software to construct the risk scores for all children and youth in the legally free population. After constructing the risk scores, mean and standard deviations are calculated for each of the three age groups to identify children and youth with the “highest risk” of emancipation. Within each group, children and youth with scores that are two deviations above the mean are categorized as having the highest emancipation risk. In the final step, the highest risk children and youth are manually exported to an Excel spreadsheet that is provided to DCW for detailed review and engagement of county partners regarding pertinent policies, practices, and resources that could be leveraged to achieve permanency for each child.

Methodological Approach

IMPAQ used a multi-step methodology to review the ERS algorithms and the associated processes for identifying high-risk children and youth. In the first step, IMPAQ reviewed the SPSS syntax and identified several opportunities for optimizing the code to improve clarity and efficiency.

Changes made to optimize the code included:

1. Automating the initial process for importing the Excel data file into SPSS
2. Updating the SPSS code to convert the “Legally Free Date” variable from a character to a numeric format
3. Generating additional code to replace missing values (i.e., null values) in the “Legally Free Date” with the mother’s or father’s “Termination of Parental Rights Date” variable (i.e., “tpr_mother_date”, “tpr_father_date”)
4. Condensing multiple lines of code into single lines when possible for a more efficient data program
5. Creating an automated process for exporting the list of high risk clients to an Excel file
6. Adding additional narrative comments throughout the SPSS file to explain each coding step

The revised code was used to replicate the process for calculating risk scores utilizing the most recent data available at the time of the analysis.¹ IMPAQ then reviewed the subpopulations of children with elevated and highest risk for the three age categories.

In the next step, IMPAQ spoke with DCW staff about the challenges associated with each of the subpopulations and potential changes that could be implemented to ensure that the processes identified children and youth who would benefit most from CDHS and its county partners’ efforts to provide pertinent policies, practices, and resources for achieving permanency.

Discussions with DCW staff provided IMPAQ with insight into the aspects of the ERS algorithms that were working successfully and those that could be improved. Staff noted that the algorithms have proven effective in DCW’s efforts to identify earlier those children with higher risk levels and subsequent efforts to leverage resources for achieving permanency. However, staff noted two existing challenges regarding the subpopulations of legally free children and youth identified by the algorithms. First, the algorithms identified several children within each of the three age groups who were 17 years or older, making efforts to reduce the risk of emancipation challenging due to the proximity of these children to the age of 18, the legal age of emancipation. Second, the algorithms identified a number of children who had been in out of home care (OOH) for 100 or more months. This extended duration of time in care also made efforts to reduce the risk of emancipation challenging for these children.

Given these challenges, IMPAQ reviewed the highest risk subpopulations to determine the percentages of children that were 17 years or older at the time that the algorithms were ran along with the percentages of children that were in out of home care for 100 months or more. Exhibit 1 provides an overview of the corresponding highest risk subpopulations that were identified under the current process of selecting children with risk scores that are two or more standard deviations above the mean.

¹ The corresponding period consisted of data pulled on January 10, 2017.

**Exhibit 1. Analysis of the Current “Highest Risk” Subpopulations
with Risk Scores 2 Standard Deviations above the Mean**

	<1-5 Age Group	6-12 Age Group	13-17 Age Group	All Age Groups
# of Children	24	28	5	57
% of Children 17+ or Older at Time of Analysis (# of children)	20.8% (5)	82.1% (23)	100% (5)	57.9% (33)
% of Children in OOH Care for 100+ Months* (# of children)	79.2% (19)	96.4% (27)	0.0% (0)	80.7% (46)
% of Children that are 17+ and 100+ Months in Care (# of children)	20.8% (5)	78.6% (22)	0.0% (0)	47.4% (27)

*Note: Following the variables in the ERS algorithms, ‘OOH Care’ includes congregate care and family-like settings for children in the <1-5 and 6-12 age groups. As time in congregate care was not a statistically significant variable in the ERS algorithm for the 13-17 age group, ‘OOH Care’ consists solely of family-like settings.

As Exhibit 1 shows, 57.9 percent of all children identified were 17 years or older, while 80.7 percent had been in out of home care for 100 or more months. While achieving permanency for legally free children that are either 17 years or older *or* have been in out of home care for 100 or more months presents a challenge, children that are 17 years or older *and* have been in out of home care for 100 or more months present an arduous challenge. IMPAQ thus examined options for minimizing the number of legally free children identified by the ERS algorithms who are 17 years or older *and* have been in out of home care for 100 or more months.

EXPLORATION OF ALTERNATIVE OPTIONS FOR IDENTIFYING SUBPOPULATIONS WITH ELEVATED RISK SCORES

The options examined included identifying children with risk scores that were between 1-1.5 standard deviations, 1.5-2 deviations, and 1-2 standard deviations. Based on these examinations, IMPAQ proposed two options for identifying alternative subpopulations with elevated risk of emancipation.

Option 1: Subpopulations with Risk Scores 1-2 Standard Deviations above the Mean

The first option consisted of a subpopulation of children with risk scores that were between one and two standard deviations above the mean risk score. This option identified a slightly smaller number of legally free children (42) with “elevated” risk scores. Exhibit 2 provides an overview of the associated subpopulations.

Exhibit 2. Analysis of the Subpopulations with Risk Scores 1-2 Standard Deviations above the Mean

	<1-5 Age Group	6-12 Age Group	13-17 Age Group	All Age Groups
# of Children	6	32	4	42
% of Children 17+ or Older at Time of Analysis (# of children)	0.0% (0)	50.0% (16)	75.0% (3)	45.2% (19)
% of Children in OOH Care for 100+ Months* (# of children)	0.0% (0)	46.9% (15)	0.0% (0)	35.7% (15)
% of Children that are 17+ and 100+ Months in Care (# of children)	0.0% (0)	21.9% (7)	0.0% (0)	16.7% (7)

*Note: Following the variables in the ERS algorithms, 'OOH Care' includes congregate care and family-like settings for children in the <1-5 and 6-12 age groups. As time in congregate care was not a statistically significant variable in the ERS algorithm for the 13-17 age group, 'OOH Care' consists solely of family-like settings.

This option significantly reduced the number of children that were 17 or older or had been in out of home care for 100 months or more. The percentage of children that were 17 or older *and* had been in out of home care for 100 or more months was reduced from 47.4 to 16.7 percent. While this option provided considerable improvements in these areas for the <1-5 and 13-17 age groups, 21.9 percent of the 6-12 age group consisted of children that were 17 or older *and* had been in out of home care 100 or more months. IMPAQ thus examined a second option intended to minimize the number of children in the 6-12 age group experiencing both factors.

Option 2: Replacing the 6-12 Age Group with Risk Scores 1-1.5 Standard Deviations above the Mean

The second option used the same approach of shifting the focus to risk scores that are between one and two standard deviations for the <1-5 and 13-17 age groups, but shifted the risk score for the 6-12 age group to one and one and a half standard deviations. This option would identify a notably smaller number of legally free children (28) with “elevated” risk scores. Exhibit 3 provides an overview of the associated subpopulations.

**Exhibit 3. Analysis of the Subpopulations
with Risk Scores of 1-1.5 Standard Deviations for the 6-12 Age Group**

	<1-5 Age Group	6-12 Age Group	13-17 Age Group	All Age Groups
# of Children	6	18	4	28
% of Children 17+ or Older at Time of Analysis (# of children)	0.0% (0)	38.9% (7)	75.0% (3)	35.7% (10)
% of Children in OOH Care for 100+ Months* (# of children)	0.0% (0)	44.4% (8)	0.0% (0)	28.6% (8)
% of Children that are 17+ and 100+ Months in Care (# of children)	0.0% (0)	16.7% (3)	0.0% (0)	10.7% (3)

*Note: Following the variables in the ERS algorithms, 'OOH Care' includes congregate care and family-like settings for children in the <1-5 and 6-12 age groups. As time in congregate care was not a statistically significant variable in the ERS algorithm for the 13-17 age group, 'OOH Care' consists solely of family-like settings.

This option further reduced the number of children that were 17 or older or had been in out of home care for 100 months or more. Most notably, the percentage of children that were 17 or older *and* had been in out of home care for 100 or more months was reduced to 10.7 percent. While this option provided considerable improvement in minimizing the percentage of children experiencing both factors, the number of children within these subpopulations (28) was approximately half the size of those identified under the existing processes.

DCW staff and IMPAQ reviewed the two proposed options and their associated benefits and limitations. Both agreed that the first option presented an effective approach while the second option produced too narrow of a focus. Accordingly, DCW expressed an interest in using the first option as a complementary approach to the existing focus on children and youth from the three age groups with the highest risk scores. This additional approach would allow DCW to identify children with “elevated risk” of emancipation earlier and leverage additional resources to prevent these children and youth from progressing to the level of “highest risk”.

UPDATING THE ERS PROCESSES TO FOCUS ON SUBPOPULATIONS OF INTEREST

In the next step, IMPAQ updated the SPSS syntax to incorporate the code used to identify children and youth with elevated risk or risk scores that were between one and two standard deviations above the mean risk score. IMPAQ tested the revised syntax to ensure that the code accurately identified the original highest risk subpopulation as well as the additional elevated risk subpopulation. The files were provided to DCW for review and discussion with CDHS leadership. After receiving approval from CDHS leadership, IMPAQ provided the finalized SPSS syntax to DCW for testing.

The resulting output file consisted of an Excel workbook with six worksheets:

1. Tab 1: High Risk, 0-5 Age Group

2. Tab 2: Elevated Risk, 0-5 Age Group
3. Tab 3: High Risk, 6-12 Age Group
4. Tab 4: Elevated Risk, 6-12 Age Group
5. Tab 5: High Risk, 13-17 Age Group
6. Tab 6: Elevated Risk, 13-17 Age Group

DISCUSSION OF THE ASSOCIATED IMPLICATIONS

Over the past 2 years, the ERS algorithms have provided an effective approach for identifying children and youth with the highest risk of emancipation, thereby allowing CDHS and county stakeholders to identify pertinent policies, practices, and resources that could be leveraged to help obtain permanency. DCW staff noted that one limitation of the current approach's focus on the highest risk subpopulation was that permanency may not be achievable for children and youth with such high levels of risk, despite an enhanced focus and efforts to leverage additional resources for achieving permanency. Staff noted that achieving permanency was especially difficult for youth that were 17 years or older and rapidly approaching the age of emancipation or children and youth that had been in out of home care for 100 or more months.

DCW faced two alternative options for updating the existing ERS processes to identify additional subpopulations with risk scores that would allow DCW and its county partners to leverage available resources at an earlier point in time. The first option involved replacing the current approach of focusing on the highest risk subpopulation with a new focus on subpopulations with elevated risk. While this approach would allow DCW to focus its efforts on children and youth with elevated risk, with most this subpopulation being under 17 and spending less than 100 months in out of home care, this approach would critically shift the focus away from the highest risk subpopulation. This alternative would also involve reducing the focus to a smaller subpopulation, which could prove problematic given the importance of achieving permanency for legally free children. The second option allows DCW to supplement its existing focus on the high-risk subpopulation with an additional focus on the elevated risk subpopulation. This alternative allows DCW to maintain its current focus on children and youth with the highest risk of emancipation while simultaneously focusing on a subpopulation that has lower risk levels but a high probability of progressing toward the highest level of risk and, ultimately, emancipating from the child welfare system. This approach would considerably expand the number of children and youth who receive an enhanced focus, while also striking an effective balance between focusing on children and youth with elevated and the highest levels of risk. Under this approach, DCW can continue to assess pertinent policies, practices, and resources that could be leveraged to achieve permanency for children and youth with the highest risk of emancipation while proactively engaging in efforts to leverage resources for children and youth with elevated risk. This new approach offers the opportunity for DCW to leverage resources at an earlier point in time during a legally free child's case to mitigate the likelihood of those with elevated levels of risk transitioning to the highest level of risk.

With this new approach, DCW can effectively target its resources toward legally free children and youth with the highest levels of risk and those who are most likely to benefit from additional resources that promote permanency. The expanded focus on the elevated risk subpopulation will allow DCW to focus on children with elevated levels of risk at an earlier point in time, with the goal of preventing a greater number of legally free children and youth from progressing to the highest level of risk. As these efforts prove effective, DCW will want to monitor changes in the size and composition of the elevated and high-risk subpopulations over time, as significant change in these subpopulations could warrant the need to shift the focus toward other subpopulations with differing risk profiles.

REFERENCES

- Elgin, D. J., Sushinsky, J., Johnson, A., Russo, G., & Sewell, T. (2015). Factors affecting permanency for legally free children & youth: A study of Colorado's legally free population across age groups, 2008–2014. *Children and Youth Services Review*, 57, 60-67.
- Keller, T. E., Cusick, G. R., & Courtney, M. E. (2007). Approaching the transition to adulthood: Distinctive profiles of adolescents aging out of the child welfare system. *Social Service Review*, 81(3), 453-484.

APPENDIX A: DETAILED OVERVIEW OF THE ERS ALGORITHMS

Risk Factors by Age Group			
<1-5			
Risk Factors	% Effect	# of Instances	Risk Score
1. African American	0.23	N/A	0.23
2. # of Involvements before Removal	0.19	#	0.19*#
3. # of Months in Congregate Care	0.04	#	0.04*#
4. # of Months in Family-Like Setting	0.03	#	0.03*#
Combined Risk Score = 0.23AfricanAmerican + 0.19#Involvements + 0.04#MonthsCongregate + 0.03#MonthsFamilyLike			
<i>Baseline Risk Score for Higher-Risk Children= 0.49</i>			
6-12			
Risk Factors	% Effect	# of Instances	Risk Score
1. African American	0.36	N/A	0.36
2. Male	0.17	N/A	0.17
3. # of involvements before removal	0.09	#	0.09*#
4. OPPLA Goal	0.97	N/A	0.97
5. # of Placements	0.06	#	0.06*#
6. # of Months in Congregate Care	0.02	#	0.02*#
7. # of Months in Family-Like Setting	0.02	#	0.02*#
Combined Risk Score = 0.36AfricanAmerican + 0.17Male + 0.09#ofInvolvements + 0.97OPPLA + 0.06#Placements + 0.02MonthsCongregate + 0.02MonthsFamilyLike			
<i>Baseline Risk Score for Higher-Risk Children=1.71</i>			
13-17			
Risk Factors	% Effect	# of Instances	Risk Score
1. OPPLA Goal	0.99	N/A	0.99
2. # of Placements	0.07	#	0.07*#
3. # of Months in Family-Like Setting	0.01	#	0.01*#
4. Age (Positive Age Effect)	-0.30	#	-0.30*#
Combined Risk Score = 0.99OPPLA + 0.07#Placements + 0.01#FamilyLike - 0.30Age			
<i>Baseline Risk Score for Higher-Risk Youth= 0.77</i>			
<i>Note: Due to the positive effect of age on permanency within the 13-17 age group (30%), age is calculated in an inverse manner by multiplying the effect of age (-0.30) by the age. Because the age value for this group is large (13 to 17), multiplying this number by a negative value tends to skew the overall Emancipation Risk Score towards a negative number. Accordingly, the Emancipation Risk Scores for this group often have negative values. Youth within this age group with the greatest risk of emancipation will have Emancipation Risk Scores that are closest (or greater than) a value of zero.</i>			