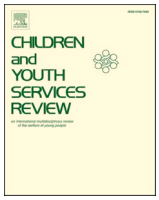
An evaluation of the racial equity of the actuarial Family risk assessment instrument used in Queensland, Australia

Children and Youth Services Review 164 (2024) 107891

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth



An evaluation of the racial equity of the actuarial Family risk assessment instrument used in Queensland, Australia

Brian Q. Jenkins *, Clare Tilbury

Griffith Criminology Institute, Griffith University, Queensland, Australia

ARTICLE INFO

Keywords:
Risk assessment
Equity
Race
SDM
Actuarial

ABSTRACT

Background: Risk assessment instruments used in child protection to guide intervention decisions should be equitable. To be considered equitable, two criteria are necessary: instruments should avoid discriminatory logic and display comparative validity for children of different races. Empirical evidence about the equity of risk assessment instruments is limited. Objective: To investigate the racial equity of a widely used risk assessment instrument, the Structured Decision Making Family Risk Assessment (FRA). Participants and setting: Administrative data were obtained for all children subject to investigation in Queensland, Australia in 2018, for whom an FRA was completed (n = 17,851).

Methods: Relationships between FRA items, race, and recurrence (subsequent investigation within 12 months) were explored using crosstabulations and Cox Proportional Hazards. Accuracy of FRA recommendations and practitioner decisions were compared using sensitivity and specificity for Indigenous and non-Indigenous children.

Results: Most FRA items predicted race better than recurrence. Differences in recurrence by race were only partially explained by FRA items. The FRA produced high rates of false positives for Indigenous children (incorrectly classifying 48.9% of children who did not recur compared to 31.7% for non-Indigenous children) and high rates of false negatives for non-Indigenous children (incorrectly classifying 59% of children who did recur compared to 39% for Indigenous children). Practitioner discretion did not mitigate inequitable FRA recommendations.

Conclusions: The FRA failed both criteria of equity, discriminating indirectly by relying heavily on factors correlated with race, and demonstrating less accuracy for Indigenous children. High rates of false positives for Indigenous children exacerbate Indigenous overrepresentation in child protection.

1. Introduction

Risk assessment instruments are used extensively in child protection practice (Brown & Packard, 2012; Mickelson et al., 2017). When children are assessed to be at unacceptable risk of harm caused by future maltreatment, the state has justification and indeed an obligation to intervene (Johnson, 2006). The assessment of risk and the instruments that guide such assessments are therefore integral to modern child protection work.

The Structured Decision Making (SDM) Family Risk Assessment (FRA) is one of the most widely used and frequently studied risk assessment instruments used in child protection, used in many states in the USA and other jurisdictions internationally (Bosk, 2018). The FRA is an actuarial (i.e., data driven) instrument developed by the National Council on Crime and Delinquency's Children's Research Center (now called Evident Change). The construction methodology for the FRA is set out by Coohey et al. (2013). The FRA uses correlations observed in large data sets to draw conclusions about individuals. As it is not possible to directly measure the distribution of maltreatment in the population, actuarial tools like the FRA draw on data about children subject to child protection involvement, such as a report, investigation, or substantiation, and identify correlates of child protection recurrence (a subsequent episode of child protection involvement), such as a re-report, re-investigation, or re-substantiation. The FRA uses factors associated with recurrence as the items on two indices, one for abuse and one for neglect, with points assigned for each factor observed in a case. Points are then

* Corresponding author at: Griffith Criminology Institute, Griffith University, Mt Gravatt Campus, Messines Ridge Road, Upper Mount Gravatt, Queensland 4122, Australia.

E-mail address: brian.jenkins@griffith.edu.au (B.Q. Jenkins).

https://doi.org/10.1016/j.childyouth.2024.107891
Received 25 August 2022; Received in revised form 10 July 2024; Accepted 28 August 2024
Available online 29 August 2024

summed to give children a score, which is converted to a risk rating (e.g. low, moderate or high) for each scale. The child is then assigned the higher of the two risk ratings. Risk ratings correspond to a recommendation to intervene, for example, no intervention for low- and moderate-risk cases and intervention for high-risk cases. Tools are refined periodically using local administrative data so there is some variation in the items, scoring, risk rating categories, and cut-off scores used in different jurisdictions. However, the items used on different versions of the FRA are generally similar (Coohey et al., 2013; Jenkins et al., 2018; Johnson & Bogie, 2009; Johnson et al., 2015; Wicke Dankert & Bogie, 2015). Actuarial tools like the FRA are often described as more accurate than consensus tools, which are developed by synthesising theory and evidence (Baird & Wagner, 2000; Bosk, 2018; van der Put et al., 2017). However, it should be noted that these studies show that actuarial tools are superior only in predicting subsequent child protection involvement, which is used as a proxy measure of child maltreatment.

Risk assessment tools need to do more than assign accurate risk ratings. Risk assessment instruments must also produce consistent results when used by different workers and over time (D'Andrade et al., 2008) and influence practitioners' decisions in meaningful and positive ways (Johnson, 2011; Russell, 2015). Risk assessment instruments must also operate in a way that is fair or equitable (Jenkins, 2021; Keddell, 2019) particularly as applied to children of different races (Baird et al., 1999; Coohey et al., 2013; Russell, 2015). Given the wide racial disparities in child protection involvement and outcomes across jurisdictions (Children's Bureau of the US Department of Health and Human Services, 2019; Hunter et al., 2020; Hyslop, 2017), interest in the ways risk assessment instruments may exacerbate or mitigate racial inequity are justified.

Equity has been conceptualised as having two dimensions: avoidance of prejudice and comparative validity (Jenkins, 2021; Slobogin, 2018). Firstly, instruments should not use observations about group averages or tendencies to draw conclusions about individuals in that group. For example, it would be prejudicial to assess an Indigenous child as high risk and assess an otherwise identical non-Indigenous child as low risk, simply because child protection recurrence is more common among Indigenous children. As Keddell (2019) notes, people should be treated as individuals, not judged by the conduct of otherwise unrelated people who happen to share a demographic characteristic. Secondly, comparative validity means tools are inequitable if they provide risk ratings for one group that are consistently less accurate than for other groups, for example, if an instrument was less accurate for Indigenous children compared to non-Indigenous children. As Slobogin (2018) noted, avoidance of prejudice and comparative validity may operate in opposition. For example, it may be possible to improve the accuracy of an instrument by using race as an item on a risk assessment tool or a variable in a predictive algorithm. Ideally, tools would meet both criteria but, in practice, it may be necessary to reach a balance between them.

The FRA has been described as equitable because it avoids unconscious practitioner bias by focusing on a set of objective family and case characteristics (Baird, 2005). However, no studies have directly examined whether the FRA does better at avoiding prejudice than practitioners' unstructured assessments. Evidence about whether it produces recommendations that are equally valid for children of different races is limited. Baird (2005) examined the comparative validity of three versions of the FRA used in Michigan, California, and Georgia, and Coohey et al. (2013) evaluated two different versions of the same tool used in Iowa. These studies were authored or co-authored by an affiliate of the Children's Research Center (now known as Evident Change), which developed and market the FRA. The studies found that rates of recurrence for African American and White children within assigned risk categories were comparable. For example, in Michigan, 28 % of African American children classified as very high risk were subject to a subsequent substantiation within 12 months compared to 30 % of White children classified as very high risk (Baird, 2005). However, in Baird's (2005) study, measures of comparative validity were only presented for two out of the three jurisdictions (California and Michigan), and data were not presented for Hispanic children who comprised the largest racial minority in the Californian data. Coohey et al. (2013) included Latino children in their analysis but there were too few cases to draw firm conclusions. It is therefore difficult to draw conclusions about overall comparative validity of the FRA used in these jurisdictions.

Further, in measuring comparative validity, evaluations relied on child protection recurrence, including subsequent reports and substantiations, as a proxy measure for future maltreatment. The relationship between child protection recurrence and maltreatment is not straightforward and the validity of recurrence as a proxy measure of maltreatment has been called into question (Jenkins et al., 2019; Keddell, 2019). These analyses assume not only that recurrence is a valid measure, but that it is equally valid for children of different racial groups, an assumption made without supporting argument or evidence.

Another important limitation of these studies is that they make comparisons between levels of recurrence within each risk category. Such a comparison is valid only when the overall base rates of recurrence are the same for children in both racial groups (Fawcett, 2006; Tharwat, 2021; Jenkins, 2021). To make comparisons between groups with unequal overall rates of recurrence, measures of both sensitivity (the proportion of children who required intervention who were correctly identified as high risk) and specificity (the proportion of children who did not require intervention who were correctly identified as low risk) should be compared across racial groups (Fawcett, 2006; Tharwat, 2021; Jenkins, 2021). While overall rates of recurrence were equivalent for African American children and White children in the Iowa data, rates of recurrence were not equivalent for Hispanic children (Coohey et al., 2013). Overall rates of recurrence were not provided for Michigan or California (Baird, 2005), although some previous studies have found equivalent rates of recurrence for Black and White children in the USA (Bae et al., 2010; Chiang et al., 2020; White et al., 2015).

Johnson (2005) explored relationships between risk assessment items and race for the FRA used in California, although no direct test of comparative accuracy was presented. Findings showed that while most individual items included on the FRA were associated with race, overall scores and risk ratings were only slightly higher for White children compared to African American and Latino children. This suggested limited indirect discrimination producing higher risk ratings for White children. Analyses were also presented to determine whether relationships between items and subsequent substantiation were "spurious", that is, if race was an underlying explanatory variable for risk factors and for recurrence. Actuarial risk assessments frequently include items that have no theoretically plausible causal relationship to increased rates of maltreatment, such as prior periods of service provision (Coohey et al., 2013; Jenkins et al., 2018; Johnson, 2005). However, this analysis is also relevant to the question of whether items operate in an equivalent way for children of different races. Correlations between items and recurrence remained unchanged after controlling for race, which is consistent with items predicting recurrence independent of race. However, there was no analysis of types of error. Risk ratings and the underlying risk factors produced the same number of errors for children of different races, but it is possible that the FRA systematically overestimated risk for children of one racial group, while systematically underestimating the risk for others. While these studies suggest complicated relationships between risk factors, race, and recurrence, given the methodological limitations, the overall equity of this version of the FRA also remains unclear.

While there is limited empirical evidence about racial equity of the FRA, problems with racial equity have been identified with other actuarial risk assessments. For example, in a widely publicised report by ProPublica, the COMPAS tool, which aimed to predict recidivism in criminal justice contexts, was found to be racially biased, displaying high rates of false positives for Black defendants (Angwin et al., 2016). While the tool's makers published their own analysis in rebuttal (Dieterich et al., 2016), this was not peer-reviewed and used metrics that

were unsuitable given the large differences in overall rates of recidivism for White and Black defendants (see Fawcett, 2006; Tharwat, 2021; Jenkins, 2021). Subsequent analysis has confirmed not only that COMPAS displays unacceptable differences in error rates for White and Black defendants (Chouldechova, 2017), but also that, overall, the instrument performs little better than predictions made by people with little or no criminal justice experience (Dressel & Farid, 2018). The issues with COMPAS highlight not only the possibility that actuarial risk assessments can produce inequitable outcomes, but that it is important for equity to be evaluated independently using appropriate metrics.

1.1. Risk assessment in Queensland, Australia

Since 2006, a version of the structured decision-making Family Risk Assessment (referred to as the Family Risk Evaluation locally) has been used state-wide in Queensland, Australia. Notwithstanding the positive evaluations of the racial equity of the Family Risk Assessment in the USA (Baird, 2005; Coohey et al., 2013), the equity of the FRA for Indigenous children in Australia has not been examined. The over-representation of Indigenous families in the child protection system in Australia is a major policy concern, related to historical removals of children from their parents (known as "the stolen generations") and ongoing social and economic disadvantage (Human Rights and Equal Opportunity Commission, 1997). Australian Indigenous children were the subjects of 33.3 % of child protection investigations in Australia in 2019–20, while being only 5.9 % of the child population (Steering Committee for the Review of Government Service Provision, 2021). Disproportionality ratios are very high: in Queensland in 2019–20, Indigenous children were 4.97 times more likely to be notified for maltreatment; 5.27 times more likely to be substantiated; and 6.52 times more likely to be in out-of-home care than non-Indigenous children (Steering Committee for the Review of Government Service Provision, 2021). To provide international context to the issue of Indigenous overrepresentation, in Canada there were 150 per 1000 maltreatment-related investigations for First Nations children compared to 42 per 1000 for non-Indigenous children, i.e. First Nations children were 3.6 times more likely to be investigated (Fallon et al., 2021). The situation in the USA varies according to State demography and policies, with Native North American children underrepresented in reports for physical and sexual abuse in 38 states and emotional abuse in 34 states, while overrepresented in reports for physical and sexual abuse in eight states and for emotional abuse in seven states. Alaska had the highest disparity ratios for maltreatment reports, depending on type of abuse, ranging from 3.20 to 5.18 (Luken et al., 2021). As in other jurisdictions, items on the FRA are known to be correlated with race, with several items on the tool known to have a disproportionate impact on Indigenous communities (e.g., past child protection involvement both for parents and children and parental criminal justice involvement), raising concerns about indirect discrimination (Harnett & Featherstone, 2020). Even though race is not explicitly included as an item on the FRA, it may nevertheless provide inflated risk scores for Indigenous children by targeting ostensibly neutral demographic and personal factors that are highly correlated with race.

Comparative validity cannot be assumed, because base rates of recurrence for Indigenous and non-Indigenous children are different (Jenkins et al., 2018). In Queensland, Indigenous children have been subject to such high rates of recurrence compared to non-Indigenous children that in a study of 9,608 children subject to child protection investigation, Indigenous status was a better predictor of a subsequent report and subsequent investigation than a rating of high risk on the FRA (Jenkins et al., 2018). It would not be surprising to find that the factors associated with recurrence were more concentrated in Indigenous populations, due to high levels of inequality. It may be hypothesised that Indigenous children recur at higher rates due to higher concentrations of risk factors. If so, the inclusion of these factors as items on an instrument could result in comparative accuracy for both Indigenous and non-Indigenous children. However, there is no empirical evidence to support this hypothesis because, to date, studies have focused on comparisons between groups with either equivalent or unreported rates of recurrence.

In response to concerns about the over-representation of Indigenous families, the child protection authority in Queensland provides practice advice to workers to consider cultural factors when making decisions about risk and the need for intervention. While it is mandatory for practitioners to complete the FRA in child protection investigations, practitioners can use discretion, supported by professional judgement, to either follow or deviate from the recommendation the tool provides. Practitioners are instructed to consider the impacts of historical injustices related to colonisation, ongoing disadvantage faced by Indigenous communities, and differences in child-rearing practices and family structure to "apply a cultural lens to every structured decision-making (SDM) tool and risk assessment" (Department of Child Safety, Youth Justice and Multicultural Affairs, 2021a). Frameworks for integrating actuarial assessment and clinical judgement usually allow for the upgrading of risk ratings in clear cases of danger but, otherwise, clinical judgement is restricted to subsequent decisions about how to intervene once the decision to intervene has been made (Schwalbe, 2008; Shlonsky & Wagner, 2005). As Bosk and Feely (2020) point out, there is an epistemological incompatibility between advice to consider information about the nature and causes of maltreatment in particular communities, and an actuarial instrument that operates purely on statistical correlations and therefore treats causality as irrelevant.

There is currently no empirical evidence about whether directions to use a "cultural lens" result in differences in the way practitioners apply discretion for Indigenous and non-Indigenous children, or whether this supports or undermines comparative validity. Cheng et al. (2022) investigated whether practitioner discretion improved or exacerbated racial bias in the Allegheny Family Screening Tool, which is used to guide decisions about whether to investigate child protection reports. They reported that practitioner discretion reduced disparities in screen-in rates and accuracy and concluded that discretion therefore helped to address racial bias. However, some results suggest discretion may have worsened racial bias. For Black children, practitioners were more likely to upgrade risk (i.e. to screen in reports the algorithm recommended be screened out), while for White children, practitioners were more likely to downgrade risk (i.e. to screen out reports the algorithm recommended be screened in). While this resulted in the total number of errors for Black and White children to be closer, this appears to have been achieved in large part by producing more false negatives for White children. It is also noted that practitioner discretion resulted in decisions that were, overall, less accurate for both Black and White children. Given this, it is difficult to conclude that practitioner discretion made a positive difference to the racial bias inherent in the Allegheny Family Screening Tool. It is not known whether a similar pattern would be observed at the close of an investigation in an Australian context.

1.2. Aim and research questions

The aim of this study was to investigate the equity of the FRA, including avoidance of prejudice and comparative validity, by understanding how the FRA predicts child protection recurrence for Indigenous and non-Indigenous children. Specifically, it considered:

1. Are the items on the FRA associated with Indigenous status?
2. Do FRA items explain the difference in recurrence for Indigenous and non-Indigenous children?
3. Does the FRA predict recurrence with equivalent accuracy for Indigenous and non-Indigenous children?
4. Does practitioner discretion affect comparative accuracy of decisions?

Despite the previously noted criticisms of recurrence as a valid proxy

for maltreatment, it remains in widespread usage as an indicator of maltreatment (Fluke et al., 2018), and continues to underpin actuarial risk assessments including the FRA (Coohey et al., 2013). It was beyond the scope of the study to add to the growing literature about the appropriateness of recurrence as a measure of maltreatment. Rather, the study aimed to test the extent to which the FRA achieves its own stated aims of accurately, equitably, and consistently classifying children at risk of child protection recurrence.

2. Methods

2.1. Design

The study used a longitudinal panel design. Administrative data were accessed for all children reported to the child protection authority in 2018 and subject to investigation that included a completed FRA. Children were followed for a standardised one-year period following this initial investigation (i.e. the same observation period was applied to all children, regardless of when during 2018 they were initially reported). A one-year period was considered sufficient given that prior research has consistently demonstrated that most recurrence takes place soon after the initial event and factors associated with recurrence do not change with longer timeframes (Bae et al., 2010; DePanfilis & Zuravin, 1999; Jenkins et al., 2018). A longer observation period was considered less desirable as this would have reduced the recency of findings without meaningful improvement in analysis. Data from initial investigations were used to identify relationships between FRA items and Indigenous status. Data about subsequent child protection investigations were used to determine rates of recurrence for Indigenous and non-Indigenous children over 12 months, whether items from the initial FRA explained differences in rates of recurrence for Indigenous and non-Indigenous children, whether the initial FRA classified Indigenous and non-Indigenous children accurately, and whether decisions to follow or deviate from recommendations resulted in improved accuracy. Griffith University Human Research Ethics Committee granted approval for the project on 16/12/2020 (GU Ref No: 2020/964).

2.2. Sample

The sample consisted of all children who were reported to the child protection authority in 2018 who were subject to an investigation with the mandatory FRA completed. This timeframe was selected to ensure the follow-up period for all children would be complete by the end of 2019 as the effects of the Covid-19 pandemic from 2020 onwards were not yet understood. The sample included 17,851 children.

2.3. Variables

2.3.1. Indigenous status

In Queensland, it is mandatory for practitioners to record Indigenous status for all children as there are explicit requirements to consider cultural factors when contacting families and conducting an investigation. Children recorded as Aboriginal, Torres Strait Islander, and both Aboriginal and Torres Strait Islander were classified as Indigenous. All other children were classified as non-Indigenous.

2.3.2. FRA risk rating and recommendations

The version of the FRA used in Queensland classifies children as low-, moderate-, and high-risk. The FRA recommends intervention for children classified as high risk and recommends case closure for children classified as low or moderate risk. The risk rating was used to construct a binary variable for intervention recommendation (intervention recommended or no intervention recommended).

2.3.3. FRA scales and items

The FRA includes a scale for abuse and a scale for neglect. Scores on the abuse scale ranged from −1 to 14. Scores on the neglect scale ranged from 0 to 13. Variables for individual items that comprised each scale were also included. Several items were included in both the abuse scale and the neglect scale, so these were collapsed into single variables indicating that the item had been selected on either the abuse scale or the neglect scale.

The following 8 items were included on both the abuse and neglect scales: number of children in household (one to three, or four or more); prior intervention with household (yes or no); primary parent has a criminal history as adult or juvenile (yes or no); primary parent was subject to abuse or neglect as a child (yes or no); primary parent had mental health problem in the last 12 months (yes or no); primary parent had mental health problem prior to last 12 months (yes or no); primary parent had drug or alcohol problem in last 12 months (yes or no); primary parent had drug or alcohol problem prior to last 12 months (yes or no).

There were 10 items included exclusively on the abuse scale: current investigation substantiated for abuse (yes or no); prior investigations related to abuse (none, one, or two or more); past child injury (yes or no); primary parent justifies abuse or neglect (yes or no); two or more domestic violence incidents in the past year (yes or no); primary parent provides insufficient emotional or psychological support (yes or no); primary parent employs excessive or inappropriate discipline (yes or no); domineering parent (yes or no); secondary parent had drug or alcohol problem in last 12 months (yes or no); secondary parent had drug or alcohol problem prior to last 12 months (yes or no).

There were 10 items included exclusively on the neglect scale: current report alleges neglect (yes or no); number of prior investigations (none, one, or two or more); prior placement in out-of-home care (yes or no); age of youngest child (under two, or two or older); child with developmental or physical disability (yes or no); child was medically fragile or displayed failure to thrive (yes or no); positive toxicology screening at birth (yes or no); family is homeless (yes or no); housing is physically unsafe (yes or no); primary parent provides physical care inconsistent with child needs (yes or no).

2.3.4. Practitioner assessment

After considering the recommendation of the FRA and other information available, practitioners are required to use professional judgement to classify children as in need of protection (triggering a statutory intervention) or not in need of protection. Note that children classified as not in need of protection are not subject to statutory intervention but may be referred to voluntary services.

2.3.5. Recurrence and time to recurrence

Recurrence was defined as a subsequent investigation within 12 months of the initial investigation, indicated by the presence of a subsequent mandatory FRA. Time to recurrence was measured as the number of days between the completion of the initial FRA and the completion of the subsequent FRA.

2.4. Analysis

2.4.1. RQ1 – Association between FRA items and Indigenous status

Cramer's V was used to calculate associations between Indigenous status and FRA items. Cramer's V represents the effect size and can vary from 0, indicating no association, to 1, indicating perfect correlation. The test is appropriate for measuring correlations between categorical variables. For binary variables, it is equal to Phi. Chi squared tests were used to determine statistical significance.

2.4.2. RQ2 – Associations between Indigenous status, FRA items, and recurrence

Bivariate associations between recurrence (yes/no) and individual items were measured using Cramer's V. Survival analysis was used to compare rates of recurrence for Indigenous and non-Indigenous children

over the 12-month observation period. Cumulative survival curves were plotted and a Cox Proportional Hazards regression was used to quantify the disparity in recurrence rates over the entire 12-month period. Control variables were then introduced into the Cox Proportional Hazards model to understand how much of the difference between the two groups could be explained by differences in risk factors. Three models with different control variables were tested: the recommendation of the FRA to intervene or close the case; the numeric scores from the abuse scale and the neglect scale; the individual items included on the FRA.

2.4.3. RQ3 – Comparative accuracy of the FRA for Indigenous and non-Indigenous children

Comparisons of accuracy must take into account two types of correct predictions (a recommendation to intervene for a child who did recur, referred to as a true positive, and a recommendation to not intervene for a child who does not recur, referred to as a true negative) and two types of errors (a recommendation to intervene for a child who does not recur, referred to as a false positive, and a recommendation to not intervene for a child who does recur, referred to as a false negative). It is important to consider each type of error separately as consequences are different; false negatives represent missed opportunities for intervention, while false positives represent unnecessary interventions. Measures must also be meaningful when making comparisons between groups with unequal overall rates of recurrence. Two measures that meet these criteria are sensitivity (true positives / all positives) and specificity (true negatives / all negatives) (Fawcett, 2006). A comparison of sensitivity and specificity is therefore the recommended approach to comparing validity of instruments that provide a recommendation to intervene or not intervene (Jenkins, 2021). Chi squared statistics were calculated to determine significance.

2.4.4. RQ4 – Comparative of accuracy of practitioner decisions for Indigenous and non-Indigenous children

First, crosstabulations were used to examine the concordance between FRA recommendations and practitioner decisions to provide intervention. It was important to take into account the different consequences of using discretion to reject advice to open a case, which may improve specificity but may degrade sensitivity, and the consequences of rejecting advice to close a case, which may improve sensitivity but may degrade specificity. Given this, crosstabulations were run for practitioner decisions by Indigenous status first for cases in which the FRA recommended intervention and second for cases the FRA recommended case closure. Chi squared statistics were used to determine significance. Second, the accuracy of practitioner decisions were compared between Indigenous and non-Indigenous children. This involved using crosstabulations to calculate and compare the sensitivity and specificity of practitioner decisions.

3. Findings

Table 1 sets out the descriptive statistics for variables included in the study. Slightly over one-third of the children sampled were Indigenous and, overall, slightly under one-quarter (23.1 %) of children subject to investigation were subject to at least one subsequent investigation within 12 months. The FRA classified just under one-quarter of children as low risk (23.1 %), around one-third as moderate risk (36.7 %) and the remaining 40.2 % of children as high risk with intervention recommended. FRA items most commonly indicated were prior investigations (18.1 % had one prior investigation and 49.8 % had two or more investigations; 17.7 % had one prior investigation related to abuse and 43.4 % had two or more prior investigations related to abuse), current report alleges neglect (41.5 %), and primary parent has criminal history as adult or juvenile (40.6 %). Items indicated least frequently were positive toxicology screening at birth (0.4 %), domineering parent (1.3 %), and housing is physically unsafe (1.5 %).

Associations between individual items of the FRA and Indigenous status are displayed in Table 2, alongside associations between FRA items and recurrence (i.e. a subsequent investigation within 12 months of the initial investigation). Out of 28 items, 21 were indicated significantly more frequently for Indigenous children than for non-Indigenous children. For these items, Cramer's V was modest but statistically significant (maximum Cramer's V of 0.192 for number of prior investigations). Relationships between FRA items and recurrence were similar with most items indicated more frequently among children who

Table 1
Descriptive statistics.

	n (mean)	% (standard deviation)
Indigenous status		
Indigenous	6,276	35.2 %
non-Indigenous	11,575	64.8 %
Recurrence		
Subsequent investigation within 12 months	4,132	23.1 %
No subsequent investigation within 12 months	13,719	76.9 %
FRA risk rating		
Low (no intervention recommended)	4,118	23.1 %
Moderate (no intervention recommended)	6,559	36.7 %
High (intervention recommended)	7,174	40.2 %
FRA risk scores		
Abuse scale	(3.84)	(3.09)
Neglect scale	(4.67)	(2.82)
FRA risk items		
Four or more children in household	5,954	33.4 %
Prior intervention with household	5,682	31.8 %
Primary parent has criminal history as adult or juvenile	7,250	40.6 %
Primary parent was subject to abuse or neglect as a child	6,452	36.1 %
Primary parent had mental health problem in last 12 months	5,540	31.0 %
Primary parent had mental health problem prior to last 12 months	6,236	34.9 %
Primary parent had drug or alcohol problem in last 12 months	5,340	29.9 %
Primary parent had drug or alcohol problem prior to last 12 months	7,376	41.3 %
Current investigation substantiated for abuse	2,081	11.7 %
Prior investigations related to abuse		
One	3,161	17.7 %
Two or more	7,748	43.4 %
Past child injury	2,042	11.4 %
Primary parent justifies abuse or neglect	1,200	6.7 %
Two or more domestic violence incidents in the past year	6,932	38.8 %
Primary parent provides insufficient emotional or psychological support	1,769	9.9 %
Primary parent employs excessive or inappropriate discipline	845	4.7 %
Domineering parent	238	1.3 %
Secondary parent had drug or alcohol problem in last 12 months	3,539	19.8 %
Secondary parent had drug or alcohol problem prior to last 12 months	4,072	22.8 %
Current report alleges neglect	7,409	41.5 %
Number of prior investigations		
One	3,228	18.1 %
Two or more	8,890	49.8 %
Prior placement in out-of-home care	3,247	18.2 %
Youngest child under two years	5,830	32.7 %
Child with developmental or physical disability	2,700	15.1 %
Child was medically fragile or displayed failure to thrive	336	1.9 %
Positive toxicology screening at birth	77	0.4 %
Family is homeless	744	4.2 %
Housing is physically unsafe	262	1.5 %
Primary parent provides physical care inconsistent with child needs	1,373	7.7 %

Table 2
Associations between FRA items and Indigenous status and between FRA items and recurrence.

	Proportion of children with item indicated		Cramer's V/ Phi	Proportion of children with item indicated		Cramer's V/ Phi
	Indigenous children (%)	Non-Indigenous children (%)		Children who recurred (%)	Children who did not recur (%)	
Four or more children in household	39.4 %	30.1 %	0.094***	35.7 %	32.6 %	0.027***
Prior intervention with household	42.4 %	26.1 %	0.167***	39.4 %	29.5 %	0.089***
Primary parent has criminal history as adult or juvenile	51.3 %	34.8 %	0.160***	47.4 %	38.6 %	0.076***
Primary parent was subject to abuse or neglect as a child	45.6 %	31.0 %	0.145***	42.1 %	34.3 %	0.068***
Primary parent had mental health problem in last 12 months	28.1 %	32.6 %	0.047***	33.9 %	30.2 %	0.034***
Primary parent had mental health problem prior to last 12 months	31.6 %	36.8 %	0.052***	38.7 %	33.8 %	0.044***
Primary parent had drug or alcohol problem in last 12 months	36.5 %	26.4 %	0.105***	35.9 %	28.1 %	0.072***
Primary parent had drug or alcohol problem prior to last 12 months	50.0 %	36.6 %	0.130***	49.2 %	38.9 %	0.088***
Current investigation substantiated for abuse	12.5 %	11.2 %	0.019*	11.0 %	11.9 %	0.012
Prior investigations related to abuse						
One	16.8 %	18.2 %		16.8 %	18.0 %	
Two or more	53.3 %	38.0 %	0.154***	53.1 %	40.5 %	0.112***
Past child injury	13.8 %	10.2 %	0.055***	13.3 %	10.9 %	0.031***
Primary parent justifies abuse or neglect	6.6 %	6.8 %	0.004	5.3 %	7.1 %	0.031***
Two or more domestic violence incidents in the past year	45.2 %	35.4 %	0.096***	43.1 %	37.6 %	0.048***
Primary parent provides insufficient emotional or psychological support	10.6 %	9.5 %	0.017*	9.1 %	10.2 %	0.015*
Primary parent employs excessive or inappropriate discipline	5.0 %	4.6 %	0.009	4.4 %	4.8 %	0.008
Domineering parent	1.4 %	1.3 %	0.003	1.4 %	1.3 %	0.003

(*continued on next page*)

Table 2 (*continued*)

	Proportion of children with item indicated		Cramer's V/ Phi	Proportion of children with item indicated		Cramer's V/ Phi
	Indigenous children (%)	Non-Indigenous children (%)		Children who recurred (%)	Children who did not recur (%)	
Secondary parent had drug or alcohol problem in last 12 months	25.3 %	16.8 %	0.144***	21.2 %	19.4 %	0.057***
Secondary parent had drug or alcohol problem prior to last 12 months	27.3 %	20.4 %	0.125***	25.2 %	22.1 %	0.069***
Current report alleges neglect	46.6 %	38.8 %	0.076***	46.6 %	40.0 %	0.057***
Number of prior investigations						
One	16.8 %	18.8 %		17.3 %	18.3 %	
Two or more	62.0 %	43.2 %	0.192***	60.1 %	46.7 %	0.122***
Prior placement in out-of-home care	24.1 %	15.0 %	0.113***	22.6 %	16.9 %	0.062***
Youngest child under two years	37.7 %	29.9 %	0.079***	35.7 %	31.7 %	0.036***
Child with developmental or physical disability	13.2 %	16.1 %	0.039***	17.1 %	14.5 %	0.030***
Child was medically fragile or displayed failure to thrive	2.3 %	1.7 %	0.022**	2.2 %	1.8 %	0.013
Positive toxicology screening at birth	0.4 %	0.4 %	0.002	0.3 %	0.5 %	0.010
Family is homeless	5.5 %	3.4 %	0.051***	3.2 %	4.5 %	0.026***
Housing is physically unsafe	2.0 %	1.2 %	0.035***	1.2 %	1.6 %	0.013
Primary parent provides physical care inconsistent with child needs	9.8 %	6.5 %	0.059***	6.5 %	8.1 %	0.025***

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

recurred than among children who did not recur. In most cases, associations between items and recurrence were significant but weaker than the associations observed between items and Indigenous status (maximum Cramer's V 0.122 for number of prior investigations).

Fig. 1 shows comparisons of recurrence over the 12-month observation period for Indigenous children and non-Indigenous children. Corresponding hazard ratios are set out in Table 3. With no control variables, hazard of recurrence was 1.43 times greater for Indigenous children than non-Indigenous children. Controlling for the recommendation of the FRA, plots of hazards were slightly closer together (hazard ratio of 1.33). Controlling for the abuse and neglect scale scores, the hazard ratio was further reduced to 1.28, and controlling for all 28 individual items from the FRA, the hazard ratio was reduced to 1.24. Hazard ratios in all models were statistically significant ($p < 0.001$).

Table 4 shows correspondence between children's recurrence and the recommendation of the FRA and between recurrence and the practitioner's decision to intervene. Figures are displayed separately for Indigenous and non-Indigenous children. For Indigenous children who recurred, the FRA recommended intervention in 61 % of cases (i.e. sensitivity was 61 %). For non-Indigenous children, the FRA's sensitivity was significantly lower at 41 % ($p < 0.001$). For Indigenous children who did not recur, the FRA correctly recommended case closure in 51.1 % of cases (i.e. specificity was 51.1 %). This compared to specificity of 68.3 % for non-Indigenous children, which was significantly higher ($p < 0.001$). Overall, the FRA recommended intervention 52.3 % of the time for Indigenous children and 33.6 % of non-Indigenous children. The FRA made the correct recommendation (i.e. intervention was recommended for children who did recur and not recommended for children who did not recur) 53.9 % of the time for Indigenous children and 62.7 % of the time for non-Indigenous children. Sensitivity of practitioner decisions was lower (20 % for Indigenous children and 14.3 % for non-Indigenous children), while specificity was higher (76.7 % and 82.9 %,

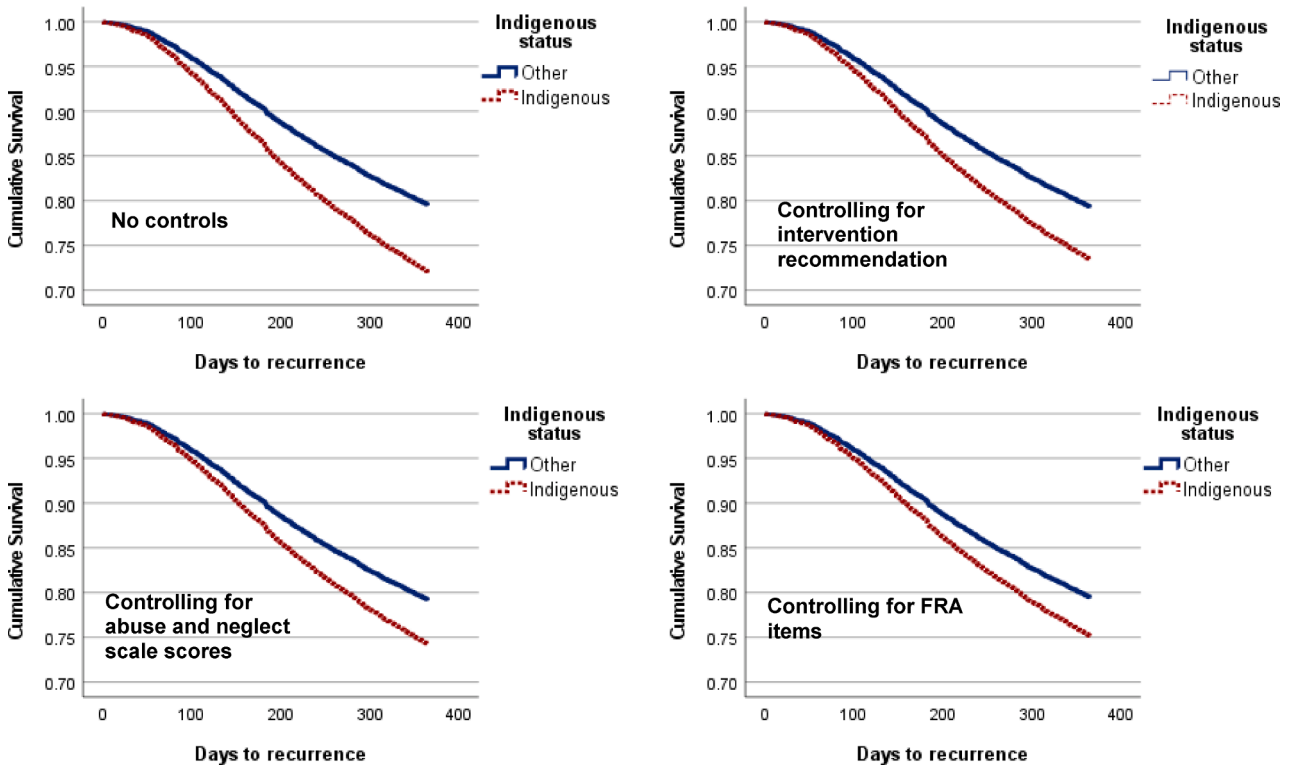


Fig. 1. Cumulative survival curves for time to recurrence by Indigenous status.

Table 3
Recurrence for Indigenous children compared to non-Indigenous children.

	Hazard ratio	Significance
No controls	1.43	0<.001
Controlling for intervention recommendation	1.33	0<.001
Controlling for abuse and neglect scale scores	1.28	0<.001
Controlling for FRA items	1.24	0<.001

respectively). Differences between Indigenous and non-Indigenous children were significant ($p < 0.001$). Overall, practitioners were more likely to open cases for Indigenous children (22.4 %) compared to non-Indigenous children (16.5 %).

Table 5 shows how often practitioners opened a case for intervention when the FRA recommended intervention and when the FRA recommended case closure. Figures are compared for Indigenous and non-Indigenous children. When the FRA recommended an intervention, practitioners provided intervention 37.1 % of the time for Indigenous children and 38.9 % of the time for non-Indigenous children. This difference was not statistically significant. When the FRA recommended case closure, practitioners opened a case for intervention 6.3 % of the time for Indigenous children and 5.2 % of the time for non-Indigenous children. This difference was statistically significant ($p = 0.03$).

4. Discussion

There have been few previous attempts to evaluate the equity of the Family Risk Assessment. While these studies have found the FRA to be equitable for White and Black children in the USA, these evaluations have considerable limitations. Attempts to compare the predictive accuracy of instruments for children of different races have been undermined by methodological problems, particularly the use of measures of comparative accuracy that are not valid for comparisons between groups with different overall rates of recurrence. In most child protection literature, little consideration has been given to the avoidance of prejudice, and it has been difficult to draw conclusions from the limited empirical research on this aspect of equity. The present study is the first independent evaluation of the racial equity of an actuarial child protection risk assessment instrument. Previous studies of the tool, with

Table 4
Accuracy of FRA recommendations and practitioner decisions by Indigenous status of child.

	Indigenous		Non-Indigenous	
	n	%	n	%
Children who recurred				
FRA recommendation**				
FRA recommended intervention	1,071	61.0 % [a]	974	41.0 % [a]
FRA recommended case closure	684	39.0 %	1,403	59.0 %
Total	1,755	100.0 %	2,377	100.0 %
Practitioner decision**				
Practitioner opened case	351	20.0 % [b]	340	14.3 % [b]
Practitioner did not open case	1,404	80.0 %	2,037	85.7 %
Total	1,755	100.0 %	2,377	100.0 %
Children who did not recur				
FRA recommendation**				
FRA recommended intervention	2,209	48.9 %	2,920	31.7 %
FRA recommended case closure	2,312	51.1 % [c]	6,278	68.3 % [c]
Total	4,521	100.0 %	9,198	100.0 %
Practitioner decision**				
Practitioner opened case	1,053	23.3 %	1,575	17.1 %
Practitioner did not open case	3,468	76.7 % [d]	7,623	82.9 % [d]
Total	4,521	100.0 %	9,198	100.0 %

** $p < 0.001$.
[a] . Sensitivity of FRA recommendation.
[b] . Sensitivity of practitioner decision.
[c] . Specificity of FRA recommendation.
[d] . Specificity of practitioner decision.

Table 5
Comparison of concordance between FRA recommendation and practitioner decisions by Indigenous status.

	Indigenous		Non-Indigenous	
	n	%	n	%
FRA recommended intervention				
Practitioner decision				
Practitioner opened case	1,216	37.1 %	1,514	38.9 %
Practitioner did not open case	2,064	62.9 %	2,380	61.1 %
Total	3,280	100.0 %	3,894	100.0 %
FRA did not recommended intervention				
Practitioner decision*				
Practitioner opened case	188	6.3 %	401	5.2 %
Practitioner did not open case	2,808	93.7 %	7,280	94.8 %
Total	2,996	100.0 %	7,681	100.0 %

* $p < 0.05$.

author or co-author affiliations to the developer of the instrument, are subject to possible researcher allegiance effect, whereby outcome variability across studies is associated with the preferences and expertise of the research teams involved (Blair et al., 2008). This effect has been found in fields including psychotherapy interventions (Blair et al., 2008) as well as actuarial risk assessment in forensic psychiatry and criminal justice (Douglas et al., 2017; Singh et al., 2013). This study addresses these limitations and responds to concerns about the fairness of the FRA for Indigenous Australian families and its role in the disproportionate entry of Indigenous children to the child protection system (Hunter et al., 2020).

4.1. Avoidance of prejudice

The FRA fails to meet the first criterion of equity, the avoidance of prejudice. The tool is constructed using explicitly prejudiced logic because it uses factors associated with above-average rates of recurrence in a data set to draw conclusions about individuals with similar demographic and personal characteristics. Children are assessed to be likely victims of maltreatment and parents are assessed to be likely maltreaters by their resemblance to other families who have been subject to repeated child protection involvement in the past. Race is not used as an item on the tool. However, findings demonstrate that most of the items included on the FRA were correlated with race. Out of 28 items, 21 were positively correlated with Indigenous status. Three other items were negatively correlated with Indigenous status (primary parent had a mental health problem in the last 12 months; primary parent had a mental health problem prior to the last 12 months; child with developmental or physical disability). Only four items were statistically independent of race (positive toxicology screening at birth; domineering parent; primary parent justifies abuse or neglect; and primary parent employs excessive or inappropriate discipline), although it should be noted that these items had no value in predicting recurrence either as they were either negatively correlated or unrelated to recurrence. Findings confirmed concerns about the inclusion of items related to contact with systems known to be characterised by Indigenous overrepresentation (Harnett & Featherstone, 2020). The five strongest correlations observed (Cramer's V between 0.145 and 0.192) were all between Indigenous status and prior involvement with child protection and criminal justice systems (number of prior investigations; prior intervention with household; primary parent has a criminal history as adult or juvenile; number of prior investigations related to abuse). These associations were stronger than any observed between FRA items and recurrence.

While effect sizes for associations between Indigenous status and FRA items were generally small, they were statistically and substantively significant. The predictive power of the FRA derives from the cumulative predictive power of many factors that are each only weakly correlated with recurrence. The correlations between risk factors and race accumulate in the same way. All 21 of the items that were positively correlated with Indigenous status were more strongly correlated with race than with recurrence. If these items can be considered sufficiently predictive of recurrence to be included in the FRA, they must also be considered too strongly associated with race for their inclusion to be equitable. If the FRA as a whole is considered useful in predicting recurrence, it must also be considered unacceptably discriminatory.

4.2. Relationships between factors

It is possible for discriminatory logic to improve the accuracy of predictions, depending on the nature of the statistical relationships between risk factors, race, and the outcome being predicted. That is, high rates of recurrence in Indigenous populations may be explained by higher concentrations of risk factors, in which case items associated with race would support accurate predictions for both Indigenous and non-Indigenous children. If so, the instrument may represent an inequitable process that nonetheless produces accurate predictions for Indigenous and non-Indigenous children. However, this hypothesis was not supported by the data. Consistent with previous findings (Jenkins et al., 2018), survival analysis showed that Indigenous children were significantly more likely to be subject to recurrence. Without any controls, over the 12-month period, recurrence was 1.43 times more likely for Indigenous children than non-Indigenous children. Controlling for the recommendation of the FRA to either provide intervention or to close the case, the hazard ratio fell only slightly to 1.33. Subsequent models controlled for the numeric scores on the abuse scale and neglect scale and for each of the 28 unique FRA items. In these models, hazard ratios were reduced to 1.28 and 1.24 respectively, with differences between Indigenous and non-Indigenous children remaining highly significant ($p < 0.001$). Different concentrations of risk factors only offer a partial explanation for the differences in recurrence between Indigenous and non-Indigenous children. This is not surprising. As per the advice given to practitioners (Department of Child Safety Youth Justice and Multicultural Affairs, 2021b), high rates of Indigenous child protection involvement are underpinned by Australia's history of colonisation including direct and indirect discrimination, assimilationist policies, forced removals, ongoing socio-economic disadvantage, intergenerational trauma, and incompatibilities between formal child protection systems and Indigenous family structures and child-rearing practices (O'Donnell et al., 2019; Tilbury, 2015). Child protection involvement is qualitatively different for Indigenous families both in its nature and causes. It is unlikely that a single set of factors could predict, with equal validity, two different phenomena in two different populations.

4.3. Comparative validity

This was confirmed by the comparison of predictive accuracy, which revealed clear differences in the frequency and types of errors the instrument produced for Indigenous and non-Indigenous children. Overall, the FRA made fewer accurate predictions for Indigenous children (53.9 %) than for non-Indigenous children (62.7 %). When types of errors are considered, the differences are starker. For Indigenous children, the FRA was risk averse, recommending intervention in the majority of cases (52.3 %). While sensitivity was high, with the tool correctly recommending intervention for 61 % of Indigenous children who recurred, it also produced many false positives. For Indigenous children who did not recur and who are therefore presumed by the FRA to not require intervention, the accuracy of the recommendation was little better than a coin toss (51.1 % specificity). In contrast, for non-Indigenous children, the FRA has a high threshold for risk, recommending intervention in relatively few cases (33.6 %). For children who did not recur and who are therefore presumed to not require intervention, the FRA provided the correct recommendation more than two-thirds of the time (68.3 %

specificity). However, the rate of false negatives was high. Most children who recurred and were therefore presumed to require an intervention were classified incorrectly (41 % sensitivity). While false positive and false negative errors are both always undesirable for the individual children affected, the distribution of these errors between Indigenous and non-Indigenous children is especially problematic in the Australian context because it would tend to exacerbate the serious and growing problem of Indigenous overrepresentation in child protection (Hunter et al., 2020).

4.4. Effect on decision making

While practitioners frequently deviated from the advice of the FRA, this use of discretion did not mitigate the bias observed in the FRA. When the FRA returned a rating of low- or moderate-risk (i.e. recommended case closure) practitioners almost always followed advice to close the case (93.7 % of the time for Indigenous children and 94.8 % for non-Indigenous children). When the FRA returned a rating of high risk, case workers appeared to consider other factors and used discretion to provide intervention in approximately two out of five cases (37.1 % and 38.9 % for Indigenous and non-Indigenous children respectively). That is, a rating of high risk on the FRA was usually necessary but not sufficient for statutory intervention to be initiated. Despite practice guidance to consider cultural factors when making decisions about Indigenous families, there was minimal difference in the use of discretion for Indigenous children compared to non-Indigenous children. When the FRA recommended intervention, there was no statistically significant difference in case opening for Indigenous and non-Indigenous children. However, when the FRA recommended case closure, practitioners were slightly and statistically significantly more likely to use discretion to initiate intervention for Indigenous children (6.3 % compared to 5.2 % for non-Indigenous children). While the difference was not numerically large, it ran in an unfavourable direction, further exacerbating the tendency of the FRA to classify Indigenous children as high-risk and classify non-Indigenous children as low-risk. Because statutory intervention was initiated in few cases, sensitivity of decisions was low and specificity was high for both Indigenous and non-Indigenous children. However, differences in accuracy remained evident. Sensitivity was significantly higher for Indigenous children compared to non-Indigenous children (20 % compared to 14.3 %) and specificity was significantly lower (76.3 % compared to 82.9 %). That is, decisions were more risk averse for Indigenous children resulting in higher rates of false positives compared to non-Indigenous children who were subject to higher rates of false negatives. Disparities in accuracy of decisions about Indigenous and non-Indigenous children were significant and consistent with disparities in advice provided by the FRA. These differences in relationships between FRA items and recurrence mean that the child protection system has a disparate, more intrusive impact on Indigenous families and communities.

4.5. Limitations

The current study examined differences in the way the FRA operates to predict child protection recurrence for Indigenous and non-Indigenous children. However, the goal of risk assessment is not to predict future involvement with statutory child protection authorities, but to predict future child maltreatment. As has been noted, the value of child protection recurrence as a proxy for child maltreatment has been contested (Jenkins et al., 2019; Keddell, 2019). The current study does not ignore these critiques or intend to endorse child protection recurrence as a measure of child maltreatment. Rather, given that recurrence continues to be used to construct, calibrate, and test risk assessment instruments, the study offers an additional line of inquiry. It contributes to the debate about whether the FRA achieves its ultimate aim of predicting maltreatment equitably by considering whether it can achieve the more modest intermediate aim of predicting recurrence equitably. The more fundamental question of whether child protection recurrence is a valid measure of maltreatment cannot be answered with administrative data alone and is therefore beyond the scope of the current study. However, this remains an important question for proponents of actuarial and algorithmic risk assessment in child protection.

For the same reasons, findings showing higher concentrations of FRA items among Indigenous children subject to investigation say little about priorities for intervention for Indigenous children subject to child protection involvement or in the community more broadly. Factors are included in the FRA based on statistical associations only, so are not necessarily causally related to child protection recurrence, much less to child maltreatment. Furthermore, several FRA items relate to child protection history. Suggestions to prevent future maltreatment or child protection involvement by preventing past child protection involvement are not helpful. Differential patterns of child protection involvement, including child protection recurrence, are important and warrant further research and policy attention. However, relationships between maltreatment, child protection involvement, correlates of child protection involvement, and interventions are not straightforward (Jenkins et al., 2018, 2019).

The findings of the current study relate to Queensland, Australia where differences in rates of recurrence have been observed for Indigenous and non-Indigenous children. Some studies have found race to be associated with recurrence (Brown et al., 2020; Holbrook & Hudziak, 2020), while others have not (Bae et al., 2010; Chiang et al., 2020; White et al., 2015). In jurisdictions where rates of recurrence are similar or equivalent for children of different racial groups, the accuracy of actuarial instrument like the FRA may be less likely to be patterned by race. However, in these places, inequity between other groups subject to differential rates of recurrence may be observed, for example children of different ages, children living in households with different compositions, and of different socio-economic statuses (Esposito et al., 2020; White et al., 2015).

5. Conclusions and future directions

The FRA instrument is not equitable because it systematically directs Australian Indigenous children towards statutory interventions at a higher rate than non-Indigenous children. It does so using items that are more strongly associated with race than with child protection recurrence, which is the outcome it claims to predict. Accuracy is lower for Indigenous children than it is for other children, which means that Indigenous children are more likely to be erroneously classified as high risk. Guidance to practitioners to consider cultural factors for Indigenous families does not mitigate the tool's racial bias. This policy guidance is set within a wider context whereby tool recommendations are characterised as more accurate. The tool provides advice based on purely statistical correlations observed in large data sets, and while practitioners see the final risk rating, they have no sense of the associations between variables that underpin that rating. It is not clear how practitioners could use professional discretion to integrate knowledge about the nature and causes of maltreatment in different communities with this kind of decision recommendation. The bias cannot be eliminated by recalibrating the scoring and cut-offs of the existing instrument because the nature of relationships between the FRA's items and recurrence are different for Indigenous and non-Indigenous children. With Indigenous overrepresentation in child protection intensifying each year in Australia (Hunter et al., 2020), there is an urgent need for the FRA to be replaced with an equitable instrument. In the Queensland child protection authority, a review of assessment and decision-making tools is underway (Queensland Government, n.d.).

Further research is required in other jurisdictions that use versions of the FRA or other actuarial or algorithmic risk assessments. Where there are differences in rates of recurrence for children of different races, data-driven risk assessment instruments are especially likely to display racial bias both in the items included on the instrument and in differential

rates and types of errors. However, the possibility of racial bias in jurisdictions with equivalent rates of recurrence across racial groups should not be discounted. Further consideration should also be given to equity as it pertains to other factors known to be associated with recurrence including children's ages, child and parental disability status, and socio-economic status.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgement

This research was supported by the Queensland Department of Children, Youth Justice and Multicultural Affairs.

References

Angwuin, J., Larson, J., Mattu, S., Kirchner, L., & , P. (2016). *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bae, H., Solomon, P. L., Gelles, R. J., & White, T. (2010). Effect of Child Protective Services System Factors on Child Maltreatment Rereporting. *Child Welfare, 89*(3), 33–55. https://www.cwla.org/child-welfare-journal/.

Baird, C. (2005). The Effect of Risk Assessments and Their Relationship to Maltreatment Recurrence Across Races. In D. Derezotes, M. Testa, & J. Poertner (Eds.), *Race matters in child welfare: the overrepresentation of African American children in the system* (pp. 131–146). CWLA Press.

Baird, C., Ereth, J., & Wagner, D. (1999). Research-Based Risk Assessment: Adding Equity to CPS Decision Making. *Children's Research Center*. http://www.nccdglobal.org/sites/default/files/publication_pdf/race_risk_1999.pdf.

Baird, C., & Wagner, D. (2000). The relative validity of actuarial- and consensus-based risk assessment systems. *Children and Youth Services Review, 22*(11), 839–871. https://doi.org/10.1016/S0190-7409(00)00122-5

Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is There an Allegiance Effect for Assessment Instruments? Actuarial Risk Assessment as an Exemplar [Doi: 10.1111/j.1468-2850.2008.00147.x]. *Clinical Psychology: Science and Practice*, *15*(4), 346-360. Doi: 10.1111/j.1468-2850.2008.00147.x.

Bosk, E., & Feely, M. (2020). The Goldilocks Problem: Tensions between Actuarially Based and Clinical Judgment in Child Welfare Decision Making. *Social Service Review, 94*(4), 659–692. https://doi.org/10.1086/712060

Bosk, E. A. (2018). What counts? quantification, worker judgment, and divergence in child welfare decision making. *Human service organizations, management, leadership & governance, 42*(2), 205–224. https://doi.org/10.1080/23303131.2017.1422068

Brown, K., & Packard, T. (2012). *Review of Child Welfare Risk Assessments*. https://theacademy.sdsu.edu/wp-content/uploads/2015/02/SACHS_Risk_Assessment_Report_and_Appendices_11_2012.pdf.

Brown, S. M., Orsi, R., & Chen, P. C. B. (2020). Child, Family, and Case Characteristics Associated With Reentry Into Out-of-Home Care Among Children and Youth Involved With Child Protection Services. *Child Maltreatment, 25*(2), 162–171. https://doi.org/10.1177/1077559519869395

Cheng, H.-F., Stapleton, L., Kawakami, A., Sivaraman, V., Cheng, Y., Qing, D., Perer, A., Holstein, K., Wu, Z. S., & Zhu, H. (2022). *How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions* Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA. Doi: 10.1145/3491102.3501831.

Chiang, C.-J., Jonson-Reid, M., & Drake, B. (2020). Caregiver physical health and child maltreatment reports and rereports. *Children and Youth Services Review, 108*, Article 104671. https://doi.org/10.1016/j.childyouth.2019.104671

Children's Bureau of the US Department of Health and Human Services. (2019). *Child Maltreatment 2017*. https://www.acf.hhs.gov/cb/resource/child-maltreatment-2017.

Chouldechova, A. (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data, 5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Coohey, C., Johnson, K., Renner, L. M., & Easton, S. D. (2013). Actuarial risk assessment in child protective services: Construction methodology and performance criteria. *Children and Youth Services Review, 35*(1), 151–161. https://doi.org/10.1016/j.childyouth.2012.09.020

D'Andrade, A., Austin, M. J., & Benton, A. (2008). Risk and safety assessment in child welfare: Instrument comparisons. *Journal of evidence-based social work, 5*(1–2), 31–56. https://doi.org/10.1300/J394v05n01_03

DePanfilis, D., & Zuravin, S. J. (1999). Epidemiology of Child Maltreatment Recurrences. *Social Service Review, 73*(2), 218–239. https://doi.org/10.1086/514415

Douglas, T., Pugh, J., Singh, I., Savulescu, J., & Fazel, S. (2017). Risk assessment tools in criminal justice and forensic psychiatry: The need for better data. *European Psychiatry, 42*, 134–137. https://doi.org/10.1016/j.eurpsy.2016.12.009

Department of Child Safety Youth Justice and Multicultural Affairs. (2021a, 5/1/2021). *Cultural differences in child-rearing practices*. Queensland Government. Retrieved 19/11/2021 from https://cspm.csyw.qld.gov.au/practice-kits/care-arrangements/working-with-aboriginal-and-torres-strait-islander/seeing-and-understanding/cultural-differences-in-childcd-rearing-practices.

Department of Child Safety Youth Justice and Multicultural Affairs. (2021b, 7/1/2021). *Over-representation*. Queensland Government. Retrieved 19/11/2021 from https://cspm.csyw.qld.gov.au/practice-kits/care-arrangements/working-with-aboriginal-and-torres-strait-islander/seeing-and-understanding/over-representation.

Dieterich, W., Mendoza, C., & Brennan, T. (2016). COMPAS Risk Scales: Accuracy Equity and Predictive Parity. *Northpointe*. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science advances, 4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Esposito, T., Chabot, M., Trocmé, N., Fluke, J. D., Delaye, A., Caldwell, J., & Mackrell, L. (2020). Recurrent involvement with the Quebec child protection system for reasons of neglect: A longitudinal clinical population study. *Child Abuse & Neglect, 104823–104823*. https://doi.org/10.1016/j.chiabu.2020.104823

Fallon, B., Lefebvre, R., Trocmé, N., Richard, K., Hélie, S., Montgomery, H. M., Bennett, M., Joh-Carnella, N., Saint-Girons, M., Filippelli, J., MacLaurin, B., Black, T., Esposito,T., King, B., Collin- Vézina, D., Dallaire, R., Gray, R., Levi, J., Orr, M., Petti, T., Thomas Prokop, S., & Soop, S. (2021). Denouncing the continued overrepresentation of First Nations children in Canadian child welfare: Findings from the First Nations/Canadian Incidence Study of Reported Child Abuse and Neglect-2019. Ontario: Assembly of First Nations. https://cwrp.ca/publications/denouncing-continued-overrepresentation-first-nations-children-canadian-child-welfare#:~:text=(PDF%2C%205635MB)-,The%20First%20Nations%2FCanadian%20Incidence%20Study%20of%20Reported%20Child%20Abuse,child%20welfare%20investigations%20across%20Canada.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fluke, J. D., Harlaar, N., Brown, B., Heisler, K., Merkel-Holguin, L., & Darnell, A. (2018). Differential Response and Children Re-Reported to Child Protective Services: County Data From the National Child Abuse and Neglect Data System (NCANDS). *Child Maltreatment, 107755951881638*. https://doi.org/10.1177/1077559518816381

Harnett, P. H., & Featherstone, G. (2020). The role of decision making in the over-representation of Aboriginal and Torres Strait Islander children in the Australian child protection system. *Children and Youth Services Review, 113*, Article 105019. https://doi.org/10.1016/j.childyouth.2020.105019

Holbrook, H. M., & Hudziak, J. J. (2020). Risk factors that predict longitudinal patterns of substantiated and unsubstantiated maltreatment reports. *Child Abuse & Neglect, 99*, 104279. https://doi.org/10.1016/j.chiabu.2019.104279

Human Rights and Equal Opportunity Commission. (1997). *Bringing them Home*. Human Rights and Equal Opportunity Commission. https://www.humanrights.gov.au/our-work/bringing-them-home-report-1997.

Hunter, S.-A., Burton, J., Blacklaws, G., Soltysik, A., Mastroianni, A., Young, J., . . . Shlonsky, A. (2020). *The Family Matters Report 2020: Measuring trends to turn the tide on the over-representation of Aboriginal and Torres Strait Islander children in out-of-home care in Australia*. https://www.familymatters.org.au/the-family-matters-report-2020/.

Hyslop, I. (2017). Child Protection in New Zealand: A History of the Future. *British Journal of Social Work, 47*(6), 1800–1817. https://doi.org/10.1093/bjsw/bcx088

Jenkins, B. Q. (2021). Measuring the equity of risk assessment instruments used in child protection. *Children and Youth Services Review, 131*, Article 106266. https://doi.org/10.1016/j.childyouth.2021.106266

Jenkins, B. Q., Tilbury, C., Hayes, H., & Mazerolle, P. (2018). Factors associated with child protection recurrence in Australia. *Child Abuse & Neglect, 81*, 181–191. https://doi.org/10.1016/j.chiabu.2018.05.002

Jenkins, B. Q., Tilbury, C., Hayes, H., & Mazerolle, P. (2019). Do measures of child protection recurrence obscure the differences between reporting and substantiation? *Children and Youth Services Review, 104*. https://doi.org/10.1016/j.childyouth.2019.104391

Johnson, K., & Bogie, A. (2009). Risk Assessment Validation: A Prospective Study. *Children's Research Center*. http://www.nccdglobal.org/sites/default/files/publication_pdf/nc_risk_asst_validation_final2009_2.pdf.

Johnson, W. (2005). Effects of Research-Based Risk Assessment on Racial/Ethnic Disproportionality in Service Provision Decisions. In D. Derezotes, M. Testa, & J. Poertner (Eds.), *Race matters in child welfare: the overrepresentation of African American children in the system* (pp. 147–161). CWLA Press.

Johnson, W. (2006). The risk assessment wars: A commentary response to "Evaluating the effectiveness of actuarial risk assessment models" by Donald Baumann, J. Randolph Law, Janess Sheets, Grant Reid, and J. Christopher Graham, Children and Youth Services Review, 27, pp. 465-490. *Children and Youth Services Review*, *28*(6), 704-714. Doi: 10.1016/j.childyouth.2005.07.007.

Johnson, W. (2011). The validity and utility of the California Family Risk Assessment under practice conditions in the field: A prospective study. *Child Abuse & Neglect, 35*(1), 18–28. https://doi.org/10.1016/j.chiabu.2010.08.002

Johnson, W., Clancy, T., & Bastian, P. (2015). Child abuse/neglect risk assessment under field practice conditions: Tests of external and temporal validity and comparison with heart disease prediction. *Children and Youth Services Review, 56*, 76–85. https://doi.org/10.1016/j.childyouth.2015.06.013

Keddell, E. (2019). Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice and the Implications for Practice. *Social sciences, 8*(10), 281. https://doi.org/10.3390/socsci8100281

Luken, A., Nair, R., & Fix, R. L. (2021). On Racial Disparities in Child Abuse Reports: Exploratory Mapping the 2018 NCANDS. *Child Maltreatment, 26*, 267–281. https://doi.org/10.1177/10775595211001926

Mickelson, N., LaLiberte, T., & Piescher, K. (2017). *Assessing Risk: A Comparison of Tools for Child Welfare Practice with Indigenous Families*. https://cascw.umn.edu/wp-content/uploads/2018/01/Risk-Assessment_FinalReport.pdf.

O'Donnell, M., Taplin, S., Marriott, R., Lima, F., & Stanley, F. J. (2019). Infant removals: The need to address the over-representation of Aboriginal infants and community concerns of another 'stolen generation'. *Child Abuse & Neglect, 90*, 88–98. https://doi.org/10.1016/j.chiabu.2019.01.017

Queensland Government. (*n.d.*). *Changing Tracks: An action plan for Aboriginal and Torres Strait Islander children and families 2017-2019*. Brisbane: Queensland Government. Retrieved from https://www.cyjma.qld.gov.au/resources/campaign/supporting-families/changing-tracks.pdf.

Russell, J. (2015). Predictive analytics and child protection: Constraints and opportunities. *Child Abuse & Neglect, 46*, 182–189. https://doi.org/10.1016/j.chiabu.2015.05.022

Schwalbe, C. S. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review, 30*(12), 1458–1464. https://doi.org/10.1016/j.childyouth.2007.11.021

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review, 27*(4), 409–427. https://doi.org/10.1016/j.childyouth.2004.11.007

Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? A systematic review and meta-analysis. *PLoS ONE, 8*(9), e72484–e. https://doi.org/10.1371/journal.pone.0072484

Slobogin, C. (2018). Principles of risk assessment for researchers and practitioners. *Behavioral Sciences & the Law, 36*(5), 507–516. https://doi.org/10.1002/bsl.2383

Steering Committee for the Review of Government Service Provision. (2021). *Report on Government Services 2021*. Productivity Commission. http://www.pc.gov.au/research/ongoing/report-on-government-services.

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics, 17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Tilbury, C. (2015). Aboriginal and Torres Strait Islander Families in Australia: Poverty and Child Welfare Involvement. In E. Fernandez, A. Zeira, T. Vecchiato, & C. Canali (Eds.), *Theoretical and Empirical Insights into Child and Family Poverty: Cross National Perspectives* (pp. 273–284). Springer International Publishing. https://doi.org/10.1007/978-3-319-17506-5_17.

van der Put, C. E., Assink, M., & Boekhout van Solinge, N. F. (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse & Neglect, 73*, 71–88. https://doi.org/10.1016/j.chiabu.2017.09.016

White, O. G., Hindley, N., & Jones, D. P. (2015). Risk factors for child maltreatment recurrence: An updated systematic review. *Medicine, Science and the Law, 55*(4), 259–277. https://doi.org/10.1177/0025802414543855

Wicke Dankert, E., & Bogie, A. (2015). *Preliminary Risk Assessment Fit Anlaysis of the SDM Family Risk Assessment*. Children's Research Center.