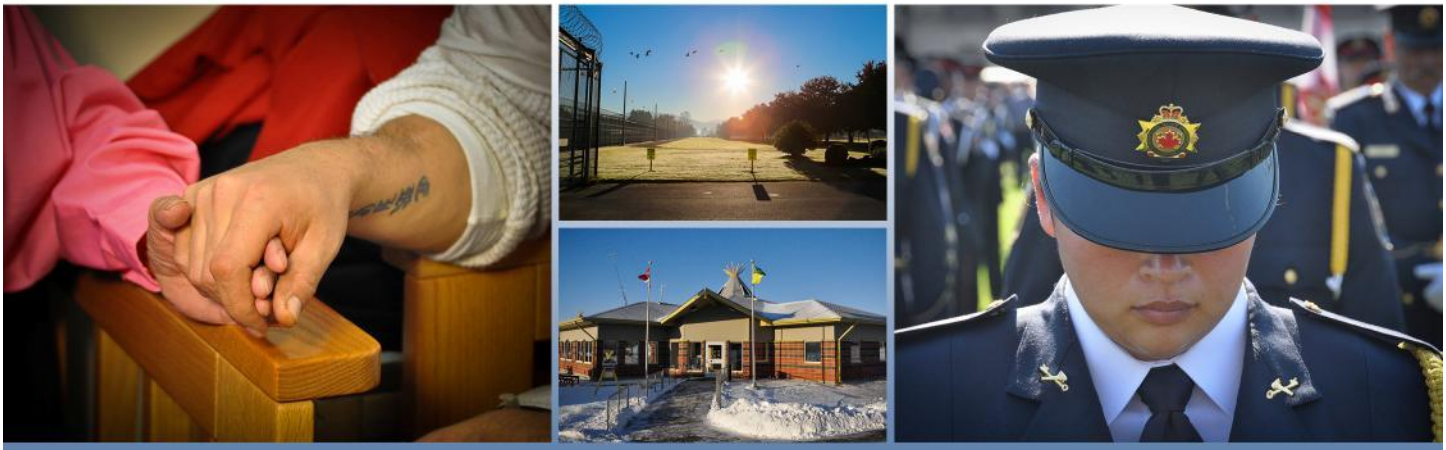
CORRECTIONAL SERVICE CANADA

CHANGING LIVES. PROTECTING CANADIANS.



RESEARCH REPORT

Optimization of Fetal Alcohol Spectrum Disorder Screening in Correctional Settings via Harmonization of Archival Datasets

2023 Nº R-447

Cat. No.: PS83-5/R447E-PDF

ISBN: 978-0-660-42690-7

Ce rapport est également disponible en français. Pour en obtenir un exemplaire, veuillez vous adresser à la Direction de la recherche, Service correctionnel du Canada, 340, avenue Laurier Ouest, Ottawa (Ontario) K1A 0P9.

This report is also available in French. Should additional copies be required, they can be obtained from the Research Branch, Correctional Service of Canada, 340 Laurier Ave. West, Ottawa, Ontario K1A 0P9.

Correctional Service Canada | Service correctionnel Canada

Canada

Optimization of Fetal Alcohol Spectrum Disorder Screening in Correctional Settings via Harmonization of Archival Datasets

Kaitlyn McLachlan, PhD [a,b] & James MacKillop, PhD [b]

[a] Department of Psychology
University of Guelph

[b] Peter Boris Centre for Addictions Research
McMaster University & St. Joseph's Healthcare Hamilton

Correctional Service of Canada

2023

Acknowledgements

This work was supported by a Memorandum of Agreement between the Peter Boris Centre for Addictions Research at McMaster University/St. Joseph's Healthcare Hamilton and Correctional Service Canada. The authors thank the staff of the Research Branch of Correctional Service of Canada for their instrumental support of the project, especially Laura Gamwell, Shanna Farrell-MacDonald, Geoff Wilton, and Dr. Lynn Stewart. The authors thank Ms. Jane De Jesus for administrative support in the Peter Boris Centre for Addictions Research.

Executive Summary

Key words: *offenders, fetal alcohol spectrum disorder, screening, assessment*

Fetal alcohol spectrum disorder (FASD) is a common neurodevelopmental disability caused by prenatal alcohol exposure (PAE) that impacts more than 1.4 million people in Canada. FASD is more prevalent in criminal justice contexts compared to the general population, with estimates for adults in correctional settings ranging from 10 to 23%. Despite overrepresentation in correctional settings, individuals with FASD frequently go unidentified, which may lead to increased risk of adverse outcomes for this population. Though several FASD screening tools have been developed to facilitate identification of individuals considered to be at risk in correctional contexts, the evidence base supporting their application in practice remains limited.

The Brief Screen Checklist (BSC) is a screening tool developed by the Correctional Service of Canada for the purposes of identifying adults at risk of having FASD in correctional settings. Early evidence points to the tool's promise, though further empirical evaluation is warranted to support evidence-based decisions made using the BSC. This report summarizes findings from a study that aimed to integrate and harmonize data from existing sponsored studies of FASD in the correctional context to develop optimized, data-driven algorithms for identifying individuals with FASD, with a focus on the BSC. We specifically sought to identify the frequency of concerns identified by individuals diagnosed with FASD in correctional contexts as assessed using the BSC, identify differences in profiles across BSC indicators between those with and without an FASD diagnosis, and assess the predictive validity of the BSC compared to "gold standard" diagnostic dispositions as a reference standard.

Three anonymized datasets from Canadian case ascertainment studies that aimed to identify rates of FASD in adult male and female offenders were integrated. Data elements for demographic characteristics (e.g., age, gender, marital status, ethnicity), diagnostic outcome (e.g., diagnosis, four-digit code), and BSC items were harmonized. Logistic regression was used to asses BSC Indicators and screening outcomes as predictors of diagnostic outcome. Sensitivity, specificity, positive predictive value, negative predictive value, and overall classification accuracy were also calculated.

The harmonized dataset comprised 194 individuals (18% female, 19-40 years), drawn from both institutional and community correctional settings. Cases were classified across four diagnostic outcomes: 14% (n = 27) were diagnosed with FASD or probable FASD; 21% (n = 40) received an 'Uncertain' classification owing to insufficient clinical information necessary for a reliable final diagnosis; 44% (n = 85) had significant CNS deficits thought to stem from causal factors other than PAE, and 21% (n = 39) had no CNS deficits identified, irrespective of PAE. On the BSC, 19% of individuals indicated a history of PAE, and 30% indicated frequent or heavy maternal alcohol use in childhood. Total scores for each of the BSC Behavioural, Historical, and Maternal Indicators significantly predicted FASD diagnosis. Final screening outcomes showed promising predictive accuracy, but suboptimal sensitivity in reference to the gold standard

clinical diagnosis.

The optimized dataset comprises the largest known sample of adults screened and assessed for FASD in a correctional context. Results suggest that conservatively estimated rates of FASD in this population are elevated relative to the general population, though further research is needed to generalize findings beyond these settings. Preliminary findings indicate that the BSC items and Indicators were present at elevated rates in the group ultimately diagnosed with FASD. A model comprising scores from three BSC Indicators resulted in promising overall screening accuracy, with high positive predictive value and specificity, though further refinements or analytic approaches may be required to improve lower than desirable sensitivity. While the current results suggest that those identified on the BSC using this approach are likely to meet criteria for FASD following a clinical diagnostic evaluation, a substantial number of cases diagnosed with FASD were not captured using the BSC. Applying advanced data-driven computational analytic approaches to further inform sensitive screening decisions may prove helpful, in addition to considering further harmonization of clinical indicators across the datasets.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

Fetal alcohol spectrum disorder (FASD) is a common neurodevelopmental disability cause by prenatal alcohol exposure (PAE) that impacts more than 1.4 million people in Canada (Lange et al., 2017; Popova et al., 2018). Though limited by a relatively small number of epidemiological studies, estimates suggest FASD prevalence ranges from 2-5% in North America (Popova et al., 2018; May et al., 2018). Individuals with FASD experience an array of difficulties related to neurocognitive deficits, problems regulating affect and behaviour, and functional deficits in their everyday living, problem-solving, and decision-making (Cook et al., 2016; Mattson et al., 2019). Rates of additional mental and physical health problems are also elevated in individuals with FASD relative to the general population, and they frequently experience a range of additional developmental insults during the pre- and postnatal periods that adversely impact outcomes (Pei et al., 2011; Popova et al., 2016). FASD is substantially more prevalent in vulnerable populations, including the criminal justice system, with contact rates derived from clinical samples estimated to range from 29-60% among individuals diagnosed with FASD (McLachlan et al., under review; Streissguth et al., 2004). Health economic estimates also indicate substantial costs linked with FASD in the Canadian context, ranging from $1.8 to 9.7 billion across sectors, with costs attributable to the criminal justice system forming one of the highest drivers (Popova et al., 2015a,b; Thanh & Jonsson, 2015).

Fetal Alcohol Spectrum Disorder in Correctional Contexts

A limited number of prevalence studies focused on ascertaining the number of youth and adults with FASD in forensic and correctional contexts have shown estimated rates ranging from 10% to 36% in Canada and Australia, including two reports detailing studies undertaken by the Correctional Service of Canada (Forrester et al., 2015; MacPherson, Chudley, & Grant, 2011). In another recently published study, the prevalence of FASD in a northern Canadian correctional population was ascertained using a parallel methodology, finding that 17.5% of the sample met diagnostic criteria for FASD, and that prevalence could have been as high as 31% with additional clinical data (McLachlan et al., 2019). Across these studies, most individuals diagnosed with FASD had not been previously identified. Though the evidence for understanding criminal justice system trajectories and risk pathways for individuals with FASD remains scarce, experts suggest that identifying their complex and specialized needs may improve their management and

treatment, and in turn, outcomes in both correctional and community contexts (Binnie, Trussler, & Jonsson, 2014).

Despite clinical concern and mounting evidence pointing to the overrepresentation of individuals with FASD in criminal justice contexts in Canada, they frequently go unidentified in justice settings more broadly (Bower et al., 2018; Burd et al., 2004, 2010; McLachlan et al., 2019). Several aspects of the disability can complicate identification, including a lack of overt physical characteristics in the majority of cases (e.g., 90%), variability in clinical presentation coupled with complex health and mental health comorbidities, relatively successful compensatory strategies that mask complex neurocognitive deficits, and a lack of knowledge and awareness in clinical and justice professionals working in legal contexts (Astley, 2010; Brown et al., 2016; Passmore et al., 2018; McLachlan et al., under review; Popova et al., 2016; Wedding et al., 2007). Recent guidance has focused on the importance of developing valid FASD screening tools appropriate for criminal justice settings, though the current evidence base evaluating the appropriateness of relevant tools remains limited (Steering Committee on FASD and Access to Justice, 2016, McLachlan et al., under review, 2019)

The Brief Screen Checklist (MacPherson et al., 2011)

Several screening instruments have been developed with the goal of identifying individuals with FASD in correctional contexts. The Brief Screen Checklist (BSC) (MacPherson et al., 2011) is a promising screening tool designed to identify adults who may be at risk of having FASD in federal correctional settings. Psychometric properties identified during the development and initial validation phase suggest that the instrument may offer a valid and reliable approach to FASD screening in this context. However, research has yet to evaluate the tool in a multisite sample of adults in different geographic regions, an important next step in confirming psychometric properties of any screening tool beyond the context in which the tool was initially developed. Identification, appropriate assessment and diagnosis, and effective provision of resources to meet needs form important protective factors for FASD (McLachlan et al., 2018; Pei et al., 2015; Rogers et al., 2013; Streissguth et al., 2004). In correctional contexts, evidence-based screening approaches may improve capacity for identifying complex needs in this population, better target limited assessment and diagnostic resources, and inform management, treatment, and planning for return to community (Brown et al., 2018; Burd et al., 2004; 2008).

Objectives of the Current Study

The current study aimed to integrate and harmonize data from existing studies of FASD in the correctional context to develop optimized, data-driven algorithms for identifying individuals with FASD. We specifically sought to identify the frequency of concerns identified by individuals diagnosed with FASD in correctional contexts as assessed using the BSC, identify differences in profiles across BSC indicators between those with and without an FASD diagnosis, and assess the validity of the BSC in the harmonized dataset.

Method

Participants

Participants in the final harmonized dataset across three sites included 193 adults (18% female) ages 19 to 40 years (Table 1). The original samples included 91 adults incarcerated in a federal correctional institution in the Prairie region, 23 adults incarcerated in a federal correctional institution in the Atlantic region, and 80 adults in a northern correctional jurisdiction comprising individuals drawn from both community and correctional settings. The three studies shared generally parallel methodologies and research goals. All participants were enrolled in prospective case ascertainment studies designed to estimate the prevalence of FASD in correctional contexts and used a multidisciplinary team in reaching diagnostic conclusions informed by the 2005 Canadian FASD Diagnostic Guidelines (Chudley et al., 2005).

Table 1

Participant characteristics

	Overall (*N* = 194)		**Site 1** (*n* = 91)		**Site 2** (*n* = 23)		**Site 3** (*n* = 80)		χ^2 /***F***	***p***	∅/ η_p^2
	n or M	% or SD	*n* or M	% or SD	*n* or M	% or SD	*n* or M	% or SD			
Age (M, SD) (19-40 years)	26.6	5.1	23.9	2.9	28.0	5.3	29.4	5.3	35.00	< .001	.27
Gender (% male)	159	82	0	0	23	100	12	15	125.02	<.001	.80
Marital Status (n, % single)	112	58	47	42	12	11	53	47	4.05	.132	.14
Diagnosis (n, %)											
FASD Diagnosis	27	14	9	10	4	17	14	18			
Uncertain	40	21	16	18	5	22	19	24	22.33	.001	.34
CNS Deficits	85	44	38	42	5	6	42	53			
No Deficits	42	22	28	31	9	39	5	6			

Note: M = Mean. SD = Standard Deviation. CNS = Central Nervous System; FASD = Fetal alcohol spectrum disorder.

Diagnostic classifications. Four levels of diagnostic classification were made using the same criteria across sites. The first classification, 'FASD Diagnosis', was made in 14% of cases (*n* = 27) across the harmonized sample. These cases received a diagnosis falling under the FASD umbrella following the 2005 Canadian FASD Diagnostic Guidelines (Chudley et al., 2005). Of these cases, the majority (*n* = 22, 81%) met criteria for Alcohol Related Neurodevelopmental Disorder (ARND) and a small number (*n* = 5, 19%) met criteria for partial Fetal Alcohol Syndrome (pFAS) (see Table 2 for diagnostic criteria). Several of these cases received a provisional diagnosis (*n* = 4, 15%) at a single site. The second classification, 'Unclear', was made for cases where there was insufficient information at time of clinical evaluation to confirm or rule out a diagnosis. A third classification, 'central nervous system (CNS) Deficits', was made in 44% of cases (*n* = 85), wherein significant CNS deficits were evident, but clinicians ascribed these difficulties to causes other than the impact of PAE. Last, a fourth classification, 'No Deficits,' was made for 21% of cases (*n* = 39), where no significant CNS deficits were detected, irrespective of PAE status.

Table 2

Diagnostic classifications

Classification	***n***	**%**	**Description**
FASD Diagnosis	27	14	Provisional or confirmed diagnosis made under the FASD umbrella following the 2005 Canadian FASD Diagnostic Guidelines (Chudley et al., 2005) for both: A) partial Fetal Alcohol Syndrome (pFAS): • Simultaneous presentation of two sentinel facial features (short palpebral fissure length; smooth or flattened philtrum; or thin upper lip) • Evidence of impairment in three CNS domains • Confirmed PAE B) Alcohol Related Neurodevelopmental Disorder (ARND) is diagnosed if: • Evidence of impairment in three CNS domains • Confirmed PAE
Uncertain	21	40	Insufficient information was available to either confirm or rule out a diagnosis for one of two reasons: a) CNS dysfunction in three or more brain domains was found, however PAE was unknown (clinicians were unable to confirm/rule out the possibility) b) PAE was confirmed, but there was not enough evidence from neuropsychological testing to meet criteria for an FASD diagnosis (i.e., CNS dysfunction was evident in only two domains)
CNS Deficits	85	44	Evidence of moderate to severe CNS dysfunction considered *not* related to PAE (as determined by clinicians)
No Deficits	39	21	No CNS deficits were identified, with or without PAE

Note. CNS = central nervous system deficits. PAE = prenatal alcohol exposure. Diagnostic criteria for FASD used across the three studies are outlined in Chudley et al. (2005).

Brief Screen Checklist (BSC, MacPherson et al., 2011)

The BSC is a screening tool developed by the Correctional Service of Canada to aid in identifying adult offenders in correctional settings who may have FASD. The BSC is a self-report screening tool that comprises 46-items[1] across three domains, including Behavioural, Historical, and Maternal Indicators (see Tables 3, 5-7). Three versions have been developed, including one that can be completed by the individual undergoing screening (BSC-Participant Version), a second for birth mothers (BSC-Maternal Version), and a third for collateral informants (BSC-Collateral). The current study focuses on the participant/self-report version only. In each of the three samples participants completed the BSC prior to undergoing diagnostic evaluation for FASD. Slight variations in BSC versions administered across the three samples were addressed based on version iteration during harmonization for the current study.

BSC Behavioural Indicator. For the current study, we achieved harmonization for 25 BSC Behavioural Indicator items. Variation in wording between versions were addressed through harmonization, and items not administered across all three samples were excluded for the current study (Table 3). Behavioral Indicator items were originally rated by participants on a 6-point Likert scale at Site 1 and Site 3, (0 = do not know; 1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree), and on a 5-point Likert scale at Site 2, (1 = disagree a lot; 2 = disagree; 3 = agree; 4 = agree a lot; 5 = do not know). For the current study, we harmonized and dichotomized the ratings (e.g., absent/do not know = do not know; strongly disagree; disagree; neither agree nor disagree; and present = agree, strongly agree). We also calculated a final cumulative total score across the 25 Behavioural Indicator items by summing individual items.

[1] The final number of items administered ranged by site with slight variations in versions.

Table 3

Item level Behavioural Indicator harmonization for discrepant wording across versions and sites

Site 1	Site 2	Site 3	Decision	Final Item Label
Has a problem with spelling?	Has trouble with spelling?	Has a problem with spelling?	Retain: Has a problem with spelling?	BEH 4
Shows poor judgment?	Makes bad choices a lot of the time	Shows poor judgment?	Retain: Shows poor judgment?	BEH 5
Has a problem budgeting or handling money?	Has a problem managing money?	Has a problem budgeting or handling money?	Retain: Has a problem budgeting or handling money?	BEH 10
Seems unaware of the consequences of his actions?	Seems unaware of the consequences of actions	Seems unaware of the consequences of actions?	Retain: Seems unaware of the consequences of actions?	BEH 11
Has a problem with arithmetic?	Has trouble with math?	Has a problem with arithmetic?	Retain: Has a problem with arithmetic?	BEH 12
Is very forgetful of everyday things?	Is always forgetting things?	Is very forgetful of everyday things?	Retain: Is very forgetful of everyday things?	BEH 15
Talks a lot but says little?	Talks a lot but has a hard time getting point across?	Talks a lot but says little?	Retain: Talks a lot but says little?	BEH 16

Table 3 (continued)

Site 1	Site 2	Site 3	Decision	Final Item Label
Has a problem with reading?	Has trouble with reading?	Has a problem with reading?	Retain: Has a problem with reading?	BEH 18
Has a poor attention span?	Has a hard time paying attention?	Has a poor attention span	Retain: Has a poor attention span	BEH 21
Is strongly opinionated?	Not administered	Is strongly opinionated?	Dropped	N/A
Has few friends?	Not administered	Has few friends?	Dropped	N/A
Is stubborn?	Not administered	Is Stubborn	Dropped	N/A
Easily gets stressed out or anxious?	Not administered	Not administered	Dropped	N/A
Does not like change?	Not administered	Not administered	Dropped	N/A
Likes to be with people?	Not administered	Not administered	Dropped	N/A
Has trouble making decisions?	Not administered	Not administered	Dropped	N/A
Has trouble staying interested in things?	Not administered	Not administered	Dropped	N/A

Historical Indicator. Nine BSC Historical Indicator items were originally administered to participants in all three samples, and six were included in the current analyses, consistent with the approach undertaken by MacPherson et al. (2011, see Table 6). Four Historical Indicator items were rated dichotomously, including: history of adoption; foster care; problems with school from an early age; and treatment for mental health problems (e.g., no/don't know vs. yes). Frequency ratings were used for two Historical Indicator items, including number of times in foster care, and number of times mental health treatment was received (e.g., 0, 1-2, or 3+ times). For the latter items, we calculated a dichotomous score reflecting number of experiences (e.g., < 3 times or ≥ 3 times). A cumulative total score across the six Historical Indicator items was then calculated by summing individual items.

Maternal Indicator. For the current study, four Maternal Indicator items were considered for analysis following harmonization owing to high rates of missing data across samples. Items considered for the current study included: history of maternal drinking in pregnancy (coded as no/unknown vs. yes); history of maternal alcohol use in childhood (coded as no/unknown vs. yes); and two indicators that were combined to yield a dichotomous score (coded as yes vs. no/unknown) following scoring procedures outlined in MacPherson et al. (2011), including indication of either frequent maternal alcohol use in childhood (2 or more times in a week), or heavy maternal alcohol use in childhood (e.g., more than 4 drinks on a single drinking occasion). Ultimately, given the high level of missing data (i.e., participants had difficulty providing details about maternal alcohol use), and potential confounding between maternal alcohol use in pregnancy and diagnostic outcome, we applied the approach adopted by MacPherson et al. (2011) and used a single dichotomous Maternal Indicator (frequent and/or heavy alcohol use in childhood) as a proxy risk variable signalling increased risk of PAE.

BSC screening decision rule. In their initial BSC validation study MacPherson and colleagues (2011) defined the criteria for a positive screen using a data-driven approach that required a 'hit' on each of the three BSC Indicators based on their final 46-item BSC and dichotomized response formats. Their recommended screening decision rule required a score of > 10 on the Behavioural Indicator, combined with both a score > 2 on the Historical Indicator, and a positive single Maternal Indicator.

Data Harmonization

Three anonymized datasets from prospective case ascertainment studies that aimed to identify rates of FASD in adult male and female offenders were harmonized (see Table 1). In each study, adults were voluntarily recruited by research teams to undergo the FASD assessment and diagnostic process, conducted by an interdisciplinary team that adhered to the 2005 Canadian FASD Diagnostic Guidelines (Chudley et al., 2005). Teams evaluated four diagnostic features, including growth restriction, sentinel facial features, CNS deficits, and PAE. Members of the various clinical interdisciplinary teams included a physician, psychologist, and research/clinical coordinator. Diagnostic decisions were made following interdisciplinary case conferences and based on all available data. For the current study, fields for demographic characteristics (e.g., age, gender, marital status, ethnicity), diagnostic outcome (e.g., diagnosis, four-digit code), and BSC items, were harmonized. The study received full approval from the Hamilton Integrated Research Ethics Board (HiREB Project # 7749-C).

Analytic Approach

Descriptive characteristics of the harmonized sample across sites, and both BSC items and Indicators, were characterized using frequency counts and percentages for categorical data and means and standard deviations for continuous data. Comparisons between sites and diagnostic outcomes were made using Analysis of Variance (ANOVA) and Pearson chi-square analyses. Pearson chi-square analyses also were conducted to compare dichotomous BSC Indicator outcomes and items across the four diagnostic groups. Pearson and point-Biserial correlations were calculated to evaluate associations between demographic characteristics, diagnostic outcomes, and BSC Indicators to assess the need to control for potential covariates.

When evaluating BSC classification accuracy we opted to exclude the ‘Uncertain’ cases owing to a lack of diagnostic clarity (e.g., FASD was neither ruled in or out for these cases, making it difficult to evaluate BSC screening accuracy). We also combined the two groups where FASD was ruled out (e.g., CNS deficits, and No Deficits), in order to derive a final dichotomous approach to diagnostic classification.

We conducted logistic regressions to assess the three BSC Indicators (e.g., Behavioural, Historical, and Maternal) using several approaches. We first independently evaluated the cumulative Behavioural and Historical Indicator total scores, and the dichotomous screening decisions for each of the three Indicators as proposed by MacPherson et al., (2011). We next

evaluated the cumulative Behavioural and Historical Indicator total scores, and the dichotomous Maternal Indicator in a single overall model using stepwise logistic regression. We also evaluated the final BSC screening decision rule as proposed by MacPherson et al. (2011). Subsequently, we evaluated the BSC Behavioural and Historical Indicators in the current, optimized sample, to identify potential refinements to the BSC decision rules. We then followed a parallel approach and evaluated these new, sample optimized BSC Indicator cut-off scores individually, in stepwise combined model with the original Maternal Indicator, and finally, together as a sample optimized final screening decision rule. Both age and gender were included in all regression models as covariates.

We also calculated indicators of classification accuracy for the individual dichotomous BSC Indicators and final screening decision rules (using the approach outlined by both MacPherson et al., 2011, and sample-optimized cut-off scores). These included: sensitivity (probability that a test result will be positive for cases receiving an FASD diagnosis); specificity (probability that a test result will be negative for cases not diagnosed with FASD, or classified as CNS Deficits/No Deficits); positive predictive value (PPV, probability that cases with a positive screening outcome truly have FASD) and negative predictive value (NPV, probability that cases with a negative screening outcome truly do not have FASD) (see Table 4).

For dichotomous prediction models (e.g., based on a single BSC Indicator cut-off score) classification accuracy parameters were computed in two ways. First, we based calculations on raw, observed frequency counts drawn from contingency tables. This method was thought to mirror real-world clinical decision-making using raw items and cut-off scores that can be manually calculated. Second, we calculated predicted classifications using logistic regression (accounting for covariates in each model) and used these to derive classification accuracy parameters (these results are reported in Appendix 1). This method paralleled the approach we used to calculate classification accuracy parameters for models that included both continuous individual predictors (e.g., a single cumulative BSC Indicator total score), or combined models using multiple predictors.

Table 4

Diagnostic classifications

		Gold Standard FASD Diagnosis		
		FASD Diagnosis (#)	Not FASD[a] (#)	TOTAL
BSC Screening Outcome	Positive Screen (#)	A (True Positive)	B (False Positive)	$T_{Positive\ Screen}$
	Negative Screen (#)	C (False Negative)	D (True Positive)	$T_{Negative\ Screen}$
		$T_{FASD\ Diagnosis}$	$T_{Not\ FASD}$	TOTAL

Note. [a]'Not FASD' combines the CNS Deficits + No Deficits groups for the current study. Sensitivity is calculated as A/(A+C) x 100. Specificity is calculated as D/(D+B) x 100. Positive Predictive Value is calculated as A/(A+B) x 100. Negative Predictive Value is calculated as D/(D+C) x 100.

Results

Preliminary Analyses

There were no significant differences in diagnostic outcome by age or gender. However, there were significant positive bivariate associations between age and both the BSC Historical Indicator total score (r = .23, p = .001), and the dichotomous Maternal Indicator (r = .14, p = .046). Gender was also significantly correlated with both the cumulative total scores for the BSC Behavioural (r = .31, p < .001) and Historical Indicators (r = .21, p = .003), as well as the dichotomous Maternal Indicator (r = .21, p = .003), with higher scores and greater endorsement in women compared to men. There were also significant correlations between site and both the cumulative Historical Indicator total score (r = .14, p = .047) and the dichotomous Maternal Indicator (r = .15, .036), though this may be partially confounded with gender. As such, we controlled for age and gender in subsequent analyses.

Behavioural Indicator

There were high rates of behavioural concerns identified across the harmonized sample, as indexed by item endorsement rates that ranged from 18% to 60% (Table 5). Several individual Behavioural Indicator items also significantly differentiated the groups, including trouble following directions, problems with spelling, problems budgeting or handling money, having trouble with math, interrupts conversations a lot, problems with reading, easily victimized, trouble completing tasks, trouble staying on topic, and poor social skills. For these items, rates were similar between the FASD and Uncertain groups, and generally elevated in these groups compared to the CNS Deficits and No Deficits groups.

Mean scores for the cumulative BSC Behavioural Indicator total score ranged from 0 to 24 (M = 8.10, SD = 6.21) with significant differences between the four diagnostic groups, F(3, 188) = 10.52, p < .001, η_p^2 = .14. Post-hoc pairwise comparisons indicated that scores were highest for the FASD group (M = 12.44, SD = 6.87), not statistically different from the Uncertain group (M = 10.15, SD = 6.09, p = .35), and significantly higher than both the CNS Deficits (M = 6.88, SD = 5.13, p < .001) and No Deficits groups (M = 5.81, SD = 6.13 p < .001). Scores in the Uncertain group were comparable to the CNS deficits group (p = .09), but significantly greater than the No Deficits group (p = .003). Scores for the CNS Deficits and No Deficits groups did not differ (p = .66). Using the originally proposed cut-off score of > 10 on the cumulative

Behavioural Indicator total score, 32% (n = 63) of the overall sample had a positive screen.

Table 5

Brief Screen Checklist Behavioural Indicators

	Total (*N* = 194)		**FASD (*n* = 27)**		**Unclear (*n* = 40)**		**CNS Deficits (*n* = 85)**		**No Deficits (*n* = 41)**		χ^2 /F	*p*	∅/ η_p^2
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
Acts impulsively	93	48	16	59	21	53	40	47	16	38	3.38	.34	.13
Has trouble following direction	58	30	16	59[a]	17	43[a]	19	23[b]	6	14[b]	21.08	<.001	.33
Is restless	90	46	17	63	16	40	39	46	18	43	3.86	.28	.14
Has a problem with spelling	60	31	18	67[a]	18	45[a]	19	22[b]	5	12[b]	29.89	<.001	.39
Shows poor judgment	67	35	14	52	13	33	26	31	14	33	4.27	.23	.15
Is easily distracted	115	60	20	74	27	68	46	54	22	52	5.33	.15	.17
Has temper tantrums	54	28	10	37	16	40	19	22	9	22	6.04	.11	.18
Has strong mood swings	70	36	13	48	17	43	30	35	10	24	5.19	.16	.16
Is hyperactive	60	31	11	42[a,b]	19	48[b]	21	25[a]	9	22[a]	9.74	.02	.23
Problem budgeting, money	90	47	20	77[a]	18	45[b]	38	45[b]	14	33[b]	12.74	.005	.26
Unaware consequences of actions	47	24	9	35	14	35	15	18	9	21	6.22	.10	.18
Has trouble with math	69	36	16	59[a]	19	48[a]	23	27[b]	11	26[b]	13.40	.004	.26
Interrupts conversation a lot	54	28	10	39[a,b]	18	45[b]	19	22[a,c]	7	17[c]	11.17	.01	.24
Is agitated	85	44	16	59	21	53	33	39	15	36	5.82	.12	.17
Is very forgetful of everyday things	69	36	11	41	16	40	30	35	12	29	1.56	.67	.09
Talks a lot but says little	47	24	11	41[a]	17	43[a]	13	16[b]	6	14[b]	16.74	.001	.30
Has a poor memory	59	30	13	48	12	31	21	25	13	31	5.31	.15	.17

Table 5 (continued)

	Total (*N* = 194)		**FASD (*n* = 27)**		**Unclear (*n* = 40)**		**CNS Deficits (*n* = 85)**		**No Deficits (*n* = 41)**		χ^2 /F	*p*	∅/ η_p^2
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
Has a problem with reading	35	18	11	41[a]	12	30[a]	12	14[b]	0	0[c]	23.41	<.001	.35
Is easily victimized	39	20	13	50[a]	11	28[a]	11	13[b]	4	10[b]	21.39	<.001	.33
Has trouble completing tasks	54	28	14	54[a]	17	43[a]	16	19[b]	7	17[b]	18.44	<.001	.31
Has a poor attention span	74	38	15	56	17	43	30	35	12	29	5.72	.13	.17
Is easily manipulated	47	24	11	41	1	28	18	21	7	17	5.98	.11	.18
Is disorganized	38	20	5	19	9	23	16	19	8	19	.27	.97	.04
Has trouble staying on topic	59	30	15	56[a]	19	48[a]	17	20[b]	8	19[b]	20.50	<.001	.33
Has poor social skills	38	20	11	41[a]	11	28[a,b]	14	17[b,c]	2	5[c]	15.65	<.001	.28
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**			
Total Score (M, SD) (range, 0 - 24)	8.10	6.21	12.44	6.87	10.15	6.09	6.88	5.13	5.81	6.13	10.09	<.001	.14
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
Positive Screens > 10	63	33	15	56[a]	19	48[a]	20	24[b]	9	21[b]	16.12	<.001	.29

Note. Each subscript letter denotes groups who did not differ significantly from each other at the $p < .05$ level. M = Mean, SD = Standard Deviation. CNS = Central Nervous System. FASD = fetal alcohol spectrum disorder.

Historical Indicator

Across the entire sample, rates of positive endorsement of individual Historical Indicator items ranged from 7% (history of adoption) to 60% (problems with school from an early age) (Table 6). Two Historical Indicator items significantly differentiated the groups, including history of foster care, and problems with school from an early age. For these items, rates were similar between the FASD and Uncertain groups, and elevated in these groups compared to the CNS Deficits and No Deficits groups.

Mean scores on the cumulative Historical Indicator total score ranged from 0 to 5 (M = 1.56, SD = 1.47), with significant differences between the four diagnostic groups, $F(3, 188)$ = 13.23, $p < .001$, $\eta_p^2 = .17$. Post-hoc pairwise comparisons indicated that scores were highest for the FASD group (M = 2.67, SD = 1.47), and significantly greater than both the CNS Deficits (M = 1.13, SD = 1.10, $p < .001$), and the No Deficits groups (M = 1.07, SD = 1.37, $p < .001$), but not significantly different from the Uncertain group (M = 2.25, SD = 1.66, $p = .99$). Scores in the Uncertain group were also significantly higher than both the CNS Deficits ($p < .001$) and No Deficits groups ($p = .001$), while the CNS Deficits and No Deficits groups did not differ (p = .99). Across the entire sample, using the originally proposed cut-off score of > 2 for the cumulative Historical Indicator, 25% (n = 49 of 194) of cases had a positive screen.

Table 6

Brief Screen Checklist Historical Indicators

	Total (*N* = 194)		**FASD Diagnosis (*n* = 27)**		**Uncertain (*n* = 40)**		**CNS Deficits (*n* = 85)**		**No Deficits (*n* = 41)**		χ^2 /F	*p*	Ø/ η_p^2
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
Ever adopted	14	7	4	15	5	13	5	6	0	0	7.55	.06	.20
Ever in foster care	77	40	18	67[a]	25	63[a]	22	27[b]	12	29[b]	25.00	<.001	.36
≥ 3 times	37	51	11	69	14	61	8	36	4	36	5.74	.13	.28
Problems with school early	113	59	25	93[a]	32	80[a]	44	52[b]	12	29[c]	36.35	<.001	.44
History mental health treatment	42	22	10	37	10	25	12	14	10	24	6.99	.07	.19
≥3 times	20	54	4	44	4	44	5	50	7	78	2.78	.43	.27
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**			
Total Score (M, SD)	1.56	1.47	2.67	1.47	2.25	1.66	1.13	1.10	1.07	1.37	14.23	<.001	.17
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
Positive Screens >2	49	25	15	56	17	43	11	13	6	14	28.94	<.001	.39

Note. Each subscript letter denotes groups who did not differ significantly from each other at the $p = .05$ level. M = Mean, SD = Standard Deviation. CNS = Central Nervous System. FASD = fetal alcohol spectrum disorder.

Maternal Indicator

Across the entire sample, individual Maternal Indicator items were variably endorsed (Table 7). In total, 68% of respondents (*n* = 121) endorsed a history of maternal alcohol use during childhood, and 30% (*n* = 59) reported frequent or heavy use patterns. Another 19% of respondents (*n* = 29) endorsed a history of maternal alcohol use in pregnancy, though they generally provided few details about use patterns, with high rates of unknown responses and/or missing data across these items. Each of the individual Maternal Indicator items significantly differentiated the groups, with similarly elevated rates between the FASD and Uncertain groups, falling above both the CNS Deficits and No Deficits groups. The only exception was the indicator for maternal alcohol use in pregnancy, which was endorsed by significantly more respondents in the FASD group compared to all other groups (Table 7).

Given the potential confounding between confirmation of alcohol use in pregnancy and diagnostic outcome (relying on confirmation of PAE for diagnosis), all subsequent analyses focused on a single dichotomous Maternal Indicator, namely, 'frequent or heavy alcohol use during childhood', consistent with the approach taken by MacPherson et al. (2011). Across the entire sample, a positive endorsement of the single Maternal Indicator resulted in 30% of cases (*n* = 59 of 194) having a positive screen.

Table 7

Brief Screen Checklist Maternal Indicators

	Total (*N* = 194)		**FASD Diagnosis (*n* = 27)**		**Uncertain (*n* = 40)**		**CNS Deficits (*n* = 85)**		**No Deficits (*n* = 41)**		χ^2 /F	*p*	∅/ η_p^2
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**			
MAU while young	121	68	21	84[a]	30	79[a]	46	59[b]	24	63[a,b]	8.30	.04	.22
*Frequent or heavy MAU while young	59	30	15	56[a]	16	40[a,b]	19	22[c]	9	21[b,c]	14.01	.003	.27
MAU pregnancy	29	19	13	54[a]	8	26[b]	5	7[c]	3	11[b,c]	27.23	<.001	.43

Note. Each subscript letter denotes groups who did not differ significantly from each other at the .05 level. MAU = Maternal alcohol use. CNS = Central Nervous System. FASD = fetal alcohol spectrum disorder. *Indicates final dichotomous Maternal Indicator used in subsequent BSC analyses.

Regression Models for Individual BSC Indicators

We first conducted a series of logistic regressions to evaluate whether individual BSC Indicators (using cumulative Behavioural and Historical Indicator total scores, and the dichotomous Maternal Indicator) predicted diagnostic outcome. Analyses excluded Unclear cases and combined the CNS Deficits and No Deficits groups to create a dichotomous final diagnostic outcome against which to reference the BSC Indicators. Age and gender were accounted for in each model as covariates.

Behavioural Indicator. The cumulative Behavioural Indicator total score significantly predicted diagnosis, b = .17 (SE = .04), p <.001, with each one-unit increase accounting for 1.2x increased risk for being diagnosed with FASD (Table 8). Using the dichotomous cut-off score (> 10), 29% of the sample (n = 44 of 154) had a positive screen, also yielding a significant prediction of diagnosis, b = -1.52 (SE = .46), p = .001. Based on the dichotomous BSC Behavioural Indicator (>10 cut-off) alone, 74% of cases were accurately classified, with 15 of the 27 FASD cases being correctly identified (sensitivity = 56%), and 29 of 127 non-FASD cases incorrectly identified (specificity = 77%).

Historical Indicator. The cumulative Historical Indicator total score significantly predicted diagnosis, b = .83 (SE = .18), with each one-unit increase accounting for 2.3x increased risk for being diagnosed with FASD (Table 8). Using the cut-off dichotomous score (> 2), 21% of cases had a positive screen (n = 32 of 154), also yielding a significant prediction of FASD diagnosis, b = -2.09 (SE = .49), p < .001. Using only the dichotomous BSC Historical Indicator (> 2) resulted 81% of cases being accurately classified. In total, 15 of the 27 FASD cases were correctly identified (sensitivity = 56%), and 17 of 127 non-FASD cases were incorrectly identified (specificity = 87%).

Maternal Indicator. Positive endorsement of the single dichotomous Maternal Indicator significantly predicted FASD diagnosis, b = -1.46 (SE = .45), p = .001, yielding a small increase in odds (.23) of being diagnosed with FASD (Table 8). In terms of classification accuracy, using only the Maternal Indicator (a positive score) resulted in 73% classification accuracy, with 15 of the 27 FASD cases being correctly identified (sensitivity = 56%), and 28 of 127 non-FASD cases being incorrectly identified (specificity = 77%).

Stepwise Logistic Regression Model for Individual BSC Indicators

We next conducted a pair of stepwise logistic regressions to evaluate the predictive accuracy of each BSC Indicator individually, and together in a combined model. Uncertain cases were again excluded, and the CNS Deficits and No Deficits groups were combined to yield a final dichotomous diagnostic outcome.

In the first model we focused on the cumulative Behavioural and Historical Indicator total scores, and the dichotomous Maternal Indicator. In the first block of the model we included age and gender as covariates. In the second and third blocks we included each of the cumulative Behavioural and Historical total scores, and in the fourth and final block we included the dichotomous Maternal Indicator. Model results indicated that neither age (p = .152) nor gender (p = .996) were significant predictors in the first block. In the second block, the cumulative Behavioural Indicator total score was a significant predictor of diagnostic outcome (p < .001). In the third block, the cumulative Behavioural Indicator remained significant (p = .003), and the cumulative Historical Indicator total score was also a significant predictor of diagnostic outcome (p < .001). In the last block, both the cumulative Behavioural (p = .005) and Historical Indicator total scores (p = .001) remained significant, as did the dichotomous Maternal Indicator (p = .013). The final model accounted for 25% of diagnostic outcome (Cox and Snell pseudo R^2 statistic). Including all three predictors in the final model resulted in the best overall classification accuracy (88%). In total, 14 of the 27 FASD cases were correctly identified (sensitivity = 52%), and only 5 of 127 non-FASD cases were incorrectly identified (specificity = 96%).

In the second model we focused on the dichotomous cut-off scores for all three BSC Indicators (Behavioural, >10; Historical >2; positive Maternal) and followed the same series of steps and blocks as in the first model. Model results indicated that neither age (p = .41) nor gender (p = .48) were significant predictors in the first block. In the second block, the dichotomous Behavioural Indicator (>10) was a significant predictor of diagnostic outcome (p = .001). In the third block, the dichotomous Behavioural Indicator remained significant (p < .02), and the dichotomous Historical Indicator was also a significant predictor of diagnostic outcome (p < .001). In the last block, both the dichotomous Behavioural (p = .04) and Historical (p = .002) Indicators remained significant, and the single dichotomous Material Indicator also predicted diagnostic outcome (p = .019). The final model accounted for 32% of diagnostic

outcome (Cox and Snell pseudo R^2 statistic). Including all three predictors in the final model resulted in 87% classification accuracy, with 12 of the 27 FASD cases being correctly identified (sensitivity = 44%), and 5 of 127 non-FASD cases were incorrectly identified (specificity = 96%).

Regression Model for Final BSC Screening Decision Rule

We next completed a logistic regression to evaluate the predictive accuracy of the final BSC screening decision rule proposed by MacPherson et al. (2011), specifically, including those who scored > 10 on the Behavioural Indicator, and > 2 on the Historical Indicator, while also having a positive dichotomous Maternal Indicator. Here, Uncertain cases were again excluded, and the CNS Deficits and No Deficits groups were combined to yield a final dichotomous diagnosis variable.

The final BSC screening decision rule was a significant predictor of diagnostic outcome, b = -4.03, (SE = .88), $p < .001$ (Table 8). This model accounted for 20% of diagnostic outcome (Cox and Snell pseudo R^2 statistic) and an overall accuracy rate of 88%. In total, 11 of 16 FASD cases were correctly identified (sensitivity = 41%), and 3 of 127 non-FASD cases were incorrectly identified (specificity = 98%).

Table 8

Predictive accuracy of Brief Screen Checklist indicators for diagnostic outcome

	B	SE	Wald	*p*-value	Odds Ratio	95% CI
Individual BSC Indicators						
Behavioural Indicator Total Score	.18	.04	17.74	<.001	1.19	1.20, 1.30
Behavioural Indicator > 10	-1.52	.46	10.87	.001	.22	.09, .54
Historical Indicator Total Score	.83	.18	21.12	<.001	2.29	1.61, 3.25
Historical Indicator > 2	-2.09	.49	18.17	<.001	.124	.05, .32
Maternal Indicator	-1.46	.45	10.70	.001	.23	.10, .56
Stepwise Model 1[a]						
Behavioural Indicator Total Score	.13	.05	7.77	.005	1.14	1.04, 1.25
Historical Indicator Total Score	.64	.20	10.56	.001	1.89	1.29, 2.78
Maternal Indicator	-1.30	.52	6.23	.013	.27	.10, .76
Stepwise Model 2[a]						
Step 2. Behavioural Indicator > 10	-1.10	.54	4.23	.04	.33	.12, .95
Step 3. Historical Indicator > 2	-1.63	.52	9.86	.002	.20	.07, .54
Step 4. Maternal Indicator	-1.17	.50	5.52	.02	.31	.12, .82
Final Screening Decision Rule[b]	-4.03	.88	20.81	<.001	.018	.003, .10

Predictive Accuracy	**% Accuracy**	**Sensitivity**	**Specificity**	**NPV**	**PPV**
Individual BSC Indicators					
Behavioural Indicator >10[i]	73.4%	55.6%	77.2%	89.1%	34.0%
Historical Indicator > 2[i]	81.2%	55.6%	86.6%	90.2%	46.8%

Table 8 (continued)

Predictive Accuracy	% Accuracy	Sensitivity	Specificity	NPV	PPV
Maternal Indicator[i]	73.0%	55.6%	77.0%	89.1%	34.0%
Stepwise Model 1[ii]	88.3%	51.9%	96.1%	90.4%	73.6%
Stepwise Model 2[ii]	87.0%	44.4%	96.1%	89.1%	70.5%
Final Screening Decision Rule[a,i]	87.7%	40.7%	97.6%	88.6%	78.5%

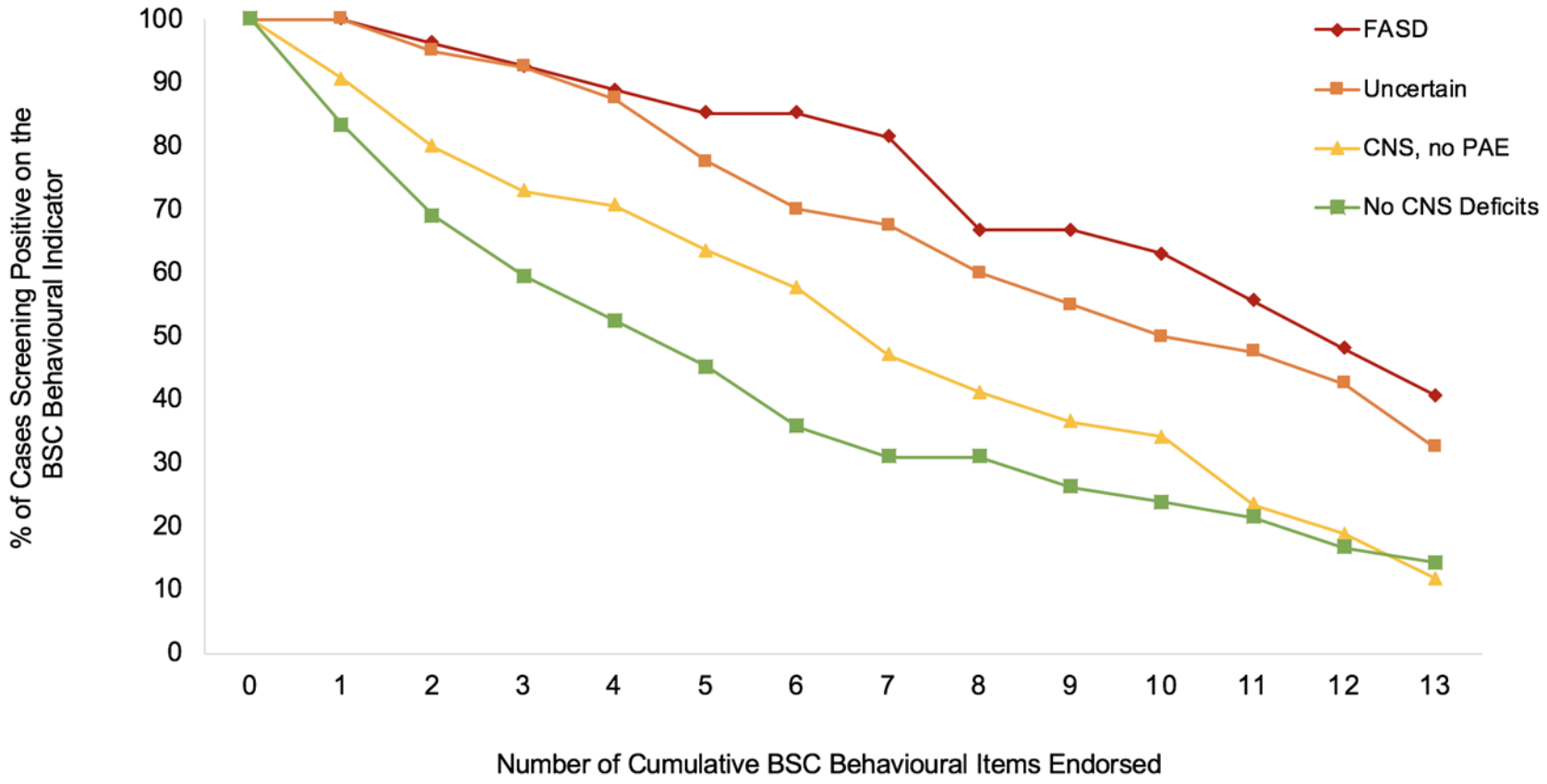
Note. *N* = 154. [a]Model indicators are derived for the final model based on all covariates and predictors. [b]Positive screen defined as having a Behavioural Indicator score > 10, + a Historical Indicator score >2, + positive Maternal Indicator. [i] Sensitivity, specificity, NPV, and PPV, are calculated from raw, observed values. These parameters were also calculated from logistic regression models that accounted for covariates, based on predicted values, and are shown in Appendix 1. [ii] Sensitivity, specificity, NPV, and PPV are calculated using predicted values derived from logistic regression models, accounting for covariates. SE = Standard Error. CI = 95% Confidence Interval. NPV = Negative Predictive Value. PPV = Positive Predictive Value.

Sample Optimized BSC Indicators

Given lower than optimal sensitivity across original models and scoring criteria, we also sought to assess alternative approaches to scoring the BSC Behavioural and Historical Indicators. We evaluated potentially improved or sample-optimized screening cut-off scores by assessing each level of cumulative item endorsement for all cases in the harmonized sample (N = 194) and then weighing the best balance between sensitivity (e.g., identifying a sufficient number of true positive cases) and specificity (not overidentifying too many true negative cases), while excluding Unclear cases (for consistency with earlier analyses). We subsequently evaluated predictive accuracy for the optimized cut-off scores following the same approaches already used for original cut-off scores. In assessing predictive accuracy, we again excluded Unclear cases, and combined the CNS Deficits and No Deficits groups to attain a dichotomous diagnostic outcome.

Behavioural Indicator. We first evaluated optimal screening cut-points at each level of item endorsement for the cumulative BSC Behavioural Indicator total score. The best balance of sensitivity and specificity appeared be > 6 items (7 or more) (Figure 1). This cut-point was a significant predictor of diagnostic outcome, -1.86 (SE = .53, p = .001, 11% Cox and Snell pseudo R^2), with overall accuracy rate of 62%. In total, 22 of 27 FASD cases were correctly identified (82% sensitivity) and 53 of 127 non-FASD cases were incorrectly identified (specificity = 58%).

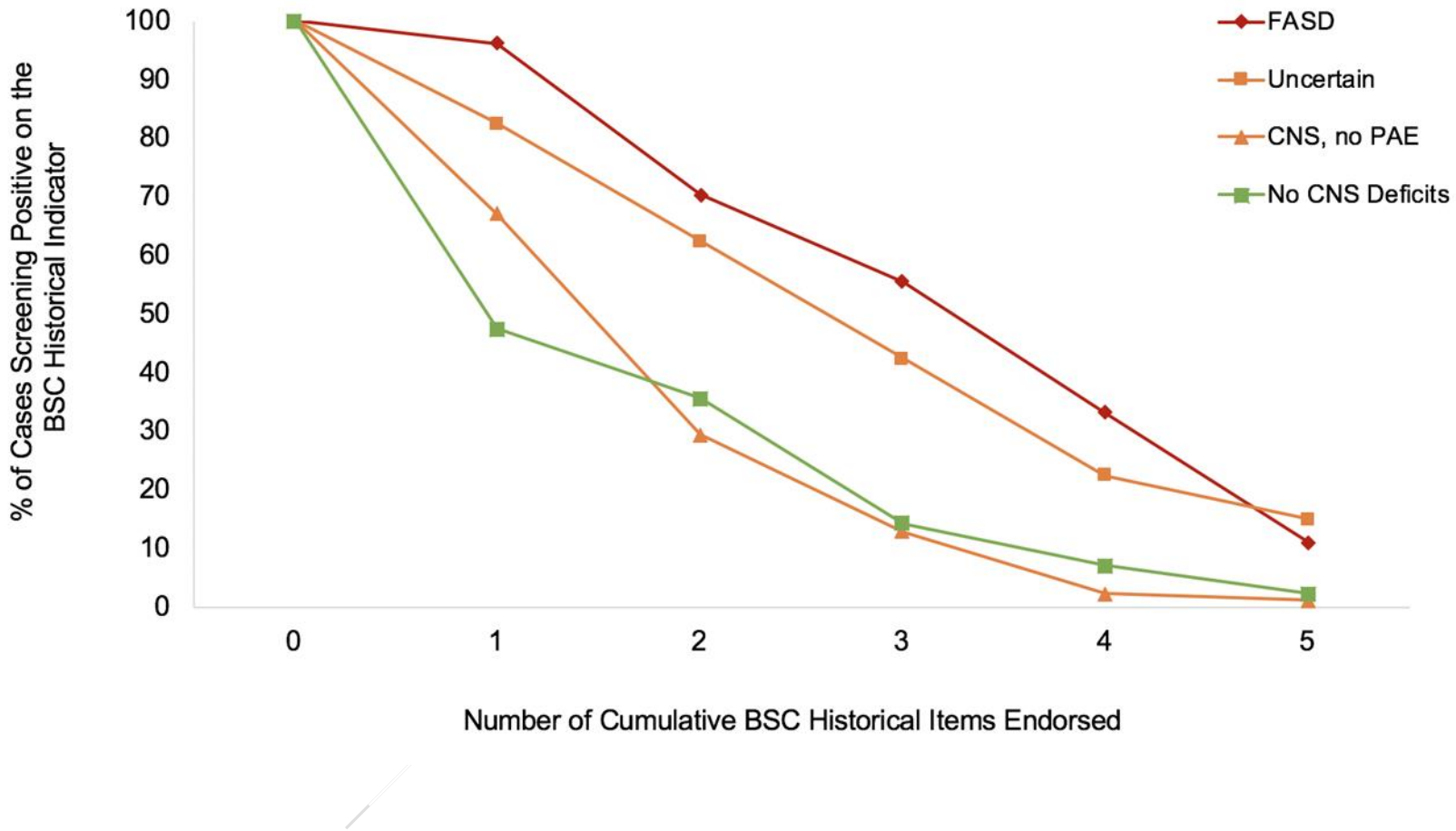
Figure 1. Percentage of cases screening positive on the cumulative BSC Behavioural Indicator for each level of cumulative items endorsed. A score of zero reflects classification of all individuals being classified positive.



Historical Indicator. We next evaluated optimal screening cut-points at each level of cumulative item endorsement for the BSC Historical Indicator total score (Figure 2). Weighing sensitivity and specificity, a clear 'optimal' screening threshold was not readily apparent. For a cut-off score of > 0 (e.g., any Historical Indicator was item endorsed) sensitivity was 96% and specificity was 39%. For a cut-off score of > 1 sensitivity was 70% and specificity was 69%. We have previously reported values for > 2 (original criteria proposed by MacPherson et al., 2011). For a cut-off score of > 3 sensitivity was to 33% and specificity was 96%.

Given the intended function of the BSC as a screening tool (vs. a diagnostic process), and the need to balance sensitivity and specificity, we opted to use the > 1 cut-off point. This cut-point significantly predicted of diagnostic outcome, b = 1.64 (SE = .48), p = .001 (15% Cox and Snell pseudo R^2 statistic) with an overall accuracy rate of 69% (Table 9). In total, 19 of 27 FASD cases were correctly identified (70% sensitivity) and 40 of 127 non-FASD cases were incorrectly identified (specificity = 69%).

Figure 2. Percentage of cases screening positive on the BSC Historical Indicator based on number of cumulative items endorsed. A score of zero reflects classification of all individuals being classified positive.



Stepwise Logistic Regression Model for Optimized BSC Indicators

We next conducted a stepwise logistic regression to evaluate predictive accuracy of each optimized BSC Indicator in a combined model. Here, Uncertain cases were again excluded, and the CNS Deficits and No Deficits groups were combined to yield a final dichotomous diagnosis variable. In the first block of the model we included age and gender as covariates. In the second and third blocks we included each of the sample-optimized Behavioural (> 6) and Historical (> 1) Indicator cut-off scores, and in the fourth and final block we included the original dichotomous Maternal Indicator.

Model results indicated that neither age (p = .15) nor gender (p = .996) were significant predictors in the first block. In the second block, the dichotomous Behavioural Indicator (> 6) was a significant predictor of diagnostic outcome (p = .001). In the third block, both the dichotomous Behavioural (p = .002), and Historical Indicators were significant predictors of diagnostic outcome (p = .002). In the last block, both the dichotomous Behavioural (p = .004) and Historical (p = .004) Indicators remained significant, and the single dichotomous Maternal Indicator also predicted diagnostic outcome (p = .007). The final model accounted for 20% of diagnostic outcome (Cox and Snell pseudo R^2 statistic). Including all three predictors in the final model resulted in overall good classification accuracy (87%), with 11 of 27 FASD cases correctly identified (41% sensitivity), and 4 of 127 non-FASD cases incorrectly identified (97% specificity).

Regression Model for Final Optimized BSC Screening Decision Rule

Last, we ran a logistic regression to evaluate the predictive accuracy of the sample optimized final BSC screening decision rule, specifically, those who scored > 6 on the Behavioural Indicator, and > 1 on the Historical Indicator, and also had a positive dichotomous Maternal Indicator. Uncertain cases were again excluded, and the CNS Deficits and No Deficits groups were combined to yield a final dichotomous diagnostic outcome.

The final sample optimized BSC screening decision rule was a significant predictor of diagnostic outcome, b = -3.28, (SE = .68), p < .001 (Table 9). This model accounted for 18% of diagnostic outcome (Cox and Snell pseudo R^2 statistic) and an overall accuracy rate of 87%. In total, 11 of 16 FASD cases were correctly identified (sensitivity = 41%), and 4 of 127 non-FASD

cases were incorrectly identified (specificity = 97%).

Results Summary

In sum, evaluating all models, the combined model relying on the cumulative Behavioural and Historical Indicator total scores, along with the Maternal Indicator, yielded the best classification accuracy (88%) with high specificity (96%) and importantly, 74% PPV, suggesting that those who were identified as being at risk for FASD via a positive screening outcome had a reasonably high probability of actually receiving an FASD diagnosis following the subsequent gold standard evaluation. Lower than optimal sensitivity was evident (52%), and clear scoring criteria cannot be readily extrapolated for everyday application using this model. However, findings suggest promise with regards to the potential for BSC items and Indicators providing valid predictions of diagnosed cases.

Table 9

Predictive accuracy of sample optimized Brief Screen Checklist indicators for diagnostic outcome

	B	SE	Wald	*p*-value	Odds Ratio	95% CI
Optimized Individual BSC Indicators						
Behavioural Indicator > 6	-1.86	.53	12.10	.001	.16	.06, .44
Historical Indicator > 1	1.64	.48	11.79	.001	5.17	2.02, 13.21
Stepwise Model[a]						
Behavioural Indicator > 6	-1.66	.57	8.44	.004	.19	.06, .58
Historical Indicator > 1	-1.47	.52	8.11	.004	.23	.08, .63
Maternal Indicator	-1.33	.49	7.31	.007	.26	.10, .69
Final Optimized Screening Decision Rule[b]	-3.28	.68	23.14	<.001	.04	.01, .14
Predictive Accuracy		**% Accuracy**	**Sensitivity**	**Specificity**	**NPV**	**PPV**
Individual BSC Indicators						
Behavioural Indicator > 6[i]		62.3%	81.5%	58.3%	93.7%	29.3%
Historical Indicator > 1[i]		68.6%	70.4%	68.5%	91.6%	32.2%
Stepwise Model[a, ii]		87.0%	40.7%	96.9%	88.5%	73.3%
Final Optimized Screening Decision Rule[b,i]		87.0%	40.7%	96.9%	88.5%	73.3%

Note. (*N* = 154) [a]Model indicators are for the final model with all covariates and predictors. [b]Positive screen defined as having a Behavioural Indicator score > 6, and a Historical Indicator score >1, and a history of substantial material alcohol use in childhood. [i] Sensitivity, specificity, NPV, and PPV, are calculated from raw, observed values. These parameters were also calculated from logistic regression models that accounted for covariates, based on predicted values, and are shown in Appendix 1. [ii] Sensitivity, specificity, NPV, and PPV are calculated using predicted values derived from logistic regression models, accounting for covariates. SE = Standard Error. CI = 95% Confidence Interval. NPV = Negative Predictive Value. PPV = Positive Predictive Value.

Discussion

The overall objective of this study was to integrate and harmonize data from existing studies of FASD in the correctional context to develop optimized, data-driven algorithms for identifying individuals with FASD. This report summarizes findings focusing on harmonized data from the BSC, a promising FASD screening tool developed for adults in the federal correctional context to aid in identifying those at risk of having the disability. In the current study, we assessed the frequency of concerns identified by individuals via the BSC who underwent diagnostic evaluation for FASD across multiple settings and regions of Canada. We also evaluated differences in profiles across BSC Indicators between different diagnostic outcomes, and assessed the predictive validity of the BSC in the harmonized dataset. Findings indicate good success harmonizing the datasets and suggest promise with respect to the predictive validity of BSC as a potentially helpful FASD screening tool for application with adults in correctional settings.

The harmonized dataset characterized in the current study comprises the largest known sample of adults screened and assessed for FASD in a correctional context. The results suggest that conservatively estimated rates of FASD drawn from three samples are elevated relative to the general population (e.g., 14% vs. 2-5%, May et al., 2018; Popova et al., 2018, 2019). A further 21% of cases were classified with diagnostic uncertainty, indicating that this rate could have been much higher in the context of clear clinical information (particularly confirmation of PAE), underscoring the challenges inherent in accurately identifying adults with FASD, particularly in justice contexts. In addition, another 44% of individuals presented with significant CNS deficits, highlighting the elevated rates of cognitive difficulties frequently observed in correctional populations (e.g., LaDuke et al., 2017; Stewart et al., 2016).

In regard to behavioural functioning, individuals diagnosed with FASD presented with generally high rates of difficulty with respect to following directions and completing tasks, deficits in their academic skills (e.g., spelling, math, reading) and adaptive functioning (e.g., budgeting money, communication, social skills), along with elevated experiences of victimization. Though these rates were generally comparable relative to those who received an ‘Unclear’ clinical classification, they were much higher than those with both CNS deficits not

related to PAE, and those without CNS deficits, suggesting important potential clinical signals for this population. Several Historical Indicators also distinguished the groups following a similar pattern, including a history of foster care, and problems with school from an early age. These findings are consistent with previous reports in the individual samples comprising the harmonized dataset, suggesting consistency across Canadian settings, and also parallel findings from other FASD studies in youth justice settings and among clinical samples (e.g., Bower et al., 2018; Lebel et al., 2019; McLachlan et al., under review). Further, screening tools that yield helpful 'here and now' clinical information that can inform practice while awaiting potentially limited future assessment and diagnostic resources also improve the everyday utility of these tools. Findings from the current study suggest areas of deficit or vulnerability that may be relevant to informing responsive management and treatment plans for individuals with FASD in community and correctional settings include academic skill needs, problems related to inattention and task completion, skills of everyday living (adaptive functioning), and possibly increased risk of victimization.

While Maternal Indicators form important clinical markers in the context of flagging cases at elevated risk for FASD, the current results suggest that individuals may have difficulty providing detailed information about maternal health behaviours, and that additional behavioural and/or historical information may be necessary to inform identification decisions made in the absence of additional confirmatory information about PAE risk. This finding also aligns with the challenges inherent in evaluating FASD in older populations where determining clear information about PAE may prove more challenging as compared to pediatric populations (Brown et al., 2018; Chudley et al., 2007).

Our findings suggest that several approaches to scoring the BSC yielded reasonably good predictive accuracy, with the strongest model combining all data derived from the Behavioural, Historical, and Maternal Indicators. This model resulted in relatively strong overall predictive accuracy, though additional refinement is required to derive appropriate clinical guidance for making ultimate selection decisions (e.g., cut-off scores) by maximizing all available Indicators and items. It is important to also consider the current findings in relation to the BSC's intended operationalization as a *screening* tool, rather than a tool used for diagnostic decision-making. Screening tests are typically administered with the aim of identifying individuals who may meet criteria for clinical diagnosis and are often applied in populations identified as being at increased

risk for having a disability or disease. Thus, screening tools such as the BSC are not used with the intention of establishing a diagnosis, but rather, to identify individuals who would benefit from additional clinical evaluation (Fletcher, 2019; Trevethan, 2017).

Ideally, a useful and effective screening tool will be able to detect most of the cases at elevated risk for having the target outcome, in this case FASD (e.g., high sensitivity) while simultaneously minimizing the number of false positives (e.g., high specificity) (Fletcher, 2019; Trevethan, 2017). The current findings suggest promise in respect to not over-identifying a high number of cases not considered to be at risk for ultimately having FASD, which proves to be an important consideration given the high costs and limited availability of comprehensive multidisciplinary gold standard FASD evaluations. Further, relatively good positive predictive value for the combined predictive BSC model suggests that a relatively high proportion of those who screened positive ultimately went on to receive an FASD diagnosis, again, an important consideration for resource-limited contexts, and where adverse impacts may be associated with stigmatization for misidentified individuals and families (Brintnell et al., 2010; Clarren & Lutke, 2008; Corrigan et al., 2018, 2019). That said, lower than optimal sensitivity for all models, including the final combined model, suggests that the BSC consistently underidentified a subset of individuals who ultimately went on to receive an FASD diagnosis, suggesting that cases may be missed relying only on this approach to screening for FASD without further adjustment and/or refinement to the instrument. This may be problematic insofar as failing to identify individuals with increased vulnerability and clinical needs associated with FASD may result in poor outcomes, particularly given the challenges already present in detecting those with the disability in correctional settings. That said, BSC item- and scale-level distinctions between clinical groups suggest that further work refining decision-making criteria using the tool may prove helpful in better identifying these cases.

Limitations

Use of archival data sources presents challenges and limitations to be kept in consideration when interpreting the current findings. While the three datasets yield a considerable overall sample size relative to previous FASD-justice studies, the number of individuals ultimately diagnosed with FASD remained a small minority. In addition, there were many unclear diagnostic outcomes likely owing to the challenges associated with obtaining clear information about PAE required to reach a firm diagnosis for many justice-involved adults who

often present with additional factors contributing to their current functioning. Excluding these cases may have resulted in somewhat inflated estimates of predictive accuracy in the current study. In addition, all participants recruited for the initial research were voluntarily enrolled in the study, thus selection bias cannot be ruled out.

Conclusions and Future Directions

Increasing evidence suggests that individuals with FASD may be overrepresented in a variety of correctional settings, though evidence to date has been limited to a small number of studies in single institutions and/or jurisdiction, with relatively conservative sample sizes. The sample harmonized for the current study provides an important picture framing the cognitive and health challenges presented by many individuals with FASD in correctional contexts, including factors and needs that may be of relevance to their effective care, management, and reintegration into community. Identifying individuals with FASD in correctional settings is important for informing whether modifications or accommodations to traditional programming and approaches may be required to effectively support their needs. Our findings point to the early promise of the BSC's potential utility as a tool for both identifying individuals with FASD, and informing current management and support approaches (e.g., understanding current behavioural needs, historical experiences). However, additional adjustments may be necessary to ensure that fewer individuals who may meet diagnostic criteria for FASD are missed, along with a need to evaluate the tool in the real-world context of 'unclear' cases. Applying more sophisticated data-driven analytic approaches to developing an optimized scoring approach for the BSC may prove helpful (e.g., latent class analysis, machine learning). As well, further harmonization of indicators beyond the BSC from the samples used in the current study may also be indicated.

References

Astley, S. J. (2010). Profile of the first 1400 patients receiving diagnostic evaluations for fetal alcohol spectrum disorder at the Washington State Fetal Alcohol Syndrome Diagnostic & Prevention Network. *Canadian Journal of Clinical Pharmacology*, *17*(1), e132-164.

Bower, C., Watkins, R. E., Mutch, R. C., Marriott, R., Freeman, J., Kippin, N. R., Safe, B., Pestell, C., Cheung, C. S. C., Shield, H., Tarratt, L., Springall, A., Taylor, J., Walker, N., Argiro, E., Leitão, S., Hamilton, S., Condon, C., Passmore, H. M., & Giglia, R. (2018). Fetal alcohol spectrum disorder and youth justice: a prevalence study among young people sentenced to detention in Western Australia. *BMJ Open*, *8*. https://doi.org/10.1136/bmjopen-2017-019605

Brintnell, E. S., Bailey, P. G., Sawhnney, A., & Kreftin, L. (2011). Understanding FASD: Disability and social supports for adult offenders. *Fetal alcohol spectrum disorders: Management and policy perspectives of FASD.* Toronto, ON: Wiley.

Brown, J., Helmstetter, D., Harr, D., & Singh, J. (2016). Perceptions of FASD by United States District Attorneys. *Journal of Intellectual Disabilities and Offending Behaviour*, *7*(4), 195–201. https://doi.org/10.1108/JIDOB-04-2016-0005

Brown, J., Freeman, N., Pickett, H., Watts, E., & Trnka, A. J. (2018). FASD in adult populations: Clinical and forensic considerations. In E. Jonsson, S. Clarren, & I. Binnie (Eds.), *Ethical and legal perspectives in fetal alcohol spectrum disorders (FASD)* (pp. 163-183). Springer, Cham. https://doi.org/10.1007/978-3-319-71755-5_10

Burd, L., Fast, D. K., Conry, J., & Williams, A. (2010). Fetal Alcohol Spectrum Disorder as a Marker for Increased Risk of Involvement with Correction Systems. *The Journal of Psychiatry and Law*, *38*(4), 559–583. https://doi.org/10.1177/009318531003800408

Burd, L., Martsolf, J. T., & Juelson, T. (2004). Fetal alcohol spectrum disorder in the corrections system: Potential screening strategies. *Journal of FAS International*, *2*(February), e1.

Clarren, S. K., & Lutke, J. (2008). Building clinical capacity for fetal alcohol spectrum disorder diagnoses in western and northern Canada. *Journal of Population Therapeutics and Clinical Pharmacology*, *15*(2).

Chudley, A. E., Conry, J., Cook, J. L., Loock, C., Rosales, T., & LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *172*(5 Suppl), S1–S21. https://doi.org/10.1503/cmaj.1040302

Chudley, A. E., Kilgour, A. R., Cranston, M., & Edwards, M. (2007). Challenges of diagnosis in fetal alcohol syndrome and fetal alcohol spectrum disorder in the adult. *American Journal*

of Medical Genetics Part C: Seminars in Medical Genetics, *145*(3), 261–272. https://doi.org/10.1002/ajmg.c.30140

Cook, J. L., Green, C. R., Lilley, C. M., Anderson, S. M., Baldwin, M. E., Chudley, A. E., Conry, J. L., LeBlanc, N., Loock, C. A., Lutke, J., Mallon, B. F., McFarlane, A. A., Temple, V. K., & Rosales, T. (2016). Fetal alcohol spectrum disorder: A guideline for diagnosis across the lifespan. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*. https://doi.org/10.1503/cmaj.141593

Corrigan, P. W., Shah, B. B., Lara, J. L., Mitchell, K. T., Simmes, D., & Jones, K. L. (2018). Addressing the public health concerns of Fetal Alcohol Spectrum Disorder: Impact of stigma and health literacy. *Drug and alcohol dependence, 185*, 266-270.

Corrigan, P. W., Shah, B. B., Lara, J. L., Mitchell, K. T., Simmes, D., & Jones, K. L. (2019). Stakeholder perspectives on the stigma of fetal alcohol spectrum disorder. *Addictive Research & Theory*, *27*(2), 170-177.

Fletcher, G.S. (2019). *Clinical epidemiology: the essentials.* Lippincott Williams & Wilkins.

Forrester, P., Davis, C. G., Moser, A., MacPherson, P., Gobeil, R., & Chudley, A. E. (2015). *Assessing Fetal Alcohol Spectrum Disorder in Women Offenders*.

LaDuke, C., DeMatteo, D., Heilbrun, K., Gallo, J., & Swirsky-Sacchetti, T. (2017). The neuropsychological assessment of justice-involved men: Descriptive analysis, preliminary data, and a case for group-specific norms. Archives of Clinical Neuropsychology, 32(8), 929-942.

Lange, S., Probst, C., Gmel, G., Rehm, J., Burd, L., & Popova, S. (2017). Global Prevalence of Fetal Alcohol Spectrum Disorder Among Children and Youth. *JAMA Pediatrics*, *171*(10), 948. https://doi.org/10.1001/jamapediatrics.2017.1919

Lebel, C. A., McMorris, C. A., Kar, P., Ritter, C., Andre, Q., Tortorelli, C., & Gibbard, W. B. (2019). Characterizing adverse prenatal and postnatal experiences in children. *Birth Defects Research*, *111*(12), 848-858.

MacPherson, P. H., Chudley, A. E., & Grant, B. A. (2011). *Fetal Alcohol Spectrum Disorder (FASD) in a correctional population: Prevalence, screening and diagnosis*. Correctional Service of Canada.

Mattson, S. N., Bernes, G. A., & Doyle, L. R. (2019). Fetal Alcohol Spectrum Disorders: A Review of the Neurobehavioral Deficits Associated With Prenatal Alcohol Exposure. *Alcoholism: Clinical and Experimental Research*, *43*(6). https://doi.org/10.1111/acer.14040

May, P. A., Chambers, C. D., Kalberg, W. O., Zellner, J., Feldman, H., Buckley, D., Kopald, D., Hasken, J. M., Xu, R., Honerkamp-Smith, G., Taras, H., Manning, M. A., Robinson, L. K., Adam, M. P., Abdul-Rahman, O., Vaux, K., Jewett, T., Elliott, A. J., Kable, J. A., …

Hoyme, H. E. (2018). Prevalence of Fetal Alcohol Spectrum Disorders in 4 US Communities. *JAMA*, *319*(5), 474. https://doi.org/10.1001/jama.2017.21896

McLachlan, K., Amlung, M., & MacKillop, J. (2019). Evidence-Based Screening, Assessment, and Intervention-Approaches for Offenders with Fetal Alcohol Spectrum Disorders and other Neurodevelopmental Disabilities: A Systematic Review. Report prepared for Correctional Service of Canada.

McLachlan, K., Gray, A. L., Roesch, R., Douglas, K. S., & Viljoen, J. L. (2018). An evaluation of the predictive validity of the SAVRY and YLS/CMI in justice-involved youth with fetal alcohol spectrum disorder. *Psychological Assessment*, *30*(12), 1640–1651. https://doi.org/10.1037/pas0000612

McLachlan, K., Mullally, K., Ritter, C., Mela, M., & Pei, J. (under review). Fetal Alcohol Spectrum Disorder and Neurodevelopmental Disabilities: An International Practice Survey of Forensic Mental Health Clinicians, submitted to the *International Journal of Forensic Mental Health Services.*

McLachlan, K., Flannigan, K., Temple, V., Unsworth, K., & Cook, J. (under review). Difficulties in daily living experienced by adolescents, transition-aged youth, and adults with Fetal Alcohol Spectrum Disorder, submitted to *Alcoholism: Clinical and Experimental Research.*

McLachlan, K, McNeil, A., Pei, J., Brain, U., Andrew, G., & Oberlander, T. (2019). Prevalence and characteristics of adults with fetal alcohol spectrum disorder in corrections: A Canadian case ascertainment study. *BMC Public Health*.

Passmore, H. M., Mutch, R. C., Burns, S., Watkins, R., Carapetis, J., Hall, G., & Bower, C. (2018). Fetal Alcohol Spectrum Disorder (FASD): Knowledge, attitudes, experiences and practices of the Western Australian youth custodial workforce. *International Journal of Law and Psychiatry*, *59*(PG-44-52), 44–52. https://doi.org/10.1016/j.ijlp.2018.05.008

Pei, J., Denys, K., Hughes, J., & Rasmussen, C. (2011). Mental health issues in fetal alcohol spectrum disorder. *Journal of Mental Health*, *20*(5), 473–483. https://doi.org/doi: 10.3109/09638237.2011.577113

Pei, J., Leung, W. S. W., Jampolsky, F., & Alsbury, B. (2015). Experiences in the Canadian Criminal Justice System for Individuals with Fetal Alcohol Spectrum Disorders: Double Jeopardy? *Canadian Journal of Criminology and Criminal Justice*, *58*(1), 56–86. https://doi.org/10.3138/cjccj.2014.e25

Popova, S., Lange, S., Burd, L., & Rehm, J. (2015a). Cost attributable to fetal alcohol spectrum disorder in the Canadian correctional system. *International Journal of Law and Psychiatry*, *41*, 76–81. https://doi.org/10.1016/j.ijlp.2015.03.010

Popova, S., Lange, S., Burd, L., & Rehm, J. (2015b). The economic burden of fetal alcohol spectrum disorder in Canada in 2013. *Alcohol and Alcoholism*, *51*(3), 367-375.

https://doi.org/10.1093/alcalc/agv117

Popova, S, Lange, S., Chudley, A. E., Reynolds, J. N., & Rehm, J. (2018). *World Health Organization International Study on the Prevalence of Fetal Alcohol Spectrum Disorder (FASD) Canadian Component*. http://store.camh.ca

Popova, S, Lange, S., Shield, K., Mihic, A., Chudley, A., Mukherjee, R. A. S., Bekmuradov, D., & Rehm, J. (2016). Comorbidity of fetal alcohol spectrum disorder: a systematic review and meta-analysis. *The Lancet*. https://doi.org/10.1016/S0140-6736(15)01345-8

Popova, Svetlana, Lange, S., Shield, K., Burd, L., & Rehm, J. (2019). Prevalence of fetal alcohol spectrum disorder among special subpopulations: a systematic review and meta-analysis. *Addiction*, *114*(7), add.14598. https://doi.org/10.1111/add.14598

Rogers, B. J., McLachlan, K., & Roesch, R. (2013). Resilience and enculturation: Strengths among young offenders with Fetal Alcohol Spectrum Disorder. *First Peoples Child & Family Review*, *8*(1), 63–81.

Steering Committee on FASD and Access to Justice. (2016). *Final report to Federal/Provincial/Territorial Ministers responsible for justice and public safety*. Retrieved from: https://scics.ca/en/product-produit/final-report-to-federalprovincialterritorial-ministers-responsible-for-justice-and-public-safety/

Stewart, L. A., Wilton, G., & Sapers, J. (2016). Offenders with cognitive deficits in a Canadian prison population: Prevalence, profile, and outcomes. *International journal of law and psychiatry*, *44*, 7-14.

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Sampson, P. D., O'Malley, K., & Young, J. K. (2004). Risk factors for adverse life outcomes in Fetal Alcohol Sydnrome and Fetal Alcohol Effects. *Journal of Developmental and Behavioral Pediatrics*, *25*(4), 228–238.

Thanh, N. X., & Jonsson, E. (2015). Costs of fetal alcohol spectrum disorder in the Canadian criminal justice system. *Journal of Population Therapeutics and Clinical Pharmacology*, *22*(1), e125-131.

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in public health*, *5*, 307.

Wedding, D., Kohout, J., Mengel, M. B., Ohlemiller, M., Ulione, M., Cook, K., Rudeen, K., & Braddock, S. (2007). Psychologists' knowledge and attitudes about fetal alcohol syndrome, fetal alcohol spectrum disorders, and alcohol use during pregnancy. *Professional Psychology: Research and Practice*, *38*(2), 208.

Appendix A

Table A1

Predictive accuracy of Brief Screen Checklist indicators for diagnostic outcome based on logistic regression with covariates

Predictive Accuracy	**% Accuracy**	**Sensitivity**	**Specificity**	**NPV**	**PPV**
Individual BSC Indicators					
Behavioural Indicator > 10	82.5%	3.7%	99.2%	82.9%	49.9%
Historical Indicator > 2	80.6%	7.4%	96.1%	83.0%	28.5%
Maternal Indicator	82.5%	0.0%	100.0%	82.5%	-
Final Screening Decision Rule[a]	87.7%	40.7%	97.6%	88.6%	78.5%
Sample Optimized Indicators					
Individual BSC Indicators					
Behavioural Indicator > 6	83.2%	7.4%	99.2%	83.5%	66.6%
Historical Indicator > 1	87.7%	40.7%	97.6%	88.6%	78.5%
Final Optimized Screening Decision Rule[b]	87.0%	40.1%	96.9%	88.5%	73.3%

Note. (N = 154). [a]Positive screen defined as having a Behavioural Indicator score > 10, a Historical Indicator score > 2, and a history of substantial material alcohol use in childhood. [b]Positive screen defined as having a Behavioural Indicator score > 6, and a Historical Indicator score > 1, and a positive Maternal Indicator. All predictive accuracy indicators are calculated from predicted values derived from logistic regression models, accounting for covariates. SE = Standard Error. CI = 95% Confidence Interval. NPV = Negative Predictive Value. PPV = Positive Predictive Value.