The Role of Early Environmental Factors in Predicting Adolescent Psychopathology: A Secondary Data Analysis of NICHD SECCYD Data using Machine Learning

Yasmine Faiad

DClinPsy Thesis (Volume 1), 2020

University College London

UCL Doctorate in Clinical Psychology

Thesis Declaration Form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: [redacted]

Name: Yasmine Issam Faiad

Date: 07/09/20

Overview

This thesis consists of three parts and generally aimed at examining the relationship between early antecedents, particularly early experiences with caregivers, in relation to important developmental outcomes.

Part one is a meta-analysis which examined the association between early maternal sensitivity and later cognitive development as well as the potential moderating effects of study- and sample-level moderators.

Part two is an empirical research paper that consisted of using Machine Learning for performing a secondary analysis on the data from the National Institute for Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD). This is a population-based, multi-phase longitudinal study which followed children and their caregivers from the age of 1 month to 15 years and primarily aimed to examine the relation of different aspects of childcare and developmental outcomes. The secondary analysis ultimately aimed to explore the use of Machine Learning for the prediction of later behaviour problems and compare it to more traditional regression techniques.

Part three consists of a critical appraisal in which I reflected on the process of working on my empirical research project and the insights I have gained as a result of this. It describes how this piece of work has influenced my views and interests as a psychologist and a researcher. Most importantly, it examines the implementation of Machine Learning in the field of mental health more critically, elaborating on its potential gains and challenges.

Impact Statement

This thesis examined early antecedent variables in relation to important developmental outcomes using two different methodologies. Part one is a meta-analysis which reviewed a large body of research to collect and synthesize data about the association between early maternal sensitivity and cognitive development. Part two is a secondary analysis of a large set of data from the multi-phase longitudinal NICHD SECCYD study to examine a wide array of potential early predictors of behaviour problems in adolescence. These two research endeavours' strengths are complementary. On one hand, the meta-analytic findings are unique for their greater potential for generalizability. On the other hand, the longitudinal findings from the NICHD SECCYD study have the added value of factoring in the simultaneous contribution of multiple predictors, and thus are unique in terms of their reduced susceptibility to confounding effects. These two types of research findings combined shed some light on the nature of the relationship between early caregiving experiences and later development and pave the way for future research in that area of interest. Such research is crucial for informing the development of early prevention and intervention programs, which aim at targeting early environmental factors to reduce the risk of later psychopathology in offspring.

Moreover, one particularly exciting aspect of the empirical research project that was undertaken is the implementation of Machine Learning for the prediction of behaviour problems in adolescence. Although more research is needed before establishing the clinical utility of the models created by the predictive modeling process, these findings provide evidence supporting the feasibility of using Machine Learning to build models for the prediction of important mental health outcomes. The clinical translation of these models could change the face of psychiatry and psychology by being used as an adjunct to diagnosis and clinical judgment for making more accurate predictions regarding best treatment options and prognosis.

Lastly, although Machine Learning is not a novel analytic approach, it hasn't been used frequently in the field of mental health due to the challenges associated with its implementation. Thus, the empirical research project in Part two extended previous literature by using an unconventional approach to analysing data. Some of these difficulties were highlighted in the project and potential solutions were suggested for future mental health research aiming at using Machine Learning for prediction. Such preliminary applications of this approach are needed as stepping-stones based on which more rigorous research will be implemented in the future.

Table of Contents

List of Tables

Part 1: Literature Review

Part 2: Empirical Paper

List of Figures

Part 1: Literature Review

Part 2: Empirical Paper

Acknowledgments

I would like to express my gratitude to my supervisor, Prof Pasco Fearon, for the guidance, understanding, and reassurance he provided me. I would also like to say a special thank you to my reviewers, Dr Liam Mason and Dr Suzanne O'Rourke, as I am aware that a delayed viva may have been an inconvenience. I want to also thank Dr John King for his reassuring and encouraging emails. I would like to also mention my gratitude for the support I have received from Dr Katrina Scior and her consistent encouragements. I can never be thankful enough for the support I have received from the DClinPsy course, which made such a big difference in my journey throughout this doctorate. I also want to acknowledge the valuable support I have received from some of my fellow trainees, whom I will miss dearly. Finally, no words can ever express how grateful I am for having my family alive and well by my side. I hope this doctorate can pave the way for a better future for them.

Part I: Literature Review

Maternal Sensitivity and Cognitive Development: A Meta-Analysis

Abstract

Background: Research shedding light on the contribution of early caregiving experiences to later development can inform future prevention and intervention efforts. Sensitive parenting has been found to be associated with a variety of developmental outcomes. Previous meta-analyses examined parental sensitivity and closely related constructs in relation to two cognitive development outcomes, executive function and language.

Aims: The current meta-analysis aimed to examine the longitudinal association between observer-assessed maternal sensitivity and cognitive development outcomes indicative of general cognitive ability in offspring younger than 19 years. Moderator analyses were conducted to examine the potential effects of socioeconomic status, child age at maternal sensitivity assessment, and type of maternal sensitivity construct on the overall effect size.

Methods: A comprehensive search of the literature examining the association between the two constructs of interest was conducted in February 2020 using the following databases: PsycInfo, MEDLINE and Web of Science. One coder completed the data extraction. Robust Variance Estimation (RVE) was used to analyse dependent effect sizes. Correlated Effects with Small-Sample Corrections models were fitted to derive an overall effect size estimate and to test for moderating effects.

Results: Across 13 studies including 4939 mother-child dyads, the association between maternal sensitivity and cognitive development was found to be significant, with a small to moderate combined effect size ($r = 0.30$, $p < 0.01$, 95% CI: 0.22 to 0.38). Child age at maternal sensitivity assessment was found to be a significant

moderator (b = 4.66E-03, $p < 0.05$). No other significant moderating effects were found.

Conclusion: These findings are consistent with previous meta-analytic findings showing significant small to moderate associations between sensitive parenting constructs and other cognitive development outcomes. Interestingly, the moderating effect of child age was found to very small, which questions the importance of considering this finding in clinical research and treatment planning.

Introduction

Maternal Sensitivity

The literature on maternal sensitivity is based in large part on Mary Ainsworth's pioneering work, which examined the features of mother-infant interactions or early parenting that facilitate the development of secure attachments (Ainsworth, 1969; Grossmann et al., 2016). Ainsworth's construct of maternal sensitivity refers to a mother's ability to perceive and interpret her child's cues regarding protection and comfort accurately and to respond to these signals appropriately and in a timely manner (Ainsworth, 1969; Ainsworth and colleagues, 1974). Bretherton (2013) added that Ainsworth also described the dyadic nature of the construct, suggesting that when the mother is acting sensitively the interaction between herself and her child is mutually rewarding.

Following Ainsworth's ground-breaking work, several closely related definitions of sensitivity emerged and conceptual issues raised. Maternal sensitivity was described by some authors as a maternal communicative process, in which the mother communicates with her offspring by following a series of steps that involve the accurate perception and interpretation of signals followed by a prompt and appropriate response (Claussen & Crittenden, 2000; Nicholls & John Kirkland, 1996). Alternatively, it was defined as a dyadic process. According to this perspective, sensitivity is a characteristic that pertains to the mother-infant/child dyad rather than to the mother alone. Hence, the mother's behaviour is deemed sensitive only when it is attuned to the child's unique needs, and the degree of attunement is inferred by observing how both members of the dyad are interacting with each other (Claussen & Crittenden, 2000; Nicholls & John Kirkland, 1996). Sensitivity was also conceptualised as a set of skills which is best described as a multidimensional construct, possibly including but not limited to dimensions such as warm responding, synchronous interactions, stimulating parental behaviour and adaptiveness-promoting parenting. Finally, several issues related to the nature of the construct were raised. It was suggested that sensitive parenting is far from being a uniform construct and its nature is likely to change depending on the child's developmental stage as well as the context in which the interaction is taking place (Claussen & Crittenden, 2000). Sensitive parenting may take different forms depending on the age of the child as a result of changing developmental needs. Classical definitions of sensitive parenting, which primarily entail providing safety and protection, may be suitable for young infants. However, such parenting may be experienced as insensitive or intrusive by a pre-schooler, who developmentally needs to be provided with ample opportunities for exercising autonomy (Claussen & Crittenden, 2000). Moreover, what constitutes maternal sensitivity may differ depending on whether the child is in a threatening situation compared to a nonthreatening situation (Claussen & Crittenden, 2000). Similarly, depending on the role the parent is meant to play with the child (e.g., teacher, playmate, protector, etc.), sensitive responding may manifest differently. For example, being sensitive may entail different behaviours when the parent is acting as a teacher versus a playmate (Claussen & Crittenden, 2000). Finally, what may be construed as sensitive parenting, in a certain socioeconomic and cultural context, may not be the case in a different context (Claussen & Crittenden, 2000; Nicholls & John Kirkland, 1996). Shin et al. (2008) suggested, based on their review of articles between 1978 to 2007, that a contemporary conceptualisation of maternal sensitivity would describe this construct as a process that: involves a caregiver's abilities; consists of the caregiver behaving in particular ways (e.g., appropriately, promptly, in an emotionally available and expressive way, with awareness); depends on the child's behaviour; and entails a level of give-and-take between both members of the dyad.

The lack of uniformity in the way parental sensitivity has been conceptualized is also reflected in the way it has been operationalized and measured in the literature (Bohr et al., 2018; Yucel and Downey, 2010). Self-report as well as observational measures have been devised to assess maternal sensitivity. Although self-report measures are easier to implement, their validity is questionable due to the social desirability bias. Observational assessment tools are most commonly used because they rely on observing parents' behaviour rather than on their potentially biased self-reports (Yucel & Downey, 2010). However, the extent to which these various observational measures are conceptually equivalent is questionable. Bohr et al. (2018) examined four commonly used standardized measures of caregiver sensitivity in the same sample of mother-infant dyads: the Emotional Availability Scales (EAS), the Parent Child Interaction – Nursing Child Assessment Feeding Scale (PCI-NCAFS), the Maternal Behaviour Q-Sort (MBQS) and the original Ainsworth Maternal Sensitivity Scales (AMSS). Medium to high correlations between AMSS and each of the remaining measures showed that the latter measures do overlap conceptually with Ainsworth's original definition of sensitivity. However, these measures can't be considered as equivalent, considering that estimates of unshared variance among them were found to be large. Similar findings were reported by Dawson et al. (2018) when the ratings on the Maternal Behavior Q-sort mini and the original Ainsworth Scale were examined for congruence in a sample of South African mother-child dyads.

Maternal Sensitivity and Development

Early caregiving plays a pivotal role in a child's development. Good quality interactions with a parent/caregiver facilitates regulation and provide the necessary foundation for healthy development to occur. The earlier these interactions are present, the better the outcomes (Nelson et al. 2019). This is supported by research on children raised in institutions, which showed that the shorter the duration of institutionalized rearing and the earlier the provision of enhanced care, the better the outcomes across most developmental domains (Zeanah et al. 2011). Moreover, several longitudinal studies examining maternal sensitivity and closely related parental characteristics in relation to brain development reported an association between these aspects of parenting and structural changes in the brain across infancy (Bernier et al., 2016), childhood (Bernier et al., 2019; Wang et al., 2019) and adolescence (Whittle et al., 2014). Although causality can't be inferred from longitudinal research, these studies suggest that further research is needed to examine whether changes in brain development could have a mediating role with respect to the effect of maternal sensitivity on cognitive development (as well as other developmental outcomes).

Parental sensitivity has been examined extensively in relation to various developmental outcomes and significant associations have been found with attachment (De Wolff & van IJzendoorn, 1997), language, executive function, cognitive development as well as physical (e.g., obesity, sleep), and socio-emotional (e.g., behavioural problems, social competence, emotionality and temperament) outcomes (Deans, 2020; Fay-Stammbach et al., 2014). Of particular interest to this current meta-analysis is the association between maternal sensitivity and cognitive development. A recent meta-analysis by Madigan et al. (2019) examined the association between sensitive parenting and child language and reported a small to moderate correlation (r = 0.27) between these two variables. Moreover, in a review by Fay-Stammback et al. (2014), sensitivity/responsiveness was one of the parenting variables that was found to be most consistently associated with executive functioning. These findings were further supported by a meta-analysis by Valcan et al. (2018), which reported a significant small to moderate effect size ($r = .25$) for the association between positive parenting behaviour (including constructs such as sensitivity and responsiveness) and executive function.

The aim of the current meta-analysis is to examine sensitive parenting in relation to other potential indicators of cognitive development, namely intellectual ability and academic achievement. Moderator analyses are crucial for clarifying the nature of this association, by examining factors that may change, strengthen or attenuate the association between these two variables of interest. In the current meta-analysis, the moderator analyses included several pre-specified variables, which were selected based on the literature and available evidence. Prespecifying potential moderators is recommended to avoid data dredging and decrease the likelihood of false positive conclusions (Thompson and Higgins, 2002).

The sample-level potential moderators that were investigated were socioeconomic status and the child age at which maternal sensitivity was assessed. Only one study-level moderator was included, which is the type maternal sensitivity construct used in the study.

Socioeconomic status has been previously examined as a potential moderator in meta-analyses investigating the association between maternal sensitivity or closely related parenting constructs and cognitive outcomes (Madigan et al. 2019; Valcan et al., 2018). Inconsistent results were reported. In a meta-analysis by Madigan et al.

(2019), socioeconomic status was a significant moderator of the relationship between sensitive parenting and language development, with effect sizes being stronger in low and diverse SES samples compared to the middle to upper ones. This was not replicated by Valcan et al. (2018), where socioeconomic status was not found to have a moderating effect on the association between positive parenting behaviour (including warmth, sensitivity and responsiveness) and executive function. However, in this meta-analysis, it was noted that there wasn't enough between-study variability in socioeconomic status, which may have reduced the possibility of detecting a significant result.

Early childhood is a time when children are actively learning through their interactions with others. Emotional support provided by educators was identified as an important factor for fostering cognitive growth during infancy and toddlerhood (National Research Council, 2015). During this phase, the child's educators tend to be the parents or alternative primary caregivers. This suggests that parental sensitivity may have a greater influence on cognitive development during this period compared to the school years when teachers are more likely to be the child's primary educators. This further implies, in relation to the current meta-analysis, that the association between early parenting and cognitive outcomes may be stronger at a younger age. Thus, it is of interest to examine the child age at maternal sensitivity assessment as a potential moderator. The child age at the time of the assessment of early parenting behaviour has been examined as a moderator in several metanalyses examining early parenting behaviour and developmental outcomes (Madigan et al., 2019; Valcan et al., 2018). In the meta-analysis by Valcan et al. (2018), child age was found to be a significant moderator, with stronger effect sizes reported when parenting was assessed earlier. Conversely, no significant moderating effect was reported for child age by Madigan et al. (2019). Thus, considering the inconsistent results obtained for the sample-level moderators, socioeconomic status and child age, further research is needed to clarify their role in the relationship between parental sensitivity and later cognitive outcomes.

Moreover, considering the lack of uniformity in the definition of parenting sensitivity, its association with cognitive outcomes may differ depending the way it is defined and operationalized. Of particular interest is whether or not the definition included a stimulation component. Stimulation generally refers to the parent's cognitively stimulating way of interacting with the child, providing a rich environment conducive to learning and the development of new cognitive skills (Claussen & Crittenden, 2000). A definition of parental sensitivity which includes stimulation is conceptually more closely related to cognitive development than a definition excluding stimulation and mainly tapping into the comforting, protective and security promoting aspects of sensitive interactions. In a study by Page et al. (2010), the unique contributions of maternal sensitivity and verbal stimulation to cognitive development were examined cross-sectionally and compared. Only verbal stimulation (including the provision of guidance, descriptions and encouragements during a teaching task) turned out to be a significant unique predictor of infant's cognitive ability. Thus, it is important to examine whether or not the nature of the association with cognitive ability varies depending on whether or not the definition encompassed stimulation.

Determining the nature of the association between sensitive parenting and cognitive development is warranted. Several reviews and meta-analyses examining the impact of parenting interventions on parental sensitivity showed significant effects (Mountain et al., 2017; O'Hara et al., 2019; Rayce et al., 2017). Whether sensitive parenting plays an influential role on cognitive development is yet to be fully determined. Further research is needed to examine whether the increases in maternal sensitivity post-intervention mediates the potential impact of parenting interventions on outcomes such a cognitive development. A meta-analysis examining the association between parental sensitivity and cognitive development represents the first step toward clarifying whether more experimental research aimed at examining cause and effect relationships is needed in this area.

Research Aims

The primary objective of the current study is to use meta-analytic methods to synthesize findings and determine the nature (strength and direction) of the correlational association between maternal sensitivity and cognitive development. A secondary objective is to examine how this relationship changes depending on various characteristics related to the extracted data (such as socioeconomic status and child age at maternal sensitivity assessment) and the design of the studies included in the meta-analysis (the type of maternal sensitivity construct used as a study-level moderator).

Method

Definitions and Constructs

In the current study, scores on standardized tests of cognitive ability (e.g., mental development, intelligence, academic achievement) were selected as indicators of cognitive development. Several parenting constructs representing various conceptualisations of maternal sensitivity, closely related to Ainsworth's original definition, were included in the study (Ainsworth, 1969; Ainsworth et al., 1974). Ainsworth's original definition described maternal sensitivity as the mother's ability to notice and interpret her child's signals accurately and respond to them appropriately and in a timely manner. Only observational measures of maternal sensitivity were included in the study to reduce threats to construct validity. Observational assessment tools examined the actual mother's interactions with the child rather than their perceptions of their own parenting. Moreover, with respect to the conceptual definition of maternal sensitivity used in the current meta-analysis, only studies that examined this variable as an intra-personal maternal characteristic rather than an interpersonal/dyadic construct (e.g., synchrony) were considered eligible. Including constructs that reflected the behaviour of the child as well as the mother's risks to threaten the construct validity of the maternal sensitivity variable.

Search Strategy

Systematic literature searches were conducted using three databases: PsycInfo, MEDLINE and Web of Science. Keyword as well as subject heading searches were carried out. Search terms related to maternal sensitivity, cognitive development and longitudinal study design were combined (refer to Appendix A for the detailed search strategy). Following the removal of duplicate citations, the titles and abstracts of 2398 articles were initially scanned for inclusion/exclusion criteria. When the information provided by the titles and abstracts was inadequate, full-text articles were reviewed for eligibility criteria.

Study Inclusion and Exclusion Criteria

Studies were considered eligible according to the following inclusion/ exclusion criteria:

- Studies examining maternal sensitivity constructs closely related to the original definition of Mary Ainsworth and her colleagues (the mother's ability to perceive the infant's signals accurately, and the ability to respond to these signals promptly and appropriately) were included. Studies that used sensitivity measures that incorporated items taping into the child's behaviour were excluded.

- Only studies using observational instruments to measure maternal sensitivity were included, whereby maternal behaviour is assessed based on observations of mother-infant interactions.
- Only longitudinal studies, which examined antecedent maternal sensitivity in relation to later cognitive outcomes, were included. Some studies used maternal sensitivity composites of several measurements over time, with the last follow-up being concurrent with the child age at the assessment of cognitive development. These studies were included in the current meta-analysis as they still do examine the association longitudinally, considering that the maternal sensitivity composite was based on multiple earlier measurements as well as the one which was concurrent with the assessment of the cognitive development.
- The current meta-analysis aims to examine the relationship between antecedent maternal sensitivity and later cognitive ability as it naturally occurs without the influence of any experimental manipulation. Thus, intervention studies would only be considered eligible if data for the nontreated control group was available for both antecedent maternal sensitivity and later cognitive ability. However, no intervention study met this criterion and was included in the current meta-analysis.
- Only studies using standardized psychometric assessment tools to evaluate cognitive ability (e.g., Bayley scales of infant development, Wechsler Intelligence Scale for Children Revised) were included. Studies using non-

standardised or subjective tools (such as school grades, or maternal report) as cognitive outcomes were excluded.

- Only data about the relationship between antecedent maternal sensitivity and later cognitive outcomes across childhood and adolescence (age 19 and younger) was included.
- Only studies providing the required data for the analysis were included.
- Only studies published in a peer-reviewed journal were included.
- Only full-text articles that were available and written in English were included.
- Studies using typically developing samples were included.
- Studies using samples of children with diagnostic language delays, intellectual disabilities, deafness (in parents or children), hearing loss or middle ear disease, autism spectrum disorders, speech anomalies, and brain injuries were excluded.
- Studies using samples of children who were born preterm were excluded.
- Studies using samples of children whose mothers have a mental health condition (e.g. depression, addiction, etc.) were excluded.

Data Extraction

Studies meeting inclusion criteria were coded using a standard data extraction form. Potential moderators included the following: the child age at the time of maternal sensitivity assessment, the type of maternal sensitivity construct used (including versus excluding stimulation) and socioeconomic status. Previous related papers in the reference lists were reviewed when information about particular moderators was lacking from the studies included in the meta-analysis. Due to limited resources, only one coder completed the data extraction. Double coding was not conducted to examine whether or not the data extraction was reliable.

Study Quality Assessment

Study quality was assessed using the quality assessment tool for observational and cross-sectional studies from the National Institutes of Health (National Institute of Health, 2014) displayed in Appendix B. It is a 14-point measure that assesses the methodological quality of studies, with higher scores indicating better quality. A score of 1 was assigned when a criterion was met. When the information was lacking or unclear regarding a particular criterion, studies were assigned a score of 0 on that item. None of the studies had a score below equal or below 5 so there was no need to exclude any on the grounds of low methodological quality.

Data Synthesis and Analysis

Multiple effect sizes derived from the same participants or overlapping samples, differing in terms of how the antecedent or outcome variables were measured (e.g., variations in the operational definitions used or time point at which the assessment occurred), were extracted and included in the current analysis. Such effect sizes have correlated error estimates because they are derived from the same pool of subjects and were treated as dependent effect sizes. Incorrectly treating them as independent effect sizes would lead to erroneous conclusions, overestimating the precision of the resulting combined effect size (Cheung et al., 2019; Fisher & Tipton, 2015).

If a study used maternal sensitivity composites derived from several follow-up assessments, the related effect sizes were included in the analysis and the age at the last follow-up was treated as the assessment age in the moderator analyses. Thus, for a few studies, the estimate used for the child age at maternal sensitivity assessment seems concurrent with the child age at cognitive development assessment. However, this is not entirely correct because the maternal sensitivity score was a composite based on prior as well as concurrent assessments. If a study (or a set of studies using overlapping samples) reported full scale scores as well as subscale scores for cognitive outcomes, only the effect sizes based on full scale composites were extracted. The effect sizes reported by the studies using overlapping data from the NICHD SECCYD, which were redundant with other effect sizes (i.e. representing the same association as other effect sizes), were excluded from the quantitative synthesis. Finally, if a study reported effect sizes for different non-overlapping subsamples (e.g., sex and ethnicity), these effect sizes were included separately in the meta-analysis, as if they were extracted from different studies.

The Pearson product-moment correlation coefficient r was extracted from most of the studies as the effect size of interest. For a few studies, in which this statistic was not available, data from regression analyses was used to compute semi-partial correlations.

The Robust Variance Estimation (RVE) meta-analytic technique which is recommended for dependent effect sizes was used in the current metanalysis. The R packages "robumeta", "metafor", "dplyr", “foreign” and “esc” were installed to run the analysis (Fisher et al., 2017; Lüdecke et al., 2017; Team et al., 2020; Viechtbauer & Viechtbauer, 2015; Wickham & Wickham, 2020). Initially, Fisher’s z-transformation of the correlation coefficients was performed to get accurate weights for each study and the corresponding variances were calculated. RVE meta-regressions which involved the fitting of Correlated Effects Models with Small-

Sample Corrections were conducted using the R package robumeta (Fisher & Tipton, 2015; Fisher et al., 2017).

A RVE intercept-only model was initially run to calculate a combined effect size without the study covariates. The fitting of the intercept model was followed by a sensitivity analysis to examine the influence of varying rho values on the model. Rho represents the common correlation between the effect sizes within a study and can range from 0 to 1, with a default value of 0.8 (Hedges et al., 2010; Fisher & Tipton, 2015). A forest plot was produced to depict the studies' effect sizes and their corresponding confidence intervals and assigned weights. Three different RVE meta-regression models were run separately for each covariate (child age at maternal sensitivity assessment, type maternal sensitivity construct, and socioeconomic status). Finally, a RVE meta-regression model was fitted including all the study covariates. The test statistic I^2 was used to quantify the amount of variability in effect size estimates due to effect size heterogeneity as opposed to random variation. Publication was assessed using Egger's regression test (Viechtbauer & Viechtbauer, 2015).). This test can be used in meta-analyses including dependent effect sizes and using the RVE technique to estimate publication bias (Pustejovsky & Rodger, 2019). Refer to Appendix C for the R code used for running the above described analyses.

Results

Studies Selected

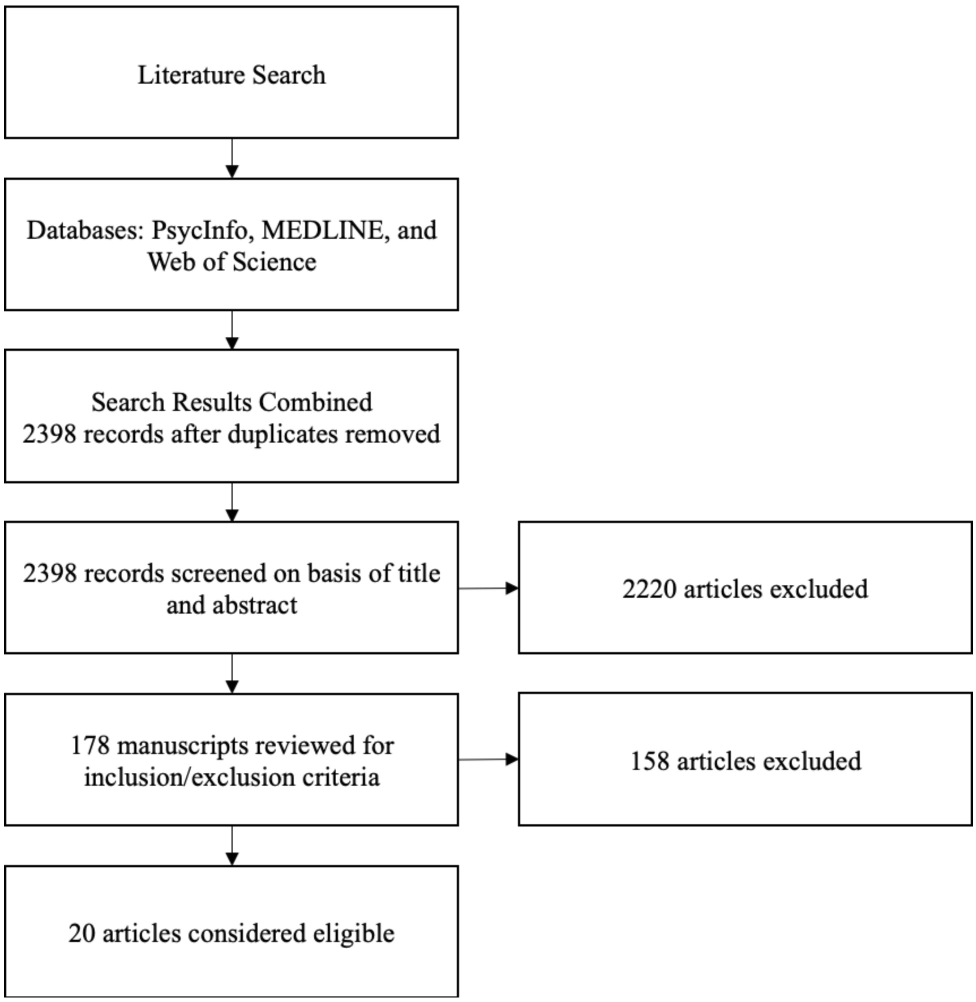
The PRISMA flow diagram in Figure 1 provides a description of the search strategy and the results it yielded. Three databases (PsycINFO, Medline and Web of Science) were searched, and duplicates were removed, yielding 2398 articles. After scanning the titles and abstracts, 178 articles were identified for full-text review, from which 20 articles were considered eligible. Effect sizes from 16 of these studies were considered suitable to be included in the quantitative synthesis based on the criteria for data extraction.

Study Characteristics

Study characteristics are reported in Tables 1 and 2 for the studies from which data was extracted. Table 1 provides brief definitions of the antecedent (maternal sensitivity) and outcome (cognitive ability) variable constructs used in each study. Table 2 presents other sample and methodological characteristics of the studies. Sample sizes ranged from 29 to 1143 mother-child dyads.

Figure 1

PRISMA Flow



With respect to Sex, the percentages of males in the sample across most of the studies varied around 50 % (with the exception of one study with two nonoverlapping samples, with 0 and 100 % males). With respect to socioeconomic status, only three studies used samples which were clearly identified as derived from disadvantaged populations, whereas the samples of the remaining studies were categorized as “other” (i.e., clearly identifiable or best described as “diverse”). With respect to the definition of maternal sensitivity, 13 studies used the definition excluding stimulation and 3 studies used the definition including stimulation. The average child age at maternal sensitivity assessment was 39 months (range 4 to 126). The average age of cognitive ability assessment was 85 months (range 15 to 216). A total of 16 studies were included in the meta-analysis: 11 studies were conducted in North America, 4 in the United Kingdom, and 1 cross-cultural study was conducted in North America and in Japan. Study quality ranged from 6 to 11, with a mean score of 8.33 and a standard deviation of 1.69. Refer to Table 3 for more details about the study’s quality assessment.

Table 1

Maternal Sensitivity and Cognitive Development Operational Definitions

Author/Research Team	Maternal Sensitivity	Stimulation Component	Cognitive Ability
Dotterer et al. (2012)	Sensitivity composite (across 12, 24 & 36 months): global sensitivity, detachment (reversed), positive regard, animation and stimulation	Yes	BBCS
Dunkel & Woodley of Menie (2019)	TSQ (items reflecting sensitivity)	No	WISC PIQ; WISC VIQ WAIS PIQ; WAIS VIQ
FCCC: Barnes & Melhuish (2017)	HOME (responsivity subscale)	No	BAS
FCCC: Sylva et al. (2011)[d]	Sensitivity composite: CIS (lack of detachment and positive relationships) and HOME (responsivity) Non-harshness composite: CIS (harshness) and HOME (avoidance of restrictions and punishment)	No	BSID MDI
Hann et al. (1996)	CABS: mother's sensitivity and responsiveness to her child's autonomous cues	No	S-B Form L-M

Hess et al. (1987)	Mother's sensitivity to the child's nonverbal messages, her awareness of the child's curiosity, and her comments about what the child was doing.	No	WISC (Vocabulary; Mathematics) ITBS (Vocabulary; Mathematics Concepts)
Ho (1987)	HOME (responsivity subscale)	No	BSID MDI g composite
Kelly et al. (1996)	MCPS: mother's ability to lead and follow the child in sensitive and responsive ways	No	WPPSI
Mills-Koonce et al. (2015)	Sensitivity composite: parental detachment (reversed), positive regard, stimulation, and animation	Yes	WPPSI (receptive verbal ability; block design) BSID MDI (6 -15 months)
Narvaez et al. (2013)	HOME (responsivity subscale)	No	BSID MDI
NICHD SECCYD: Cottrell et al. (2015)	Sensitivity composite of supportive presence, respect for autonomy of the child, and reflected hostility (reversed)	No	WJ (Math; Vocabulary; and Reading)
NICHD SECCYD: Fraley et al. (2013)	Sensitivity composite (across 6, 15, 24, and 36 months) of supportive presence, respect for autonomy, and hostility (reversed)	No	WJ composite

NICHD SECCYD: NICHD ECCRN (1999)	Sensitivity composite (across 6, 15, 24, and 36 months): sensitivity at 6, 15, and 24 months defined as distress, positive regard, and intrusiveness (reversed); and sensitivity at 36 months defined as supportive presence, respect for autonomy, and hostility (reversed)	No	BBCS
Pearson et al. (2011)	TIM: positive nonverbal responses items reflecting the quality and warmth of the mother's non-verbal responses towards her infant	No	GMDS DQ; WPPSI VAI; WPPSI NVI
Propper et al. (2012)[a]	Sensitivity composite (across 6, 12, 24, and 36 months): sensitivity/responsiveness, positive regard, stimulation, animation, and detachment/disengagement (reversed)	Yes	WJ subtest (reading letter-word) WJ subtest (math applied problems)
Raby et al. (2015)	Sensitivity composite (across 3, 6, 24, and 42 months): Ainsworth's sensitivity at 3- and 6-months follow-ups; supportive presence at 24- and 42-months follow-ups	No	PIAT (math; reading comprehension; reading recognition; and spelling) WJ (comprehension; calculation)

Note. BBCS = Bracken Basic Concepts Scale; TSQ = Teaching Strategies Q-Set; WISC = Wechsler Intelligence Scale for Children; WAIS: Wechsler Adult Intelligence Scale; PIQ = Performance IQ; VIQ = Verbal IQ; FCCC = The Family, Children, and Child Care study; HOME = Home Observation for the Measurement of the Environment; BAS = General Cognitive Ability; CIS = Caregiver Interaction Scale; BSID MDI = Bayley Scales of Infant Development Mental Development Index; g = General Cognitive Ability; CABS = Control-Autonomy Balance Scales; S-B = Stanford-Binet; ITBS = Iowa Test of Basic Skills; PCPS = Mother Child Play Scales; WPPSI = Wechsler Preschool and Primary Scale of Intelligence; NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; WJ = The Woodcock–Johnson Tests of Cognitive Abilities; NICHD ECCRN = National Institute of Child Health and Human Development Early Child Care Research Network; TIM = Thorpe Interaction Measure; GMDS = Griffiths Mental Development Scale; DQ = Developmental Quotient; VAI = Verbal Acquisition Index; NVI = Nonverbal Index; PIAT = Peabody Individual Achievement Test.

Table 2

Study Characteristics

Research Team	n	Country	Sex (% males)	SES	Age at maternal sensitivity assessment	Age at cognitive development assessment	Quality
Dotterer et al. (2012)	94[a]	North America (African Americans)	50	Disadvantaged	36[b]	36	8
Dotterer et al. (2012)	70[a]	North America (European Americans)	50	Disadvantaged	36[b]	36	8
Dunkel & Woodley of Menie (2019)	101[c]	North America	NA	Other/diverse	60	132, 216	6
FCCC: Barnes & Melhuish (2017)	1012[d]	United Kingdom	50	Other/diverse	10, 18	51	12

FCCC: Sylva et al. (2011)[d]		United Kingdom	50.2	Other/diverse	10	18	8
Hann et al. (1996)	69	North America	55	Other/diverse	20	30	6
Hess et al. (1987)	44	Japan	NA	Other/diverse	48	132	7
Hess et al. (1987)	47	North America	51	Other/diverse	48	144	7
Ho (1987)	274[c]	North America	NA	Other/diverse	12	24, 36, 48	6
Kelly et al. (1996)	29	North America	57	Other/diverse	20	60	6
Mills-Koonce et al. (2015)	629	North America	50.5	Disadvantaged	6, 24	15, 36	7
Narvaez et al. (2013)	336[d]	North America	NA	Disadvantaged	4, 8, 18, 30	24, 36	10
NICHD SECCYD: Cottrell et al. (2015)	1143[d]	North America	52[f]	Other/diverse	54	78, 102, 126, 180	10

NICHD SECCYD: Fraley et al. (2013)					36, 54, 78, 102, 126	54, 78, 102, 126, 180	10
NICHD SECCYD NICHD ECCRN (1999)					36[b]	36	11
Pearson et al. (2011)	732	United Kingdom	52	Other/diverse	12	18, 49	8
Propper et al. (2012)[a]	60	United Kingdom	100	Other/diverse	36	78	8
Propper et al. (2012)[a]	57	United Kingdom	0	Other/diverse	36	78	8
Raby et al. (2015)	243[c]	North America	55	Disadvantaged	42	78, 90, 102, 138, 192	10

Note. FCCC = The Family, Children, and Child Care study; NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (SECCYD); NICHD ECCRN = National Institute of Child Health and Human Development Early Child Care Research Network.
[a] These studies reported effect sizes based on nonoverlapping samples.
[b] The extracted effect size was based on a maternal sensitivity composite across several follow-ups and the current meta-analysis used the child age at the last follow-up as the age at maternal sensitivity assessment for the moderator analyses.
[c] Sample size ranges were reported, and the midpoint (rounded up) was used as the sample size for the study.
[d] Different sample sizes based on overlapping within the same study were reported and the average of these sample sizes was used as the overall sample size for the study.
[e] Pearson correlation coefficients were not reported, thus semi-partial correlations were computed instead from available regression coefficient estimates.
[f] Average percent males was computed across all NICHD SECCYD studies

Table 3

Study Quality Assessment

Author/ Research Team	1. Research Question or Objective Stated	2. Study population specified & defined	3. Was the participation rate of eligible persons at least 50%?	4. Subjects from the same or similar populations incl. same time period	5. Sample size justification, power description or variance and effect estimates	6. Exposure of interest measured prior to the outcome	7. Timeframe sufficient	8. different levels of the exposure in relation to outcome	9. Exposure measures or independent variables clearly defined, valid and reliable	10. Exposure assessed more than once over time	11. Outcome measures clearly defined, valid and reliable	12. Outcome assessors blinded to the exposure status of participants	13. Loss to follow-up after baseline 20% or less?	14. Confounding variables measured and adjusted statistically	QA score
Dotterer et al. (2012)	Yes	No	NR	NR	NR	Yes	Yes	Yes	No	Yes	Yes	NR	Yes	Yes	8
Dunkel & Woodley of Menie (2019)	Yes	NR	NR	NR	No	Yes	Yes	Yes	No	No	Yes	No	NR	Yes	6
FCCC - Barnes & Melhuish (2017)	Yes	Yes	Yes	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	Yes	No	12
FCCC - Sylva et al. (2011)	Yes	No	Yes	Yes	NR	Yes	No	Yes	Yes	No	Yes	NR	CD	Yes	8
Hann et al. (1996)	Yes	No	NR	NR	NR	Yes	No	Yes	Yes	No	Yes	NR	NR	Yes	6
Hess et al. (1987)	Yes	Yes	No	CD	No	Yes	Yes	Yes	No	No	No	NR	Yes	Yes	7
Ho (1987)	Yes	No	NR	NR	NR	Yes	Yes	Yes	Yes	No	Yes	NR	CD	No	6
Kelly et al. (1996)	Yes	No	NR	NR	NR	Yes	Yes	Yes	Yes	No	Yes	NR	No	No	6
Mills-Koonce et al. (2015)	Yes	No	NR	NR	NR	Yes	Yes	Yes	Yes	Yes	No	NR	NR	Yes	7
Narvaez et al. (2013)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	10
NICHD SECCYD, Cottrell et al., 2015)	Yes	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	NR	No	Yes	10
NICHD SECCYD (Fraley et al., 2013)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	10
NICHD SECCYD (NICHD ECCRN, 1999)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	Yes	Yes	11

Pearson et al. (2011)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	No	No	Yes	NR	NR	Yes	8
Propper et al. (2012)	Yes	No	NR	Yes	Yes	Yes	Yes	Yes	No	Yes	No	NR	No	Yes	8
Raby et al. (2015)	Yes	Yes	NR	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	10

Note. FCCC = The Family, Children, and Child Care study; NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (SECCYD); NICHD ECCRN = National Institute of Child Health and Human Development Early Child Care Research Network.

Meta-Analyses

An effect size which deviates 3 standard deviations from the overall mean of effect sizes was considered an outlier. One effect size reported by Propper et al. (1998) was identified as an outlier and discarded from subsequent analyses. A sample of 54 effect sizes extracted from 13 studies were used for all subsequent analyses. The number of effect sizes extracted per study was 4 on average, with a range of 1 to 16.

Intercept-Only Model

The fitting of a RVE intercept-only model was done initially to obtain an overall estimate of the effect size, without taking into account the effect of the study's covariates. The weighted average effect size of the association between maternal sensitivity and cognitive ability was estimated to be 0.31 (SE = 0.04, $p <$ .01, 95% CI: 0.22 to 0.40). A small to moderate combined effect size was found for the association between maternal sensitivity and cognitive development, when the estimates resulting from Fisher'z transformation were converted to correlation coefficients ($r = 0.30$, $p < 0.01$, 95% CI: 0.22 to 0.38).

The I^2 statistic was found to be 92.41%, indicating that 92.41% of the total variability in effect sizes is due to true heterogeneity or between-studies variability rather than random variation. An I^2 statistic of 75% and above indicates considerable heterogeneity and suggests that potential moderators should be examined in subsequent analysis (Del Re, 2015; West et al. 2010). The degrees for freedom were above 5 (11.6), which indicates that the model results are reliable.

Sensitivity Analysis

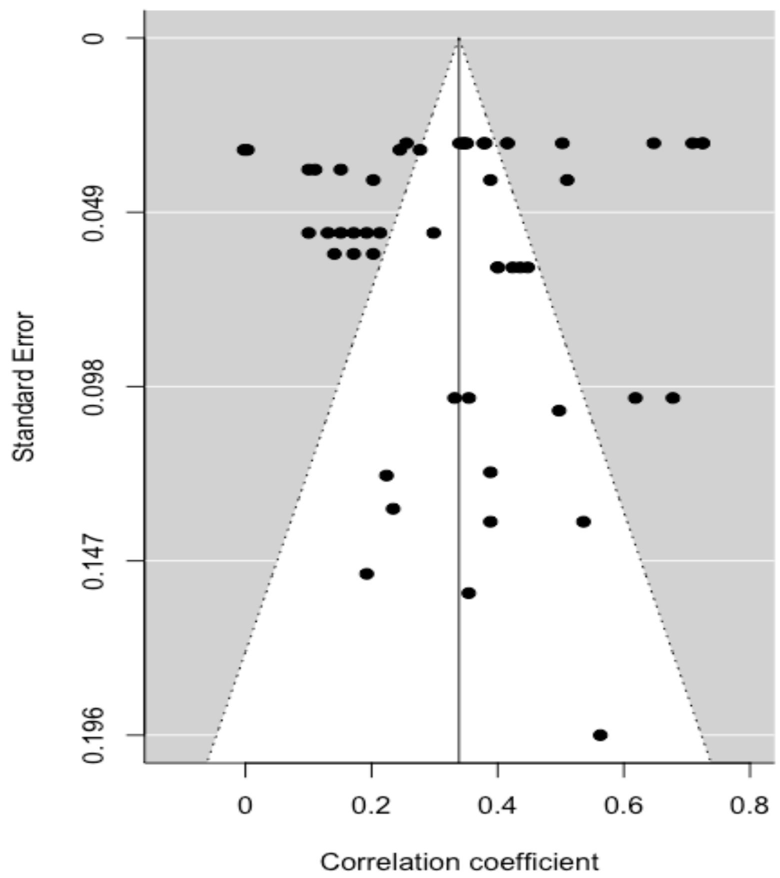
Sensitivity analyses showed that using different values for the within-study effect size correlation did not affect the intercept-only model results (test values ranged from 0 to 1 in 0.20 intervals). The findings remained consistent, with a combined effect size of 0.31, despite the change in the value of the within-study effect size correlation (rho), ranging from 0 to 1.

Publication Bias

A visual inspection of the funnel (Figure 2) showed that the effect sizes were spread out almost symmetrically around the pooled effect size (the striped line), suggesting the absence of publication bias, which was confirmed by the insignificant result obtained on Egger's regression test. This implies that there was no bias toward publishing studies with high effect sizes.

Figure 2

Funnel Plot

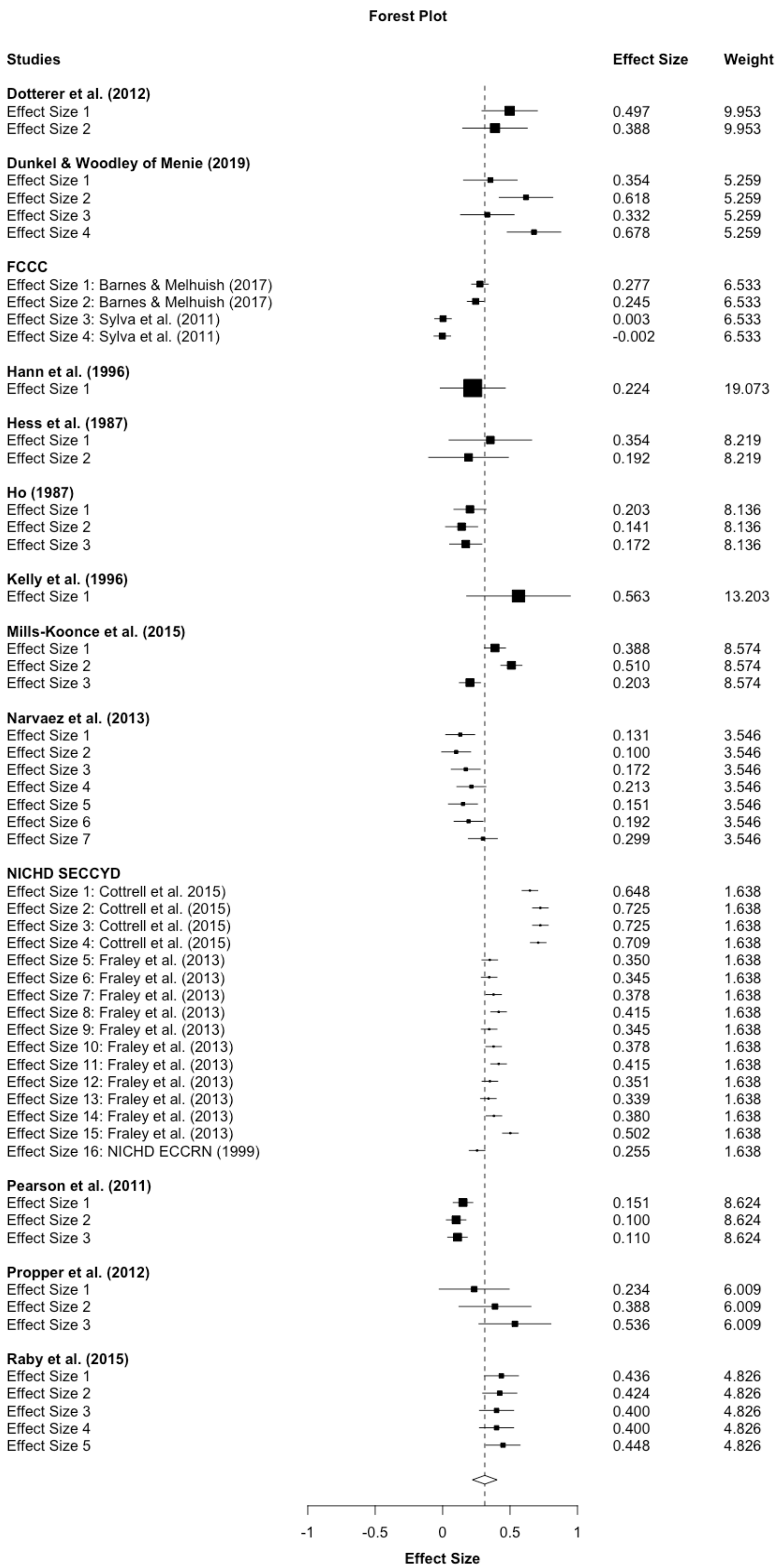


Forest Plot

The forest plot (Figure 3) displays the effect size estimates as well as their corresponding confidence intervals and weights, taking into account the correlated effects dependence structure of the extracted data, with effect sizes derived from overlapping samples being subsumed under the same study.

Figure 3

Forest Plot



Moderator Analyses

The moderator analyses consisted of two steps. First, potential moderators were examined one at a time. A RVE Correlated Effects with Small-Sample Corrections model was fitted for each potential moderator, to examine the effect of these variables separately on the association between maternal sensitivity and cognitive ability. Second, the fitting of a RVE meta-regression model was implemented including all the potential moderators simultaneously. The moderators examined in the current study were child age at maternal sensitivity assessment, socioeconomic status (0 = disadvantaged; 1 = other/diverse), and maternal sensitivity construct (0 = excluding stimulation; 1 = including stimulation).

Initially, when the moderators were examined separately, no significant moderating effects were found. Table 4 displays the output of the Correlated Effects with Small-Sample Corrections models for each moderator. None of the moderators had significant coefficient estimates. However, the Satterthwaite degrees of freedom values below 4 indicate that the current findings for two moderators (child age at maternal sensitivity assessment and type of maternal sensitivity construct) are not reliable and should be interpreted with caution

Table 4

RVE Correlated Effects with Small-Sample Corrections Models per Moderator

Models	Covariate (moderator)	Estimate	SE	t	df	95% CI	
						Lower	Upper
RVE Model 1	Age at maternal sensitivity assessment	4.36E-03	1.59E-03	2.74	3.73	-1.85E-04	8.90E-03
RVE Model 2	Maternal sensitivity construct (0 = excl. stimulation; 1 = incl. stimulation)	0.11	0.06	1.91	3.19	-0.07	0.28
RVE model 3	Socioeconomic status (0 = disadvantaged; 1 = other/diverse)	-0.05	0.08	-0.63	6.06	-0.25	0.15

Note. Categorical variables were dummy coded and the category with a value of 0 was treated as the reference group.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5

RVE Correlated Effects with Small-Sample Corrections Meta-Regression Model

	Estimate	SE	t	df	95% CI Lower	95% CI Upper
Intercept	0.18	0.06	3.23	2.88	-1.80E-03	0.36
Age at maternal sensitivity assessment	4.66E-03*	1.33E-03	3.50	4.04	9.77E-04	0.01
Maternal sensitivity construct (0 = no stimulation; 1 = stimulation)	0.11	0.04	2.77	2.88	-0.02	0.23
Socioeconomic status (0 = disadvantaged; 1 = other/diverse)	-0.04	0.05	-0.90	3.08	-0.20	0.11

Note. Categorical variables were dummy coded and the category with a value of 0 was treated as the reference group. **p* < .05. ***p* < .01. ****p* < .001.

When the effects of the moderators were examined simultaneously, different results were found. Table 5 displays the findings of the RVE Correlated Effects with Small-Sample Corrections meta-regression model, which was fitted including all the moderators as covariates. The child age at maternal sensitivity assessment was found to be a significant moderator (b = 4.66E-03, $p < 0.05$), with a Satterthwaite degrees of freedom value above 4 (4.04), indicating that this finding is reliable. The other two moderators, type of maternal sensitivity construct and socioeconomic status, were not found to be significant. However, the Satterthwaite degrees of freedom values for these two variables were below 4, indicating that the reliability of these insignificant results is questionable.

Discussion

The main objective of this meta-analysis was to examine the magnitude and direction of the correlation between maternal sensitivity and cognitive development. A variety of measures have been used to assess these two variables in the literature. The current meta-analysis focused on examining maternal sensitivity as an intra-personal construct in relation to several indicators of cognitive development that are more reflective of a general cognitive ability (e.g., intelligence, academic achievement, etc.).

The findings of the intercept-only model reliably showed a significant small to moderate effect size ($r = 0.30$, $p < .05$) for the association between maternal sensitivity and cognitive development. The direction of the effect size indicates that the greater the sensitivity of the mother, the better are the child's cognitive outcomes. This is in line with the findings of two previous meta-analyses by Madigan et al. (2019) and Valcan et al. (2018), which examined early sensitivity in relation to language development (r = 0.27) and executive function (r = 0.25). This effect size estimate is based on the intercept-only model results, which were found to be reliable as indicated by a Satterthwaite degree of freedom value above 4 (11.6). Tests of publication bias did not yield significant results lending more confidence in the validity of the current findings (Petticrew and Gilbody, 2004). Sensitivity analysis showed that the effect size estimate is robust and was insensitive to changes in the value of the within-study effect size correlation.

This meta-analysis also examined the possible moderating effects of sample and study design characteristics on the association between maternal sensitivity and cognitive development. However, only a limited number of prespecified variables were examined as potential moderators. Potential moderator variables were prespecified based on the literature and empirical evidence to avoid data dredging. Extending moderator analyses to include variables that were specified post hoc, based on patterns in the data extracted, is not recommended because it may increase the risk of type 1 error (Thompson and Higgins, 2002). Moreover, limiting the number of moderators as much as possible to get robust results was necessary considering the small sample of studies included in the current meta-analysis (Thompson and Higgins, 2002).

Child age at maternal sensitivity assessment was found to be the only significant moderator. Interestingly, when moderators were initially examined separately, the effect of age was not significant. However, when it was entered simultaneously with the other moderators in the meta-regression model, it emerged as reliably significant (b = 4.66E-03, $p < .05$), with a Satterthwaite degrees of freedom value above 4. The model coefficient estimate indicated that for every unit increase in the child age at maternal sensitivity assessment, there was an average

0.005 increase in the overall effect size estimate. This suggests that the effect size of the association between maternal sensitivity and cognitive development was greater when maternal sensitivity was assessed at an older age. This is inconsistent with prior meta-analytic findings examining positive parenting constructs in relation to cognitive outcomes. Valcan et al. (2018)'s meta-analysis did report significant moderating effect for child age, yet the direction of the effect was the opposite, suggesting that effect sizes were stronger in younger children. Conversely, in another meta-analysis by Madigan et al (2018), no significant moderating effect for this variable was found. It is worth noting that different indicators of cognitive ability were used as the developmental outcomes of interest in the current as well as these two previous meta-analyses (Madigan et al. 2018; Valcan et al., 2018). In the current meta-analysis, standardised measures of intelligence, mental development and academic achievement were used as indicators of general cognitive ability, whereas in the two meta-analyses by Madigan et al. (2018) and Valcan et al. (2018), language development and executive function were used as cognitive outcomes, respectively. Thus, the nature of the relationship between maternal sensitivity and cognitive ability may vary depending on the type of cognitive outcome used or cognitive domain being assessed. Different aspects of cognitive development may have different developmental trajectories, with some possibly maturing at a younger age and others showing significant improvements until adulthood. A narrative review by Best and Miller (2010) suggested that distinct facets of executive functions, such as inhibition and working memory, mature somewhat differently from early childhood (as early as age 4) to adulthood (to the early twenties). Specifically, inhibition seemed to mature mostly in preschool and undergo less fundamental improvements in later years, whereas the development of working memory seemed to be more linear from early childhood to adulthood. Finally, with respect to the moderating role of child age, it is important to mention that the effect was extremely small, bringing into question the utility of taking into consideration this finding in clinical practice and research.

Moreover, the shift in the significance of the results related to child age at maternal sensitivity assessment, when examined individually versus simultaneously with other moderators in the current study, indicate the potential presence of a negative confounder. A negative confounder is related to another variable and masks its effect on an outcome, because it is also related to the outcome, however in the opposite direction. It may be that socioeconomic status was the negative confounder since the average child age at maternal sensitivity assessment tended to be higher in diverse study samples (M = 45) compared to the disadvantaged ones (M = 24 months). One would speculate that the effects of both variables compensated for each other, whereby a unit increase in child age at maternal sensitivity assessment led to an increase in the combined effect size estimate, while this was the opposite for socioeconomic status (when it shifted from 0 to 1, i.e. from disadvantaged to other/diverse). This highlights the importance of examining the effects of potential moderators simultaneously in meta-regression models to reduce the bias that may be introduced by such confounding effects.

The type of maternal sensitivity construct was not found to be a significant moderator, when it was examined alone as well as simultaneously with other moderators. This is inconsistent with a study by Page et al. (2010), in which maternal sensitivity and stimulation were found to be differentially predictive of cognitive development, with stimulation being the only significant predictor. Yet, it is worth noting that the insignificant findings with respect to this moderator are unreliable and should be interpreted with caution, as indicated by Satterthwaite degrees of freedom values below 4. The unreliability of these findings could have been caused by the fact that the type of maternal sensitivity construct variable was an unbalanced moderator (Fisher and Tipton, 2015). Only 8 out of 54 effect sizes were associated with a maternal sensitivity operationalization, which included a stimulation component. More research is needed to examine maternal sensitivity empirically as a multidimensional construct. Previous studies reported evidence suggesting that maternal emotional support and cognitive stimulation are independent predictors of cognitive outcomes (Hubbs-Tait et al., 2002; Leerkes et al., 2011; Quittner at al., 2012). It would be interesting to examine and compare the contribution of different components of maternal sensitivity, including but not limited to emotional support and cognitive stimulation, on development and other important health outcomes. Such research would shed light on the ingredients of maternal sensitivity that potentially may need to be targeted in parenting interventions depending on particular outcomes of interest.

Similarly to the findings related to type of maternal sensitivity construct, socioeconomic status was not found to be a significant moderator. This is inconsistent with previous findings suggesting that positive parenting may have a protective buffering effect on the negative impact of socioeconomic status on cognitive development (Lee et al., 2019; Madigan et al., 2019). However, it is worth noting that in the current studies, the examination of socioeconomic status as a moderator consisted of comparing effect sizes derived from “disadvantaged” versus “other” (mainly diverse samples). It is possible to find larger differences when using samples that are vary more in terms of socioeconomic status, as shown by prior meta-analytic findings (Madigan et al, 2019).

One of the main strengths of the current meta-analysis is its reliance on longitudinal studies for the data extraction. Such design has more methodological rigor and is more likely to yield robust estimates of effects. Although it does not provide the grounds for inferring causality, it is closer to examining the nature of the effect of an antecedent variable on an outcome compared to the cross-sectional design. In addition, a cautious approach toward choosing the variables of interest was adopted. All of the studies included were of acceptable quality, as assessed based on the National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (National Institute of Heatlh, 2014). Moreover, careful considerations were made regarding how the variables of interest should be defined and operationalized. To minimize threats to construct validity, only studies that used observational tools for maternal sensitivity and standardized psychometric measures for cognitive development were included. In addition, to avoid any confounding of maternal sensitivity with other interpersonal constructs tapping into the mother-child relationship quality, only studies that defined this variable as an intra-personal maternal characteristic were included. All these measures were taken to maximise the likelihood of obtaining a truer effect size estimate.

However, several limitations with the design and statistical analyses should be mentioned. First, this meta-analysis only included observational studies which are correlational in nature and cannot provide the evidence required to make causal inferences. To examine causality, experimental research is needed to test whether or not the change in maternal sensitivity due to parenting interventions mediate the impact of these interventions on cognitive development. Second, the data extracted was based on typically developing samples. Thus, the current findings do not generalise to children with or those at risk of cognitive delay. Research aimed at clarifying the nature of the association between maternal sensitivity and cognitive development in these populations is extremely valuable as it might shed light on the clinical utility of parenting interventions in improving cognitive health in the context of cognitive delay. Third, the generalisability of the findings is also limited because the data was only based on mother-child dyads and the influence of fathers' sensitive parenting on cognitive development may be different. Fourth, the current study did not comprehensively examine possible sources of heterogeneity in effect sizes. Based on a closer examination of the patterns in the extracted data, a few variables stood out as potential moderators (e.g., whether the indicator of cognitive ability was a composite score of various subtests or a subtest score; whether the type of tasks used as part of the observational measure of maternal sensitivity were structured versus unstructured tasks; maternal sensitivity based on one assessment versus or multiple assessments over time). However, due to the small sample size, it was not possible to include these variables as moderators in the meta-regression model. Sixth, various aspects of the data extracted as well as the statistical analysis may have undermined the validity of the results including but not limited to the small sample size and unbalanced moderators. The sample of studies included was small and did not allow for the examination of a wide array of moderators. This is a likely limitation of many studies conducting meta-regression (Thompson & Higgins, 2002). To obtain meaningful results from a meta-regression, a large ratio of studies to moderator is needed, with some studies recommending a ratio of at least ten studies per moderator (Borenstein et al., 2009). The current meta-analysis potentially examined too many moderators (3 moderators) for the total number of studies included (13 studies) which may have compromised the robustness of the results. The potential presence of unbalanced moderators may have been the reason why the model fitting led to unreliable results as indicated by the Satterthwaite degrees of freedom below 4.

In conclusion, the current meta-analysis showed that there is a significant small to moderate association between early maternal sensitivity and cognitive development. Sensitivity analysis as well as Egger's test of publication indicated that this finding is reliable. Moreover, the results also revealed a high level of heterogeneity, which further highlights the importance of examining potential moderators to clarify the potentially complex nature of this association. With respect to the moderator analysis, the validity of the results is questionable, potentially due to the small sample of studies and the presence of unbalanced moderators. However, for child age at maternal sensitivity assessment, the findings were more reliable and shed light on the potential greater impact of later compared to early maternal sensitivity on subsequent cognitive outcomes. This may have significant clinical implications, for example with respect to designing parenting interventions aimed at enhancing maternal sensitivity and ultimately promoting cognitive health. Finally, given that the generalisability of the current findings is limited, meta-analyses aimed at examining sensitive parenting in relation to cognitive outcomes in atypically developing samples or father-child dyads are warranted to further clarify the nature of this association.

References

Ainsworth, M. D. S., & Bell, S. M. (1969). Some contemporary patterns in the feeding situation. In A. Ambrose (Ed.), Stimulation in early infancy (pp. 133–170). London: Academic Press.

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. (1974). Infant–mother attachment and social development: "Socialization" as a product of reciprocal responsiveness to signals. In P. M. Richards (Ed.), The integration of a child into a social world (pp. 99–135). Cambridge: Cambridge University Press.

Barnes, J., & Melhuish, E. C. (2017). Amount and timing of group-based childcare from birth and cognitive development at 51 months: A UK study. *International Journal of Behavioral Development*, *41*(3), 360-370.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, *81*(6), 1641-1660.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Chichester, West Sussex, U.K.; Hoboken: John Wiley & Sons.

Bretherton, I. (2013). Revisiting Mary Ainsworth's conceptualization and assessments of maternal sensitivity-insensitivity. *Attachment & Human Development*, *15*(5-6), 460-484.

Bohr, Y., Putnick, D. L., Lee, Y., & Bornstein, M. H. (2018). Evaluating Caregiver Sensitivity to Infants: Measures Matter. *Infancy: the official journal of the International Society on Infant Studies*, *23*(5), 730–747.

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology review*, 1-10.

Claussen, A. H., & Crittenden, P. M. (2000). Maternal sensitivity. In P. M. Crittenden and A. H. Claussen (Eds). *The organization of attachment relationships: Maturation, culture, and context* (pp. 115-124). New York: Cambridge University Press.

Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the black–white gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology*, *100*(6), 1713.

Dawson, N., Bain, K., & Mesman, J. (2018). Comparing two measures of maternal sensitivity: goodness of fit with a South African cultural context. *Attachment & Human Development*, 1-9.

Deans, C. L. (2020). Maternal sensitivity, its relationship with child outcomes, and interventions that address it: a systematic literature review. *Early Child Development and Care*, *190*(2), 252-275.

De Wolff, M. S., & Van Ijzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child development*, *68*(4), 571-591.

Del Re, A. C. (2015). A practical tutorial on conducting meta-analysis in R. *The Quantitative Methods for Psychology*, *11*(1), 37-50.

Dotterer, A. M., Iruka, I. U., & Pungello, E. (2012). Parenting, race, and socioeconomic status: Links to school readiness. *Family Relations*, *61*(4), 657-670.

Dunkel, C. S. (2019). Maternal sensitivity and performance and verbal intelligence in late childhood and adolescence. *Journal of Biosocial Science*, *51*(1), 48.

Fay-Stammbach, T., Hawes, D. J., & Meredith, P. (2014). Parenting influences on executive function in early childhood: A review. *Child Development Perspectives*, *8*(4), 258-264.

Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv preprint arXiv:1503.02220*.

Fisher, Z., Tipton, E., Zhipeng, H., & Fisher, M. Z. (2017). Package 'robumeta'. *Retrieved from* http://cran.uni-muenster.de/web/packages/robumeta/robumeta.pdf

Fraley, R. C., Roisman, G. I., & Haltigan, J. D. (2013). The legacy of early experiences in development: formalizing alternative models of how early experiences are carried forward over time. *Developmental psychology*, *49*(1), 109.

Grossman, K. E., Bretherton, I., Waters, E., & Grossman, K. (Eds.). (2016). *Maternal Sensitivity: Mary Ainsworth's Enduring Influence on Attachment Theory, Research, and Clinical Applications*. Routledge.

Hann, D. M., Osofsky, J. D., & Culp, A. M. (1996). Relating the adolescent mother—child relationship to preschool outcomes. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, *17*(4), 302-309.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, *1*(1), 39-65.

Hess, R. D., Azuma, H., Kashiwagi, K., Holloway, S. D., & Wenegrat, A. (1987). Cultural variations in socialization for school achievement: Contrasts between Japan and the United States. *Journal of Applied Developmental Psychology*, *8*(4), 421-440.

Ho, H. Z. (1987). Interaction of early caregiving environment and infant developmental status in predicting subsequent cognitive performance. *British Journal of Developmental Psychology*, *5*(2), 183-191.

Hubbs-Tait, L., Culp, A., Culp, R., & Miller, C. (2002). Relation of Maternal Cognitive Stimulation, Emotional Support, and Intrusive Behavior during Head Start to Children's Kindergarten Cognitive Abilities. *Child Development, 73*(1), 110-131.

Kelly, J. F., Morisset, C. E., Barnard, K. E., Hammond, M. A., & Booth, C. L. (1996). The influence of early mother—child interaction on preschool cognitive/linguistic outcomes in a high-social-risk group. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, *17*(4), 310-321.

Lee, D. B., Assari, S., Miller, A. L., Hsieh, H. F., Heinze, J. E., & Zimmerman, M. A. (2019). Positive Parenting Moderates the Effect of Socioeconomic Status on Executive Functioning: A Three-Generation Approach. *Journal of child and family studies*, *28*(7), 1878-1885.

Leerkes, E. M., Blankson, A. N., O'Brien, M., Calkins, S. D., & Marcovitch, S. (2011). The Relation of Maternal Emotional and Cognitive Support During Problem Solving to Pre-Academic Skills in Preschoolers. *Infant and child development*, *20*(6), 353–370.

Lüdecke, D., Lüdecke, M. D., & Calculator'from David, B. W. (2017). Package ‘esc’. Retrieved from https://mran.microsoft.com/snapshot/2018-04-29/web/packages/esc/esc.pdf

Madigan, S., Prime, H., Graham, S. A., Rodrigues, M., Anderson, N., Khoury, J., & Jenkins, J. M. (2019). Parenting behavior and child language: a meta-analysis. *Pediatrics*, *144*(4).

Mills-Koonce, W. R., Willoughby, M. T., Zvara, B., Barnett, M., Gustafsson, H., Cox, M. J., & Family Life Project Key Investigators. (2015). Mothers' and fathers' sensitivity and children's cognitive development in low-income, rural families. *Journal of Applied Developmental Psychology*, *38*, 1-10.

Mountain, G., Cahill, J., & Thorpe, H. (2017). Sensitivity and attachment interventions in early childhood: A systematic review and meta-analysis. *Infant Behavior and Development*, *46*, 14-32.

Narvaez, D., Gleason, T., Wang, L., Brooks, J., Lefever, J. B., Cheng, Y., & Centers for the Prevention of Child Neglect. (2013). The evolved development niche: Longitudinal effects of caregiving practices on early childhood psychosocial development. *Early childhood research quarterly*, *28*(4), 759-773.

National Institutes of Health. Study Quality Assessment Tools. (2014). Available at: www.nhlbi.nih.gov/health- pro/guidelines/in-develop/ cardiovascular-risk-reduction/tools/ cohort. Accessed October 15, 2018

National Research Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. National Academies Press.

National Institute of Child and Human Development Early Child Care Research Network. (1999). Chronicity of Maternal Depressive Symptoms, Maternal Sensitivity, and Child Functioning at 36 Months. *Developmental Psychology*, *35*(5), 1297-1310. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=yrovftd&NEWS=N&AN=00063061-199909000-00012.

Nelson III, C. A., Zeanah, C. H., & Fox, N. A. (2019). How early experience shapes human development: The case of psychosocial deprivation. *Neural Plasticity*.

Nicholls, A., & Kirkland, J. (1996). Maternal sensitivity: A review of attachment literature definitions. *Early Child Development and Care*, *120*(1), 55-65.

O'Hara, L., Smith, E. R., Barlow, J., Livingstone, N., Herath, N. I., Wei, Y., ... & Macdonald, G. (2019). Video feedback for parental sensitivity and attachment security in children under five years. *Cochrane Database of Systematic Reviews*, (11).

Page, M., Wilhelm, M. S., Gamble, W. C., & Card, N. A. (2010). A comparison of maternal sensitivity and verbal stimulation as unique predictors of infant social–emotional and cognitive development. *Infant Behavior and Development*, *33*(1), 101-110.

Pearson, R. M., Heron, J., Melotti, R., Joinson, C., Stein, A., Ramchandani, P. G., & Evans, J. (2011). The association between observed non-verbal maternal responses at 12 months and later infant development at 18 months and IQ at 4 years: A longitudinal study. *Infant Behavior and Development*, *34*(4), 525-533.

Petticrew, M., & Gilbody, S. (2004). Planning and conducting systematic reviews. *Health psychology in practice. Oxford: BPS Blackwell Publishers*, 150-79.

Propper, C. B., Shanahan, M. J., Russo, R., & Mills-Koonce, W. R. (2012). Evocative gene–parenting correlations and academic performance at first grade: An exploratory study. *Development and psychopathology*, *24*(4), 1265-1282.

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. Research synthesis methods, 10(1), 57–71.

Quittner, A. L., Cruz, I., Barker, D. H., Tobey, E., Eisenberg, L. S., Niparko, J. K., & Childhood Development after Cochlear Implantation Investigative Team (2013). Effects of maternal sensitivity and cognitive and linguistic stimulation on cochlear implant users' language development over four years. *The Journal of pediatrics*, *162*(2), 343–8.e3.

Raby, K. L., Roisman, G. I., Fraley, R. C., & Simpson, J. A. (2015). The enduring predictive significance of early maternal sensitivity: Social and academic competence through age 32 years. *Child development*, *86*(3), 695-708.

Rayce, S. B., Rasmussen, I. S., Klest, S. K., Patras, J., & Pontoppidan, M. (2017). Effects of parenting interventions for at-risk parents with infants: A systematic review and meta-analyses. *BMJ Open, 7*(12).

Shin, H., Park, Y. J., Ryu, H., & Seomun, G. A. (2008). Maternal sensitivity: A concept analysis. *Journal of Advanced Nursing*, *64*(3), 304-314.

Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L. E., & FCCC-team. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. *British Journal of Developmental Psychology*, *29*(1), 18-45.

Team, R. C., Bivand, R., Carey, V. J., DebRoy, S., Eglen, S., Guha, R., ... & Pfaff, B. (2020). Package ‘foreign’. Retrieved from https://cran.rapporter.net/web/packages/foreign/foreign.pdf

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted?. *Statistics in medicine*, *21*(11), 1559-1573.

Valcan, D.S., Davis, H. & Pino-Pasternak, D. (2018). Parental Behaviours Predicting Early Childhood Executive Functions: A Meta-Analysis. Educ Psychol Rev, 30, 607–649.

Viechtbauer, W., & Viechtbauer, M. W. (2015). Package ‘metafor’. *The Comprehensive R Archive Network. Package ‘metafor’. Retrieived from http://cran. r-project. org/web/packages/metafor/metafor. pdf*.

West, S. L., Gartlehner, G., Mansfield, A. J., Poole, C., Tant, E., Lenfestey, N., ... & Viswanathan, M. (2010). Comparative effectiveness review methods: Clinical heterogeneity [Internet]. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK53317/table/ch3.t2/

Wickham, H., & Wickham, M. H. (2020). Package ‘plyr’. *Obtenido de https://cran. rproject. org/web/packages/dplyr/dplyr. pdf*.

Yucel, D., & Downey, D. B. (2010). Assessing the advantages of a multi-method approach: Measuring mothering with data from the Early Childhood Longitudinal Study—Birth Cohort. *Social Science Research*, *39*(6), 894-911.

Zeanah, C. H., Gunnar, M. R., McCall, R. B., Kreppner, J. M., & Fox, N. A. (2011). VI. Sensitive periods. *Monographs of the Society for Research in Child Development*, *76*(4), 147-162.

Part 2: Empirical Paper

The Role of Early Environmental Factors in Predicting Adolescent Psychopathology:

A Secondary Data Analysis of NICHD SECCYD Data using Machine Learning

Abstract

Background: Emotional and behavioural problems in adolescence often persist into adulthood. The early identification of those at risk of developing these problems is crucial for successful prevention efforts. A myriad of child, parent and contextual characteristics have been examined in relation to these outcomes, mostly using traditional statistical techniques.

Aims: The aim of the current study is to extend previous research by using a less conventional analytic approach called Machine Learning for the prediction of two response (outcome) variables, namely externalizing and internalizing problems at 15 years of age. The examined features (predictors) encompassed variables representing characteristics of the child and his/her early infancy/childhood caregiving/interpersonal environment as well as various contextual risk factors.

Methods: A secondary analysis of the data (N = 1364) from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD) was conducted. A set of supervised Machine Learning algorithms were used to train several predictive models and select the one with the best performance. The 5-fold cross-validation scheme was used to train and then test the model's performance making sure that different partitions of the data are used for the training and testing phases. Embedded feature selection using Automatic Relevance Determination (ARD) was implemented to identify the features which are most relevant to the prediction of the response variables.

Results: Gaussian Process Regression (exponential GPR) models had the best performance with the lowest RMSE and highest R-Squared values with respect to the prediction of both response variables. Eighteen and 11% of the variance in externalizing and internalizing problems were explained by the obtained models, respectively. The most influential features for the prediction of externalizing problems were 36-month attachment (ambivalent and insecure-controlling/insecure-other or disorganized classifications) and gender. The most influential features for the prediction of internalizing problems were 36-month attachment (ambivalent and avoidance-insecure classifications), non-family childcare hours, and ethnicity.

Conclusions: The current findings are to some extent consistent with previous research. The low amount of variance explained in externalizing and internalizing problems indicate that there are other important predictors that were not included in the models. Moreover, further research is needed to test the models obtained on previously unseen data to be able to determine their clinical utility.

Introduction

Mental health conditions affect a significant proportion of young people and represent 16% of the global burden of disease for adolescents (age 10 to 19). Half of these problems have an onset at the age of 14 and remain mostly undetected (World Health Organization, 2019). The impact of these untreated conditions extends into adulthood, causing physical and mental ill-health (World Health Organization, 2019). Research has shown that behavioural and emotional problems in adolescence persist into adulthood to a considerable degree (Hofstra et al., 2000, 2001, 2002; Reef et al., 2011; Narusyte et al., 2017). Thus, focusing on early interventions to prevent the development of these problems in adolescence is important to improve adult outcomes in children who are at risk of developing mental health difficulties.

Previous research examined early predictors of later externalizing and internalizing problems in an attempt to identify the factors that may need to be addressed in early intervention programs (Kjeldsen et al., 2014). Carneiro et al. (2016) reviewed the literature on the predictors of these behaviour problems during the preschool period. The findings supported the view that, in order to develop an accurate understanding of the emergence and course of these mental health problems in young people, one should approach the child as a part of a system. A variety of environmental factors, including early interpersonal experiences as well as contextual risk factors, were identified as potential predictors of later behaviour problems in addition to the child's individual characteristics.

Early Interpersonal Experiences

The role of early interpersonal experiences in the development of long-term emotional and behavioural problems has been a popular area of research in developmental psychology. There has been great controversy regarding the nature of the impact of these experiences on developmental outcomes, specifically whether it is enduring or transient (Fraley & Roisman, 2015).

Some of the variables that were previously used in research to examine these experiences in relation to long-term socioemotional adjustment include early attachment, early maternal sensitivity, and early child-care.

Early Attachment

Attachment theory provided an influential framework for understanding how early experiences could be operationalised and tested empirically (Bowlby, 1969, 1973, 1980). According to attachment theory, the environment, specifically the quality of the emotional bond between the parent/caregiver and the child early on in development, affects the child's subsequent socio-emotional development and mental health (Newman et al., 2015).

So far, research has shown that the quality of early attachment does predict long-term socioemotional adjustment outcomes. Insecure attachment has been repeatedly found to be predictive of behavioural problems in early and late childhood. Three meta-analyses using observational assessments of attachment examined the associations between insecurity and internalizing as well as externalizing behaviour during childhood and the findings were significant (Fearon et al., 2010; Groh et al., 2012; Madigan et al., 2013). With respect to externalizing behaviour, Fearon et al. (2010) reported a significant small to moderate effect size ($d = 0.31$; 95% CI: 0.23, 0.40) based on 69 independent samples with a total of 5,947 children and their families. With respect to internalizing behaviour, two meta-analyses were conducted by Groh et al. (2012) and Madigan at al. (2013). Groh et al. (2012) reported a small effect size ($d = 0.15$, CI: 0.06, 0.25) based on 42 independent samples with a total of 4,614 children and their families. Madigan et al. (2013)

reported small to moderate effect sizes (*d* = .37, 95% CI: 0.27, 0.46) based on 60 independent samples and a total of 5,236 participants. A meta-analysis based on a more comprehensive search of the literature, including both representational and questionnaire measures of attachment and a wider age group for participants (children aged 3 to 18 years) was conducted by Madigan et al. (2016). Small to moderate effect sizes were reported for the contrasts between secure and insecure attachment in terms of internalizing (165 studies; 48,224 children and their families; *d* = .58; 95% CI: 0.52, 0.64) and externalizing behaviour problems (116 studies; 24,689 children and their families; *d* = .49; 95% CI: 0.42, 0.56).

Prior studies using data from the National Institute of Child Health and Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) examined early attachment in relation to externalizing and internalizing behaviour problems in offspring and found significant associations, whereby offspring with early insecure attachments, were more likely to have increased behaviour problems in early (Belsky & Fearon, 2002a; McCartney et al., 2004) and late childhood (O'Connor, 2012)". Conversely, children with a development advantage, consisting of early secure attachment and increased maternal sensitivity, were more likely to have fewer behaviour problems (Belsky & Fearon, 2002b).

Moreover, given that the nature of early interpersonal experience cannot be fully represented by early attachment, other constructs such early parenting operationalized as early maternal sensitivity as well as early child-care were also examined in relation to long-term socioemotional outcomes (Lorber & Egeland, 2009).

Early Maternal Sensitivity

Maternal sensitivity was first conceptualized by Mary Ainsworth as a mother's ability to perceive and interpret her child's cues accurately and to respond appropriately and in a timely manner (Ainsworth, 1969; Ainsworth and colleagues, 1974). In addition to predicting attachment security – the purpose for which it was originally developed - parental sensitivity has been examined extensively in relation to various developmental outcomes and significant associations have been found with various indicators of socio-emotional development (Deans, 2020).

Lorber and Egeland (2009) conducted one of the earliest studies which examined an early parenting quality construct closely related to maternal sensitivity in relation to behaviour problems longitudinally. Significant associations were found between low quality parenting during infancy (i.e., defined as low sensitivity, low positive regard and high negative regard) and externalizing behaviour at follow-up (i.e., kindergarten, first grade as well as ages 23 and 26) throughout the period from kindergarten to early adulthood.

Structural equation modelling was used to examine the nature of the relationship between maternal sensitive parenting and internalizing behaviour problems during the preschool period using data from the Generation R study (N = 886), a large longitudinal population-based study. Model pathways, which featured multiple measurements of these two variables across time, revealed a consistent association between both constructs. These findings were replicated when the same statistical analysis methodology was implemented using data from the NICHD SECCYD (N = 935) collected during the preschool period lending further support to their validity (Kok et al., 2013). Structural equation modeling was performed again on data from the NICHD SECCYD (N = 1,306) to examine maternal sensitivity early in life (first three years) in relation to not only internalizing but also externalizing problems, during a period extending from preschool years to adolescence (age 15). The modeling included repeated assessments of both variables over time. The results supported an enduring effects model suggesting that early sensitive parenting may have a long-term effect on later internalizing as well as externalizing problems (Haltigan at al., 2013).

Early Child-Care

With respect to early child-care, different related aspects have been found to be associated with adolescent socioemotional adjustment outcomes (Burchinal et al., 2014; Vandell et al., 2016). A greater number of hours spent in child-care was shown to predict more impulsivity and externalizing behaviour problems at the age of 15 for those who experienced low maternal sensitivity during middle childhood (Burchinal et al, 2014).

Contextual Risk

In addition to early experiences with caregivers, other variables related to the parent/primary caregiver as well as the sociocultural context in which the relationship between the parent/primary caregiver and the child develops may also increase the risk of behaviour problems. Research has shown repeatedly that early contextual risk factors are significantly associated with long-term developmental outcomes, possibly due to their influence on the quality of early parenting (Belsky & Fearon, 2002b; McLoyd, 1998).

Contextual characteristics, including child's minority status, poverty, younger maternal age, poorer maternal education, single parenthood, maternal mental health difficulties, parenting stress, low marital relationship quality, and low mother-reported social support, were found to be associated with unfavourable socioemotional development outcomes in offspring (Ashford et al., 2008; Bayer et al., 2012; Carneiro et al. 2016; Connell & Goodman, 2002; Heberle et al., 2005; Lansford et al., 2019; Leve et al. 2005; McCarty & McMahon, 2003; McLaughlin et al., 2007; Morgan et al., 2009; Palmer et al., 2018; Vaez et al., 2015; Wang et al., 2018).

In prior studies using data from the NICHD SECCYD (Belsky & Fearon, 2002b; Haltigan et al., 2013), the importance of partialing out the contribution of contextual risk factors when examining the relationship between early experiences with primary caregivers and later socio-emotional outcomes was addressed. Belsky and Fearon (2002b) showed that cumulative contextual risk had a significant moderating effect on the relationship between attachment and socio-emotional outcomes, whereby attachment insecurity was most predictive of later behaviour problems in the presence of contextual risk. Thus, it is important to isolate the effects of these risk factors from the effect of early interpersonal experiences on subsequent development and adjustment.

The Child's Characteristics

Central to developmental psychology is the nature versus nurture debate, which questions to what extent developmental outcomes are caused by biological versus environmental factors. Thus, in an attempt to get a better understanding of the role of environment in later development and adjustment, the impact of early interpersonal experiences as well as other contextual factors needs to be examined while taking into consideration the effect of the child's predispositional characteristics.

Based on a systematic review of the literature, temperament was identified as one of the characteristics, potentially predictive of socioemotional development during the preschool years (Carneiro et al. 2016). According to the diathesis-stress model, psychopathology emerges as a result of an interaction between biological vulnerabilities and stress caused by life experiences (Broerman, 2020). In line with this model, the impact of the environment on later behaviour problems was found to be moderated by temperament. In a cohort study following-up a community-based sample of children and their families (N = 373) from childhood to adolescence, the findings of latent growth curve modelling showed that temperament was predictive of age 17 externalizing and internalizing problem (Leve et al., 2005). In a similar vein, a study using data from the NICHD SECCYD showed that children who had a difficult temperament in infancy were more likely to show compromised socio-emotional adjustment later on (at 54 months and ages 11 and 15) if they had low-quality childcare (Pluess & Belsky, 2009). Thus, considering the findings of previous research, it is worth examining early temperament in addition to other early environmental factors (i.e., early attachment, maternal sensitivity, and early non-family child-care hours as well as early contextual risk factors) in relation to long-term internalizing and externalizing psychopathology.

Finally, sex differences have been reported in the literature (Bayer et al., 2012; Carneiro et al., 2016; Leve et al., 2005; Morgan et al., 2009; Wang et al., 2018), showing a tendency toward girls being at a greater risk of internalizing problems and boys being at a greater risk of externalizing problems. A systematic review by Carneiro et al. (2016) reported that, although these sex differences were not substantial, such findings were more or less consistent.

Machine Learning

The current study was a secondary analysis of data collected in the NICHD SECCYD, with the primary aim of examining the predictive power of the variables described above (including early interpersonal experiences, early contextual risk factors, and child's characteristics) in anticipating the emergence of behaviour problems later in life. Prior studies using data from the NICHD SECCYD have examined these associations using traditional statistical methodologies. However, the current study extended previous work by using Machine Learning predictive modeling instead of more traditional statistical techniques. Specifically, supervised Machine Learning was used in the current analysis. When supervised Machine Learning is used, a model is trained on available input and output data. During the training phase, the model learns from the data i.e. making predictions and evaluating its performance by comparing the predicted responses to the actual true responses iteratively. This learning process ceases when the model reaches an acceptable level of performance, as indicated by a minimum prediction error when the predicted responses are compared to the correct responses (Ciaburro, 2017; MathWorks, 2016b).

Early detection of children at risk of developing emotional and behaviour problems in adolescence is important for designing early intervention and prevention programs. The risk factors with the most predictive power can be then identified through Machine Learning predictive modeling. This analytic approach has been recently used in psychological research and was found generally to be a suitable technique for predicting oppositional defiant behaviour during the first five years of life (Na et al., 2020), mental health difficulties (Tate et al., 2020), self-harm and suicidal behaviour (Burke et al., 2019), and violent offending in patients with schizophrenia (Kirchebner et al., 2020). A Machine Learning approach can be superior to traditional statistical techniques for a variety of reasons. With traditional statistical methods, linear models are generally employed to make sense of the data, whereas Machine Learning-based predictive modeling can work with more complicated non-linear relationships between variables. Moreover, unlike standard regression, Machine Learning predictive modeling can examine a wide range of predictors simultaneously (Orrù et al., 2020). Thus, the current study will allow for the examination of a large number of potential early predictors of long-term internalizing and externalizing psychopathology. A systematic review was conducted by Burke at al. (2019) to examine the value of using a Machine Learning to predict self-harm and suicidal behaviour. The review showed that analytic methods using Machine Learning not only confirmed but also augmented previous findings in the literature about related risk factors and were associated with enhanced predictive accuracy compared to traditional statistical techniques.

The primary aim of the current study is to use Machine Learning predictive modeling to examine early interpersonal experiences, contextual risk factors, and child characteristics in relation to two outcome variables, mother-reported internalizing and externalizing problems at age 15 years. Specifically, the current study aimed to build a model that can be used to predict behaviour problems from new input data in the future. The secondary aim of the study was to identify the predictors with the greatest predictive power, which contributed the most to the variance explained in the outcome variables.

Methods

Participants

This study used data from the NICHD SECCYD, a multi-site, multi-phase, and longitudinal study conducted from 1991 to 2007. It primarily aimed to examine the relationships between child-care characteristics and later cognitive, physical, and socio-emotional developmental outcomes. The study consisted of four phases including the following: (a) Phase I from birth through three years of age, (b) Phase II from 54 months through first grade, (c) Phase III from second through sixth grades, and (d) Phase IV from seventh through ninth grades. The data used in the current analyses were collected during the first and last phases of the study.

The recruitment process was initiated in January 1991 and was finalized in November 1991. Throughout 1991, 8,986 mothers, who were giving birth in hospitals across 10 different locations in the United States (Little Rock, AK; Irvine, CA; Lawrence, KA; Boston, MA; Philadelphia, PA; Pittsburgh, PA; Charlottesville, VA; Morganton, NC; Seattle, WA; and Madison, WI), were screened for eligibility. Initial exclusion criteria consisted of the following: (a) the mother was below the age of 18, (b) the mother was not an English speaker, (c) the family had plans to relocate within the first year, (d) the new-born had serious medical complications or was born from a mother who used substances during pregnancy, (d) the mother was seriously ill, (e) the new-born was placed for adoption, (f) the mother refused to participate in further screening, (g) the mother was living more than an hour away from the collection site, or (h) the mother was residing in an unsafe neighbourhood.

As a result, 5416 families and their new-borns were recruited as potential participants for further follow-up and screening done via 2-week phones calls. Conditional random sampling was used to select those to be enrolled in the study from this sample of potential participants. This sampling procedure aimed to ensure that that the sample was diverse in terms of important sociodemographic characteristics (including income, education, marital status, and race) and child-care arrangement plans (full-time maternal care, full-time non-maternal care, or part-time maternal care). As a result, a random sample of 3,015 children and their families was selected for further participation in the enrolment process. At this stage, additional families were excluded if they met any of the following criteria: (a) the infant was hospitalized for more than seven days, (b) the family had plans to relocate within the next three years, (c) three unsuccessful attempts at contacting the family by phone, or (d) the family refused to participate. As a result, a total of 1,526 families were considered eligible, of which 1364 (89 % response rate) officially enrolled by giving their consent to participate in the study and successfully completing the one-month interview (NICHD Early Child Care Research Network [ECCRN], 2005a).

Table 1 provides descriptive statistics for the enrolment (N = 1364), the analysis (n = 973), and the attrition samples (n = 391). A total of 1,364 healthy children and their families enrolled in the study following the successful completion of the 1-month interview. The enrolment sample comprised 52% male and 48% female new-borns. Most of the sample consisted of White American children (80 %), with the remaining participants coming from non-White American ethnic backgrounds (12.9% Black or Afro-American; 1.6% Asian or Pacific Islander; 0.4% American Indian, Eskimo, Aleut). The mothers, who participated in the study, were on average 28 years of age (SD = 5.63) and had 14 years of education (SD = 2.51). The mean income-to-need ratio (M = 3.34, SD = 2.69) indicated that the sample was on average above the poverty line.

The study's sampling procedures aimed to include families from different locations in the United States, who varied in terms of socioeconomic status, ethnic background, child-care arrangements, and maternal employment plans. Based on the U.S. Census Tract data, the sample of participants recruited for this study was representative of the children and their families who lived in the areas from which recruitment occurred with respect to most demographic characteristics, with the exception of education and income. Parents' education level was higher and household income was lower than those reported based on the U.S. Census Tract data (NICHD ECCRN, 2005a; Watamura, Phillips, Morrissey, McCartney, & Bub, 2011).

At age 15, scores on the two outcome variables used in the current study (mother's report of internalizing problems; mother's report of externalizing problems) were obtained for 973 adolescents (71 % of the enrolment sample). Table 1 also compares the age 15 analysis and attrition samples with respect to the variables included in the current study (refer to Appendices D and E for additional details). Differences were found between the analysis sample at age 15 (n = 973) and the attrition sample (n = 391). The adolescents who did not participate in the study at age 15 were more likely to be male and had on average a lower attachment Q-sort security score at 24-month follow up. On average, the nonparticipating mothers at age 15 were younger, had less years of education, and reported a lower income-to-need ratio. They also scored lower on early maternal sensitivity (first 36 months), which is consistent with the finding on their offspring scoring lower on Attachment Q-sort Security. Interestingly, they seemed to have had greater support from their partners/husbands, reporting on average higher scores on relationship intimacy and a lower frequency of early single parenthood status (during the first 36 months)

Procedure

Children (and their families) were followed from birth to the age of 15. The data used in the current study was collected at 1, 6, 15, 24, and 36 months as well as age 15. Data collection occurred at multiple time points: at 1, 6, 15, 24, 36 month follow-up in phase one; at 54 months, K, and grades 1 in phase two; at grades 2, 3, 4, 5, and 6 in phase three, and at grade 7 and age 15 in phase four. Assessment were conducted in a variety of settings (home, lab, childcare, and/or school visits), with different informants (children, peers, parents, caregivers, and/or teachers), and using different measurement methodologies (interviews, self-report measures, observations).

Table 1

Descriptive Statistics

	Enrolment sample (N = 1364)			Analysis sample (n = 973)			Attrition sample (n = 391)			p value
	N	M or %	SD	N	M or %	SD	N	M or %	SD	
Sex	1364			973			391			0.04
Male	705	51.70		486	49.90		219	56.00		
Female	659	48.30		487	50.10		172	44.00		
Child's Ethnicity	1364			973			391			0.08
American Indian, Eskimo, Aleut	5	0.40		2	0.20		3	0.80		
Asian or Pacific Islander	22	1.60		15	1.50		7	1.80		
Black or Afro-American	176	12.90		114	11.70		62	15.90		
White	1097	80.40		794	81.60		303	77.50		
Other	64	4.70		48	4.90		16	4.10		
Mother's Age (years)	1364	28.11	5.63	973	28.58	5.57	391	26.94	5.63	<.0011
Mother's Education (years)	1363	14.23	2.51	973	14.45	2.45	390	13.69	2.58	<.001
Income-to-Need Ratio	1355	3.34	2.69	971	3.52	2.62	384	2.90	2.81	<.001
Maternal Characteristics										
Single Parenthood	1364	3.20	1.39	973	3.41	1.26	391	2.69	1.57	<.001
Depression	1363	9.86	6.88	973	9.72	6.58	390	10.19	7.55	0.28
Psychological Adjustment	1272	59.00	13.95	946	59.02	13.68	326	58.96	14.73	0.95
Parenting Stress	1363	0.00	0.80	973	0.03	0.76	390	-0.06	0.86	0.07
Social Support	1363	4.98	0.61	973	4.98	0.58	390	4.99	0.70	0.69
Marital/Partner Intimacy	1288	4.85	0.97	931	4.77	0.93	357	5.05	1.03	<.001

Maternal Sensitivity	1306	-0.02	0.77	967	0.03	0.75	339	-0.15	0.81	<.001
Child Characteristics										
15-month Attachment (Strange Situation)	1191	87.30		921	94.70		270	69.10		0.60
Insecure - Avoidant (A)	160	11.70		128	13.20		32	8.20		
Secure (B)	710	52.10		552	56.70		158	40.40		
Insecure - Resistant (C)	102	7.50		75	7.70		27	6.90		
Disorganized/ Disoriented (D & U)	219	16.10		166	17.10		53	13.60		
36-month Attachment (Strange Situation)	1140	83.60		901	92.60		239	61.10		0.33
Avoidant (A)	55	4.00		40	4.10		15	3.80		
Secure (B)	701	51.40		562	57.80		139	35.50		
Ambivalent (C)	197	14.40		149	15.30		48	12.30		
Insecure Other/ Controlling (D)	187	13.70		150	15.40		37	9.50		
24-mo Attachment Q-sort Security	1197	0.29	0.21	927	0.30	0.21	270	0.27	0.21	0.03
Temperament (mother's report)	1279	3.18	0.40	952	3.17	0.41	327	3.21	0.40	0.11
Early Non-family Child Care										
Total Child-Care Hours (per week)	1035	28.35	14.66	781	27.96	14.84	254	29.54	14.04	0.12
CBCL internalizing				973	46.64	9.86				
CBCL externalizing				973	45.51	10.46				

Regular phone interviews with mothers (and children at later ages) were also carried out to update and add to the data already collected. Children (and their families) were followed from birth to the age of 15. The data used in the current study was collected throughout first three years of life (at 1-, 6-, 15-, 24-, and 36-months during phase I) and at the age of 15 (during phase IV). Data collection took place during home visits for almost all the variables included in the analysis, with the exception of the assessments for attachment at 15- and 36-month follow-up, which occurred during lab visits. Data was collected using interviews, self-report questionnaires, and observations. An emphasis was placed on training research assistants as well as the use of standardized assessments and reliability testing to enhance the quality of the data collected.

Measures

Contextual Risk

Sociodemographic Data. Data on the following sociodemographic variables was collected at one-month follow-up: child's ethnicity; mother's age; mother's education; the poverty income-to-need ratio; and single parenthood status. The child ethnicity variable originally consisted of five categories (American Indian, Eskimo, Aleut; Asian or Pacific Islander; Black or Afro-American; White; and Other). The categories other than 'White' were combined into one category referred to as 'Other', resulting in a variable with two levels only. Mother's education level represented number of years of education.

The income-to-need ratio was used for assessing poverty level. It examines total family income in relation to the poverty threshold for a household. Total family income includes mother's income, other sources of income (including government funds), and husband/partner's income if he lives at the home. The poverty threshold for a household is based on the year the income is earned, the total number of household members, and the number of children living within the household full-time. Data on the income-to-need ratio was collected at five time points (at 1, 6, 15, 24, and 36 months) throughout the first three years of life. The average income-to-need ratio over this period was used for the current analysis.

Mother's single parenthood status is the number of times from 1 to 36 months, when the mother reported not living with the father or the partner at follow-up. Specifically, it represents the frequency of reporting being a single parent across all time points (1, 6, 15, 24, and 36 months).

Maternal Characteristics

Several maternal characteristics were also examined as potential (predictors) within this study. The variables that are related to the mother's psychological well-being included the following: Maternal Depression; Maternal Psychological Adjustment; and Parenting Stress.

Maternal Depression was examined using the Center for Epidemiological Studies Depression Scale (CES-D). The CES-D is a 20-item self-report measure aimed at assessing depressive symptoms in the general population. Participants are required to report how frequently they experienced 20 depressive symptoms during the past week. Research has shown that the scale has satisfactory psychometric properties across a variety of samples and settings. Satisfactory internal consistencies and test-retest reliabilities were reported as well as evidence for structural, convergent and criterion validity (Radloff, 1977; Roberts, 1980; Roberts & Vernon, 1983; Orme, Reis, & Herz, 1986). Standard psychometric analyses using the current study sample showed that the scale was highly reliable, with Cronbach alphas at follow-up ranging between 0.88 to 0.91. The scores for the five time points (1, 6, 15, 24, and 36 months) were aggregated to yield an overall average estimate of maternal depression to be used for the current analysis.

Maternal Psychological Adjustment was a composite score computed using data on three personality traits (agreeableness, extraversion, and neuroticism) collected at 6-months follow-up. The composite score was obtained by subtracting neuroticism from the sum of agreeableness and extraversion. The assessment was conducted at home using the Neuroticism and Extraversion scales of the NEO Personality Inventory (NEO PI) and Agreeableness scale of the NEO Five-Factor Inventory (NEO-FFI). The NEO PI is the gold-standard measure for assessing personality based on the Five Factor Model (FFM), which posits that personality can be best explained in terms of five major factors (i.e., Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness). The psychometric properties of the NEO PI have been extensively examined and satisfactory empirical evidence is available supporting the use of this instrument worldwide (Costa & McCrae, 1985; Costa & McCrae, 2008). The NEO FFI is briefer personality assessment measure derived from the NEO PI (Costa & McCrae, 1989). Acceptable to excellent internal consistencies and test-retest reliabilities were reported for both measures. Moreover, a wealth of research has lent support to the structural validity of the NEO PI across various cultural contexts (Costa & McCrae, 2008). The NEO PI neuroticism subscale assesses emotional instability. The NEO PI extraversion subscale assesses the degree to which an individual is sociable, active, optimistic, fun-loving, and affectionate. The NEO FFI agreeableness subscale assesses the extent to which an individual is soft-hearted, trustworthy, humble, complaisant, and altruistic. An overall composite score representing psychological adjustment was derived by summing the scores on Agreeableness and Extraversion followed by subtracting the score on Neuroticism.

Maternal Parenting stress was assessed using two different measures. The Parenting Stress Index (PSI) (Feelings about Parenting) was administered at one- and six-month follow-up during home visits. Specifically, a shorter (30-item) version of the 101-item PSI was used and comprised three Parent subscales: Attachment, Restrictions of Role, and Sense of Competence. The Attachment subscale measures the degree to which the mother feels invested in her parenting role. The Restrictions of Role subscale assesses the extent to which the mother feels restricted due to her parenting responsibilities. The Sense of Competence subscale taps into the mother's sense of competence with respect to her parenting skills (Abidin, 1983; Abidin, 2012). Acceptable to excellent internal consistencies were reported for the Parent subscales (0.75-0.87) and Total Parenting Stress scale (0.96 or greater). Adequate empirical evidence was reported supporting the use of this measure across a variety of samples (American Psychological Association, 2011). In a study by Belsky and Fearon (2002a) using data from the NICHD Study for Early Child Care, high internal consistencies were reported for the Total Parenting Stress scale score (Cronbach alphas > 0.65) at each follow-up assessment.

The second measure that was used as an assessment tool for maternal parenting stress was the Parent Role Quality Scale (PRQS) (Parenting Experiences). This measure is aimed at evaluating the quality of parenting and is suitable for use with parents of toddlers, pre-schoolers and older children. It was administered at 15-, 24-, and 36-month follow-up during home visits. A modified version of the Parent-Role Quality scale was used in the current study, including 10 negative (concerning) items and 10 positive (reinforcing) items. Participating mothers were required to rate on a 4-point scale the degree to which they perceived each negative item as a concern, and each positive item as a rewarding experience (Barnett & Marshall, 1991). Previous research reported good internal consistencies and test-retest reliabilities as well as some evidence supporting the scales criterion validity (Barnett & Marshall, 1991; Barnett, Marshall, & Pleck, 1992).

Scores on the PSI at one- and six-month follow-up and those on the PRQS at 15-, 24-, and 36-month follow-up were standardized and then averaged to yield an overall estimate of early parenting stress for the first three years of life.

The six-item emotional intimacy subscale of the Personal Assessment of Intimacy in Relationships Inventory (PAIR) (Schaefer & Olson, 1981) was used to assess the quality of the mother's relationship with the partner/husband. The subscale was administered during home visits at one- and 36-month follow-up. Using data from the NICHD SECCYD, high internal consistencies were reported for measurements at one- (Cronbach alpha = 0.80) and 36-month (Cronbach alpha = 0.86) follow-up (Belsky & Fearon, 2002a).

Social support was assessed by using the Relationships with Other People instrument. Measurements were carried out at all five time points during home visits 1, 6, 15, 24, & 36 months. It is an 11-item self-report measure, which requires participating mothers to rate their relationships in the past month (Marshall & Barnett, 1993). Good internal consistencies and test-retest reliabilities were previously reported. In the sample used by the NICHD Study of Early Child Care, excellent Cronbach alphas were reported exceeding 0.9 (Belsky & Fearon, 2002a).

Early Interpersonal Experiences

Maternal sensitivity was assessed using an observational measure. Videotapes of 15-minute semi-structured interactions between the mother and the child were assessed for maternal sensitivity. For half of the time, the mother and the child were required to play freely as they usually do. For the second half, the mother was asked to engage her child in play using a standard set of toys.

Qualities of mother-child interaction were rated from the videotapes obtained at 6 and 15 months in the home setting and those obtained at 24 and 36 months in the lab setting.-Four-point global qualitative rating scales were used for videotapes obtained at the first three measurements time points (6, 15, and 24 months) and were derived. These scales were used to rate the following qualities of mother-interactions with the child: the mother's sensitivity/responsiveness to distress, sensitivity/responsiveness to nondistress, intrusiveness, detachment/disengagement, stimulation of cognitive development, positive regard for the child, negative regard for the child, and flatness of affect.

Seven-point global qualitative rating scales were used to rate mother-child interactions in videotapes at 36-month follow-up. These scales were based on the work of Egeland and Heister (1993). Ratings for the flowing qualities were provided: maternal supportive presence, respect for child’s autonomy, stimulation of cognitive development, hostility, and confidence, and child enthusiasm, negativity, persistence, and affection for mother. Check NICHD SECCYD (2001) and Vandell et al. (2010) for additional details about the coding procedures.

At 6, 15, and 24 months, the maternal sensitivity total scores were computed by summing 3 ratings on the following subscales: sensitivity to nondistress, positive regard, and intrusiveness (reversed). At 36 months, the maternal sensitivity total score was constructed by summing ratings on the following subscales: supportive presence, respect for autonomy, and hostility (reversed) scales. Maternal sensitivity composites at 6, 15, 24, and 36 month had satisfactory internal consistencies and inter-coder reliabilities. Maternal sensitivity scores obtained at 6, 15, 24, and 36 months were standardized, and a mean score was computed to represent observed early maternal sensitivity.

Infant-mother attachment assessments at both 15 months and 36 months were used in the current analysis. Attachment at 36 months will be examined given the evidence showing that later attachment may be more predictive of behaviour problems than earlier ones (Madigan et al., 2013).

The Strange Situation (Ainsworth et al., 1978) was conducted in the 15-month laboratory visit. See NICHD Early Child Care Research Network (1997) for a full description of the procedures at 15 months.

At 15 months, the study employed Ainsworth and Wittig's (1969) standard Strange Situation procedure to assess attachment. Mother-child interactions were videotaped in a laboratory playroom setting in a series of three-minute episodes, aimed at activating the child's attachment system. The first episode consisted of the mother and the child getting familiar with the playroom while a 'stranger' joined them, followed by the mother leaving for three minutes (first separation episode) then returning for another three minutes (first reunion episode). After the first reunion episode with the mother, there was a second separation episode, in which the child is left alone, without the mother and the stranger in the room for three minutes. After the second separation episode, there was a second reunion, with the stranger first, followed by the mother. The duration of the separation episodes had to be cut short by the return of the mother, when the stranger wasn't successful at comforting the child. Ainsworth and colleagues' (1978) coding system was used to classify the children as either secure (B), insecure-avoidant (A), insecure-resistant (C), disorganized (D), and unclassifiable (U). Secure (B) infants seek contact with the mother at reunion, to get comfort in case the separation was distressing or to satisfy their desire to communicate/ connect with her. Infants with insecure-avoidant (A) attachment generally remain indifferent to the separation and do not seek closeness with the mother upon reunion. Infants with insecure-resistant attachment are distressed by the separation, and although they seek proximity upon reunion, they also display angry behaviour resisting closeness. The behaviour of infants with disorganised/ disoriented (D) attachment is confusing and difficult to explain. Infants with this type of attachment display simultaneously or sequentially contradictory behaviours, such as seeking closeness and yet fighting it or running away from it. Such contradictory behavioural patterns can only be justified if one assumes that the child is having to maintain a bond with a threatening/unsafe parent to be able to survive. Behavioural patterns that fall under category (D) primarily has the characteristics of disorganized/ disoriented attachment but also can have a secondary classification falling within either categories (A), (B), or (C). If the child's behaviour is not classifiable within the latter conventional categories, then it is classified as Unclassifiable (U). Behaviour categorized as Unclassifiable (U) is almost always disorganized/ disoriented, thus this category was combined with category (D) to yield an overall category referred to as primary Disorganized/ Disoriented (D + U) attachment.

At 36 months, a modified Strange Situation procedure was employed (Cassidy et al., 1992b). The modification excluded the stranger reunion following the second separation episode. Moreover, the second separation episode was also modified to last longer (five than 3 minutes) than in the original strange situation procedure.

The criteria set by MacArthur Working Group on attachment were used to classify the child's attachment as either Secure (B) or Insecure (A, C, or D). The insecure cluster includes Avoidant (A), Ambivalent (C), or Insecure-controlling/ Insecure-other subclassifications (Cassidy et al., 1992a). The definitions of the attachment categories (A), (B), and (C) are similar to the conceptualisations derived from the 15-mo Strange Situation assessment procedure. Insecure-controlling/insecure-other subclassifications (D) children are either controlling (e.g., reversing roles or acting punitive with the mother) or show combinations of attachment strategies (e.g., avoidant and ambivalent or avoidant and controlling attachment behavioural patterns). The Attachment Q-Set was used to assess attachment security at 24-month follow up (Waters and Deane, 1985). It consists of 90 cards and each card represents a behavioural characteristic of a child from 18 to 24 months. Observers have to sort these cards into different sets, which range from "most descriptive" to "least descriptive" of the child. Scores of nine and one are assigned for items classified as "most characteristic" and "least characteristic", respectively. The procedure ultimately yields three scores for three attachment related concepts (security, dependency, and sociability) and a social desirability score to control for this confounding variable. In the NICHD SECCYD, everyday routine as well as semi-structured mother-child interactions were observed during a two-hour home visit by one observer (with the exception of visits which were aimed at examining reliability whereby two observers were present) at 24-month follow-up. The cards were sorted immediately following the 2-hour home visit based on memory and notes taken throughout the observation. A meta-analysis by van Ijzendoorn et al. (2004) reported evidence supporting the validity of the Attachment Q-Sort Security score as an attachment measure. In the NICHD SECCYD, the

Pearson correlation coefficient between the child's sort and the "Security Criterion Sort" was used as the security score at 24-month follow-up. High inter-observer reliability was reported for this measure (r = 0.92).

Non-family-child care hours was also examined as a variable representing one important aspect of early interpersonal experiences or early experiences with caregivers. It represented the total numbers per week spent in child-care that did not involve any member of the immediate or extended family. Mother-reported child-care total number of hours per week were collected at 3-month intervals (or epochs) throughout the first three years of life to represent the typical total number of hours spent in non-family child-care per week for the past three months or epoch. The average total number of hours per week was computed from the weekly number of hours obtained at three-month intervals throughout the first 36 months of life.

Child's Characteristics

Two child characteristics have been examined as predictors, early temperament and sex. The Revised Infant Temperament Questionnaire (RITQ) (Carey and McDevitt, 1978) was used to assess temperament at six-month follow-up. Mother were required to complete five subscales, namely Approach, Activity, Intensity, Mood, and Adaptability. The total battery composite was used in the current analysis. Satisfactory internal consistencies were reported in previous analyses using data from the NICHD Study for Early Child Care for the battery's total (Cronbach alpha = 0.81) as well as subscale scores (Cronbach alphas ranging from 0.52 to 0.75).

Adolescent Mental Health

Two outcome variables were examined, reflecting the two domains of behaviour problems, namely mother-reported internalizing and externalizing problems.

The Child Behaviour Checklist completed by the mother or alternate caregiver was used as an indicator of socioemotional development at the age of 15 (CBCL, Achenbach, 1991). The CBCL is recognized worldwide as the gold-standard for behaviour problems and social competence in young people. The CBCL consists of items representing a broad range of emotional, behavioural, and social problems and respondents have to rate the extent to which each item is representative of the child behaviour using a three-point scale. Specifically, standardized T scores on two Total scales were using in the current study, namely the Internalizing and the Externalizing scales. The Internalizing scale consisted the Withdrawn, Somatic Complaints, and Anxious/Depressed syndrome subscales. The Externalizing scale comprised the Delinquent Behaviour and Aggressive Behaviour syndrome subscales. Higher scores indicate greater behavioural and emotional problems. High internal reliabilities were found for the externalizing (Cronbach's alpha = 0.91) and internalizing (Cronbach's alpha = 0.86). Total scales in previous standard psychometric analyses using the data from the NICHD SECCYD.

Data Analysis

Machine learning was used to examine the overall contribution of a set of 17 predictors to the explained variance in two response variables. The predictors (referred to as features in the Machine Learning literature) included four categorical variables (Child's Sex; Child's Ethnicity; 15-mo Attachment; and 36-mo

Attachment) and 13 continuous variables (Mother's Age; Mother's Education; Income-to-Need Ratio; Mother's Single Parenthood Status; Maternal Depression; Marital/Partner Relationship Intimacy, mother-report; Mother's Social Support; Mother's Psychological Adjustment; Mother's Parenting Stress; Maternal Sensitivity; Non-Family Child Care Hours; Child Temperament, mother-report; and 24-mo Attachment Q-Sort Security). The two outcome variables (referred to as response variables in the Machine Learning literature) were Internalizing Problems and Externalizing Problems (mother/alternate caregiver's report).

Specifically, supervised Machine Learning was used in the current analysis. This technique consists of training models on available input (predictors or features) and output (outcome or response variables) data, which could potentially be used in the future for predicting outcomes based on new input data. Unlike supervised Machine Learning, which uses both input and output data, unsupervised learning uses input data solely, with the aim of identifying underlying patterns within it (Bali et al., 2016; Ciaburro, 2017; MathWorks, 2016a, 2016b;).

Given that the response (outcome) variables in the current study were continuous, the following supervised Machine Learning algorithms for regression were used for model building: Linear Regression, Regression Trees, Support Vector Machines (SVM), Gaussian Process Regression (GPR), and Ensemble Trees. These are the algorithms available in the Regression Learner app of the Statistics and Machine Learning toolbox in MATLAB and some of the most common algorithms used for predictive modeling of continuous outcomes (Ciaburro, 2017; MathWorks, 2016a, 2016b). Table 2 presents brief descriptions of these algorithms, including their strengths, weaknesses and common uses.

Table 2

Descriptions of Supervised Machine Learning Algorithms

Type of Algorithm	How It Works	Strengths	Weaknesses	Best Used
Multiple Linear Regression	Describes a continuous response variable as a linear function of one or more features[a]	Easy to interpret[b]	Very low to medium flexibility[b]	For the ease of fitting and interpretation[a]
	To identify the beta coefficients that are associated with the least prediction error[c]	Easy to train and fit[c]	Based on strong assumptions about the data[c]	As a baseline against which other more complex regression models are compared[a]
		Provides estimates clarifying the nature of the association between the features and the response variables[c]	The model's form must be prespecified[c]	
Regression Trees	Decision trees for regression are similar to decision trees for classification, but they are modified to be able to predict continuous responses[a]	Easy to interpret[b]	Difficult to interpret with large trees[c]	With categorical features or non-linearity[a]
		Medium (medium tree) to high flexibility (fine tree)[b]	Low flexibility for coarse tree[b]	
		No requirement for model pre-specification[c]	A large amount of training data required[c]	

		May fit some type of data better than regression[c]	Difficult to determine the predictive effect of individual features[c]	
Support Vector Machines (SVM)	SVM regression algorithms find a model that is associated with the least amount of prediction error while at the same time using parameters that reduce sensitivity to error[a]	Easy to interpret for linear SVMs[b]	Hard to interpret for kernels other than the linear[b]	For high-dimensional data[a]
		Medium (Quadratic, Cubic or Medium Gaussian kernels) to high (Fine Gaussian kernel) flexibility[b]	Low flexibility for Linear and Coarse Gaussian kernels[b]	
		Less likely to be influenced by noisy data[c]	Slow to train with large datasets[c]	
		Less risk of overfitting[c]		
Ensembles	Ensemble-based methods involves training several strong models and combining them into all-purpose predictive model[c]	Medium to high flexibility[b]	Hard to interpret[b]	For high-dimensional data as well[a]
		Better generalizability[c]	Model tuning can be time consuming[c]	Can handle small datasets[c]
		Suitable for noisy or missing data as well as categorical or continuous features[c]		
Gaussian Process Regression	Nonparametric[a]	Automatic flexibility: optimizing accuracy while reducing risk of overfitting[b]	Hard to interpret[b]	Widely used in the field of spatial analysis[a]

Note. [a]MathWorks (2016b); [b]MathWorks (2020g); [c]Bali et al. (2016)

The plan of the analysis consisted of four main steps (a) data pre-processing, (b) training regression models and identifying the one with the best performance, (c) identifying the most predictive features, and (d) improving the model by tuning its parameters.

Pre-processing

The statistical packages SPSS (version 26) and R (version 1.2.5033) were used for pre-processing the data. The pre-processing of the data consisted of the following steps: exploring assumptions; examining descriptive data; detection of univariate outliers; missing values analysis (MVA); and imputation of missing values. SPSS was used to perform all of these steps with the exception of imputing missing values, which was completed by using the VIM R package (Field, 2009; Kowarik & Templ, 2016).

Exploring violations of normality were examined by quantifying normality using standardized scores of skewness and kurtosis. Descriptive data included measures of central tendency (incl. mean, median and mode) and variability (incl. standard deviation and range). Manifest univariate outliers were detected using boxplots (Field, 2009; Rumsey, 2016).

SPSS MVA was performed to test for MCAR (missing completely at random) using the Little MCAR's test (Tabachnick & Fidell, 2013). The VIM R package was used to perform data imputation using the k-nearest neighbor algorithm. k-nearest neighbor imputation is a single imputation technique, whereby each missing value is replaced by one plausible value estimated based on values from closely matching other cases (nearest neighbors), resulting in a single dataset that is used for further analysis (Jadhav et al., 2019). k is the number of neighbors used for the imputation and is a tuneable hyperparameter. By default, it is set at five. A neighbor is a case which is similar to the one with a missing value based on two criteria. First, they are similar based on the available values on other features included in the dataset/the pattern of values obtained on other features. Second, similarity is determined by using a distance function. Distance functions are used to calculate the distance between two observations (or cases) represented by some feature vectors, as two points in a multidimensional feature space (Cordeiro et al., 2010; Jadhav et al., 2019). Gower's distance is the recommended distance function to use for determining nearest neighbors when the variables used for distance calculations (referred as distance variables) are a mix of qualitative and quantitative features (Kowarik & Templ, 2016).

After identifying the k nearest neighbors, the missing value on a particular feature for an observation (or a case) will be imputed based on the corresponding k values obtained from the k nearest neighbors, which do not have a missing value on that particular feature. These k values have to be aggregated to produce a single imputed value. If the missing value is on a continuous variable, then the default aggregation method is to use the median of the k values of the nearest neighbors. If the missing value is on a categorical variable, the default method would be to use the mode of the k values obtained (Kowarik & Templ, 2016).

Machine Learning Predictive Modeling

The Regression Learner app of the Statistics and Machine Learning Toolbox in MATLAB R2020a (9.8.0) was used to perform Machine Learning predictive modeling which consisted of the following steps: training several regression models and identifying the model with the best performance; determining the most predictive features (predictors); and improving the model by tuning its hyperparameters (Ciaburro, 2017; MathWorks, 2016c).

Before selecting the model with the best performance, several models have to be trained using a dataset with input and output data, referred to as training data. The performance of the resulting model is checked by testing it using a new dataset. Testing each model using validation in this way protects against overfitting, which occurs when the model excessively fits the training data to the point that it can't be generalized to new data and its accuracy in predicting future outputs would be significantly compromised (Ciaburro, 2017; MathWorks, 2020a, 2020b). The hold-out validation method, which involves training the model on a portion of the data and then validating it using the rest of the data, is suitable when datasets are large. Considering that the study's dataset is not sufficiently large, the 5-fold cross-validation method is recommended in this case and was used for the current analyses. This validation scheme consists of initially dividing the data into five separate sets (or folds). Subsequently, for every set (or fold), it trains the model using the out-of-fold data and then tests its performance using the in-fold data. Finally, it yields an overall error rate, which consists of the mean of the test error estimates obtained across all folds. Following the initial training and testing of the model using separate portions of the dataset, the model is then trained on the full data set (MathWorks, 2020b).

To select the best model for future predictions of output data, the performance of the various trained models using different algorithms was compared by referring to the Root Mean Square Error (RMSE), which provides an estimate of the model's predictive accuracy. RMSE measures the prediction error in predicted values when compared to the actual responses. It has the same units as the response variable. Lower values indicate better model performance (Ciaburro, 2017; MathWorks, 2020c). Moreover, the Predicted vs. Actual Response plot and the

Residuals plot were visually inspected to examine how well the model predicted the true response values. The former plots the predicted response against the actual response. The latter shows the discrepancy between the predicted response values and the actual ones (Ciaburro, 2017; MathWorks, 2020c). Finally, model performance was also assessed by examining to the R-squared statistic. The R-squared value represents the proportion of variance in the response variable explained by the trained model. The closer the R-squared statistic is to one, the better the model performance. A Negative R-squared value indicates that the performance of the trained model is worse than the constant model, in which the response equals the average training response (Ciaburro, 2017; MathWorks, 2020c).

After the training phase, feature selection was performed to identify the features with the most predictive power. Specifically, the embedded type of feature selection was used, whereby feature relevance is determined based on the learning that occurs during model training (Guyon & Elisseeff, 2003; MathWorks, 2020d). Embedded feature selection was implemented by manually generating and running a MATLAB code. It involved training the selected model using Automatic Relevance Determination (ARD). ARD is a regularization technique, which applies a penalty to the model as it becomes more complex, removing superfluous features and identifying those that are most relevant to the prediction of the response variable (Wipf & Nagarajan, 2008).

The final step consists of improving the selected model by the automated hyperparameter optimization option in the Regression Learner app (Ciaburro, 2017; MathWorks, 2020e). The hyperparameter optimization function was used to automate the process of tuning the hyperparameters of the chosen model. This function automatically chooses multiple combinations of the model's internal parameters iteratively, examines their impact on the model's performance to ultimately yield a model with the best hyperparameters. which is associated with the lowest model Mean Squared Error (MSE) estimate. The default option Bayesian optimization was used for the tuning process. The aim of Bayesian Optimization is to identify the set of internal parameters that would minimize the mean squared error (MSE). The Minimum MSE Plot was used to display the findings of the hyperparameter optimization. The trained model statistics were then compared to the new model statistics resulting from the tuning process (Ciaburro, 2017; MathWorks, 2020e, 2020f).

Results

Pre-processing

Descriptive Statistics

The sample consisted of 1364 observations (or cases) with 17 predictors referred to as features and two outcome variables referred to as response variables. The features included four categorical variables (Child's Sex; Child's Ethnicity; 15-mo Attachment; and 36-mo Attachment) and 13 continuous variables (Mother's Age; Mother's Education; Income-to-Need Ratio; Mother's Single Parenthood Status; Maternal Depression; Marital/Partner Relationship Intimacy, mother-report; Mother's Social Support; Mother's Psychological Adjustment; Mother's Parenting Stress; Maternal Sensitivity; Non-Family Child Care Hours; Child Temperament, mother-report; and 24-mo Attachment Q-Sort Security). The two response variables were: Internalizing Problems and Externalizing Problems (mother/alternate caregiver's report).

The sample consisted of 48 % females and 52 % males. The majority of the sample consisted of White American children (80 %) while 20 % had a non-White American background. On average, the children's mothers were 28 of age and had 14 years of education at one-month follow-up. Income-to-need ratio had a mean of 3.34, a standard deviation of 2.69 and a median of 2.71. With respect to this variable, the median fell below the mean, indicating that more than half of the participants fell below the average. This suggests that the sample consisted of a more disadvantaged population than indicated at first glance by a mean that is worth 3-times the poverty line. For more details on the data's descriptive statistics, refer to Table 1 in the methods section.

Assumptions of Parametric Testing

Violations of normality were identified, as shown by the skewness and kurtosis statistics (refer to Appendix F). Z-scores of skewness and/or kurtosis exceeding 3.29 ($p < .001$) indicated violations to normality. A conservative threshold was used for detecting deviations from normality considering the sample size ($N = 1364$) used in the current study (Field, 2009). The following variables were identified as positively skewed: Income-to-Need Ratio; Maternal Depression; Mother's Parenting Stress; Internalizing Problems (mother/alternate caregiver's report); and Externalizing Problems (mother/alternate caregiver's report). The variables that were identified as negatively skewed were: Mother's Single Parenthood Status; Marital/Partner Relationship Intimacy; Mother's Social Support; Maternal Sensitivity; and Attachment Q-sort Security. Two variables were identified as platykurtic included: Mother's Age and Non-Family Child Care Hours. Seven variables were identified as leptokurtic: Income-to-Need Ratio; Mother's Single Parenthood; Maternal Depression; Mother's Social Support; Maternal Sensitivity;

and Child's Temperament (mother's report). Although most of the variables included in the study were not normally distributed, transformations were not performed because Machine Learning predictive modeling is known to be robust to violations of normality (Mueller & Massaron, 2016).

Detection and Treatment of Outliers

The data was examined for univariate outliers using boxplots, which are more appropriate in the context of violations to normality (Rumsey, 2016). One hundred and twenty-two outlying cases were identified (refer to Appendix G for the list of cases that were identified as univariate outliers). Removing outliers is recommended since their presence can bias the learning process that occurs during the training of a model (Mueller & Massaron, 2016). However, the removal of these outliers would significantly reduce the sample size. The sample used for the current study is not considered large enough for Machine Learning predictive modeling and thus further reducing it might compromise the robustness of subsequent analyses. Moreover, since the current study is a secondary data analysis, it was not possible to confirm whether some or all of these outliers are erroneous data and, thus, no strong justification was available for removing any of them. Due to these various considerations, the presence of outliers was addressed by running the subsequent analyses twice, with and without the outlying cases, to check for any discrepancies in findings.

Treatment of Missing Values

SPSS Missing value analysis yielded significant ($p < 0.001$) little MCAR's tests for the two datasets used in this study (the data with and without the outliers), which indicated that the data was not missing completely at Random (Tabachnick & Fidell, 2013). Refer to Appendix H for more information on the little MCAR's tests.

The k-nearest neighbor (kNN) algorithm was used for imputing missing values. This technique is nonparametric and thus appropriate in the context of normality violations. Moreover, it works with any type of variables, whether categorical or continuous (Kowarik & Templ, 2016). The kNN imputation function in the R package VIM was used to impute missing values using the complete dataset and the one excluding the outliers. Refer to Appendix I for the manual codes used to perform the imputations (Kowarik & Templ, 2016). The kNN imputation method imputes missing values on a particular feature by using values on other features in the dataset. There is some reservation regarding using available values on response variables to impute missing values on the features within the dataset (Mueller & Massaron, 2016). However, other resources recommend using as many variables as possible for imputing missing data regardless whether these variables are to be used as predictors or outcomes (Kontopantelis et al., 2017; Meng, 1994). In the current study, all the variables, regardless of whether they were features or response variables, were used for imputing missing values. However, given that the k-Nearest Neighbor algorithm assesses similarity among observations (or cases) based on a distance metric, it tends to be sensitive to the scale used for numeric features and thus standardization of these variables was necessary before performing the imputation (Mueller & Massaron, 2016).

Machine Learning Predictive Modeling using Data Including Outliers

The Regression Learner App in the Statistics and Machine Learning Toolbox in MATLAB was used to build models, which would predict responses on the current study's two outcome variables, namely Externalising Problems and Internalising (mother/alternate caregiver's report). This process was repeated twice, including and then excluding outliers. The first set of analyses used data including the outlying cases.

Training Regression Models. All the regression models within the Regression Learner app were trained in parallel and then compared for best performance for each response variable (Externalizing versus Internalizing Problems). The 5-fold cross validation scheme was used for all the models being trained.

Table 3 shows the statistics related to the models that were trained for the prediction of externalizing problems (mother/alternate caregiver's report). The RMSE values for the trained models ranged between 0.84936 to 1.1255. R-squared ranged between -0.45 and 0.18. R-squared values closer to one indicate better model performance whereas negative values indicate that the trained model's performance is worse than the constant model (Ciaburro, 2017; MathWorks, 2020c). The Gaussian Process Regression (exponential GPR) model had the lowest RMSE (0.84936) and the highest value for R-squared (0.18), indicating it was the model with the best performance. Interestingly, a standard Linear Regression (Linear) model performed almost as well as the above mentioned GPR model (RMSE = 0.85399, R-squared = 0.17).

Figures 1 and 2 depict plots based on the Gaussian Process Regression (Exponential GPR) model, which was selected as the best model for predicting Externalizing Problems. Figure 1 shows the Plot of the Predicted vs. Actual Response. The points represent the true responses, and the diagonal line depicts the predicted responses. The further away the points are from the diagonal line, the greater the errors and the poorer is the performance of the model. As is evident in the figure, the points were spread out asymmetrically around the diagonal line, which indicated that the model, which was selected as the best model, still performed comparatively poorly.

Table 3

Model Statistics for the Prediction of Externalizing Problems

Model Type (Kernel Function)	RMSE	R-squared
Gaussian Process Regression (Exponential GPR)	0.84936	0.18
Gaussian Process Regression (Matern 5/2 GPR)	0.85283	0.17
Gaussian Process Regression (Rational Quadratic GPR)	0.85334	0.17
Gaussian Process Regression (Squared Exponential GPR)	0.85359	0.17
Linear Regression (Linear)	0.85399	0.17
Linear Regression (Robust Linear)	0.85534	0.16
Support Vector Machines (Coarse Gaussian SVM)	0.85833	0.16
Support Vector Machines (Linear SVM)	0.86201	0.15
Support Vector Machines (Medium Gaussian)	0.86555	0.14
Ensemble (Boosted Trees)	0.86785	0.14
Linear Regression (Stepwise Linear)	0.878	0.12
Ensemble (Bagged Trees)	0.88081	0.11
Support Vector Machines (Quadratic SVM)	0.88394	0.11
Tree (Coarse Tree)	0.91428	0.04
Linear Regression (Interactions Linear)	0.92637	0.02
Support Vector Machines (Fine Gaussian SVM)	0.93299	0.01
Support Vector Machines (Cubic SVM)	0.95443	-0.04
Tree (Medium Tree)	0.99571	-0.13
Tree (Fine Tree)	1.1255	-0.45

Figure 2 displays the Residuals Plot. which depicts the discrepancies between the predicted and actual responses Residuals were asymmetrically scattered around zero and varied in magnitude from left to right, which suggests that the model was performing poorly. Table 4 shows the statistics of the models that were trained for the prediction of Internalizing problems (mother/alternate caregiver's report). The RMSE values ranged between 0.87275 to 1.1218. R-squared represents the percent of variance explained in the response variable and it ranged between -0.47 and 0.11.

Figure 1

Plot of Predicted vs. Actual Response for Externalizing Problems

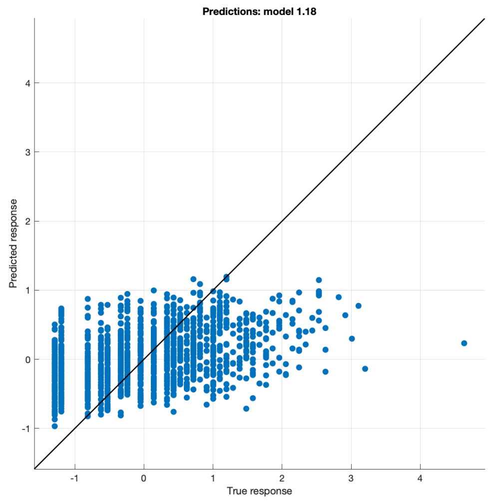
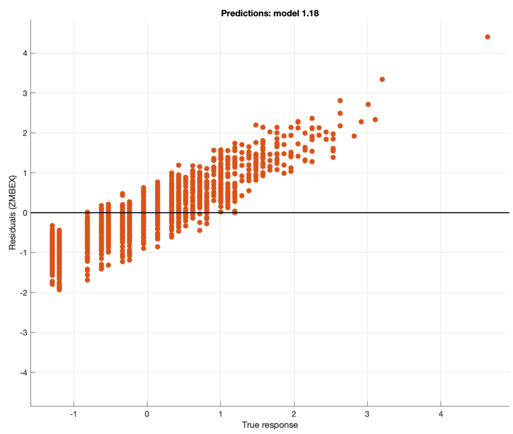


Figure 2

Plot for Residuals Plot for Externalizing Problems



The Gaussian Process Regression (exponential GPR) model had the lowest RMSE (0.87275) and the highest value for R-squared (0.11), indicating it was the model with the best performance and lowest prediction error. This model was followed closely by the Gaussian Process Regression (Relational Quadratic GPR) Model with a slightly higher RMSE (0.87421) and similar R-squared value (0.11). Interestingly, two Linear Regression models (with Linear and Robust Linear Kernel functions) as well as a Support Vector Machines (Coarse Gaussian SVM) model performed almost as well with approximately similar RMSE values and slightly lower R-squared value (0.1).

Table 4

Model Statistics for the Prediction of Internalizing Problems

	RMSE	R-squared
Gaussian Process Regression (Exponential GPR)	0.87275	0.11
Gaussian Process Regression (Rational Quadratic GPR)	0.87421	0.11
Gaussian Process Regression (Matern 5/2 GPR)	0.87581	0.1
Gaussian Process Regression (Squared Exponential GPR)	0.87657	0.1
Support Vector Machines (Coarse Gaussian SVM)	0.87662	0.1
Linear Regression (Linear)	0.87701	0.1
Linear Regression (Robust Linear)	0.87817	0.1
Support Vector Machines (Linear SVM)	0.87929	0.1
Ensemble (Boosted Trees)	0.88301	0.09
Support Vector Machines (Medium Gaussian SVM)	0.8876	0.08
Ensemble (Bagged Trees)	0.89028	0.07
Linear Regression (Stepwise Linear)	0.91287	0.03
Support Vector Machines (Quadratic SVM)	0.9135	0.02
Support Vector Machines (Fine Gaussian)	0.92244	0.01
Tree (Coarse Tree)	0.93814	-0.03
Support Vector Machines (Cubic SVM)	0.96511	-0.09
Linear Regression (Interaction Linear)	0.97416	-0.11
Tree (Medium Tree)	1.01	-0.19
Tree (Fine Tree)	1.1218	-0.47

Figures 3 and 4 present plots from the Gaussian Process Regression (Exponential GPR) model, which had the best performance in predicting Internalizing Problems. Figure 3 shows the Plot of the Predicted vs. Actual Response. Similarly, to the results obtained for the prediction of Externalizing Problems, the points representing the true responses were not scattered symmetrically along the diagonal line, which indicates that the model performed poorly in predicting Internalizing Problems.

Figure 3

Plot of Predicted vs. Actual Response for Internalizing Problems

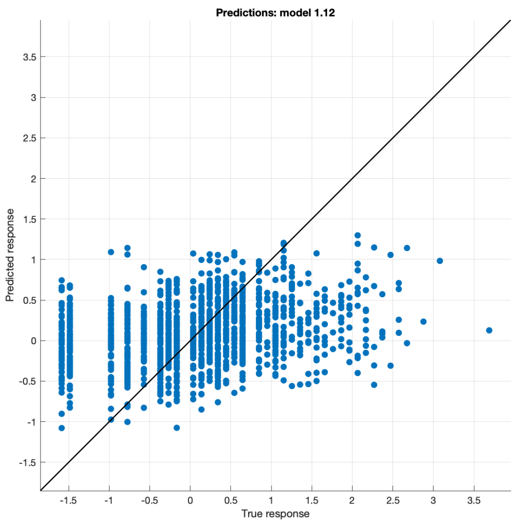
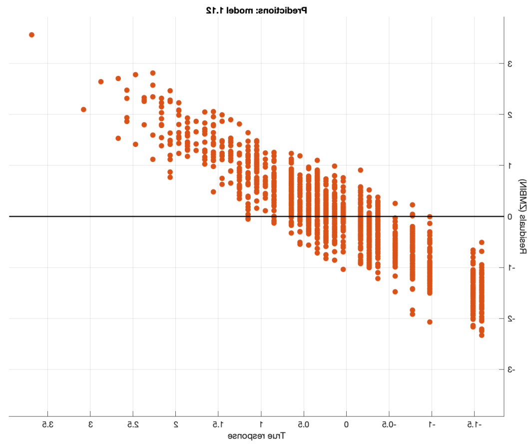


Figure 4 shows the Residuals Plot. Similar to the findings obtained for Externalizing Problems, residuals were not symmetrically scattered around zero and their size varied significantly from the left of the graph to the right, which suggests that the model's performance was far from satisfactory.

Figure 4

Residuals Plot for Internalizing Problems



Feature Selection. Feature selection was conducted based on the Gaussian Process Regression model, which was selected as the best model. The embedded type feature selection was implemented by generating a MATLAB code that allows for the examination of the features' predictive power within that model (see Appendix J for the code). This involved training the selected GPR model again by using the fitting function 'fitrgp' after modifying one of its parameters by using an Automatic Relevance Determination (ARD) kernel ('ardexponential' instead of 'exponential'). This modification allows for the determination of the relevance of each predictor to the prediction task and the assignment of predictor weights accordingly. The larger the weight assigned to a particular feature, the more influential that feature is in predicting the response variable. Features with weights equal or close to zero are considered as irrelevant input data (MathWorks, 2020d, 2020i; Rasmussen & Williams, 2006). Note that MATLAB, when working with the fitting function (fitrgp), uses full dummy coding, whereby each level in a categorical variable is converted to a single binary dummy variable (MathWorks, 2020h).

With respect to the prediction of externalizing problems, the plot of normalised predictor weights (Figures 5 and 6) shows the weights assigned to each feature included in the model. Normalised weights ranged from 0 to 0.2736. Table 5 shows the features and their associated normalised weights in a descending order.

With respect to the prediction of internalizing problems, the plot of normalised predictor weights (Figure 7) shows the weights assigned to each feature included in the model. Normalised weights ranged from 0 to 0.1743. Table 6 displays the features and their associated normalised weights in a descending order.

Figure 5

Plot of Normalised Weights for Features Predicting Externalising Problems (Complete Plot)

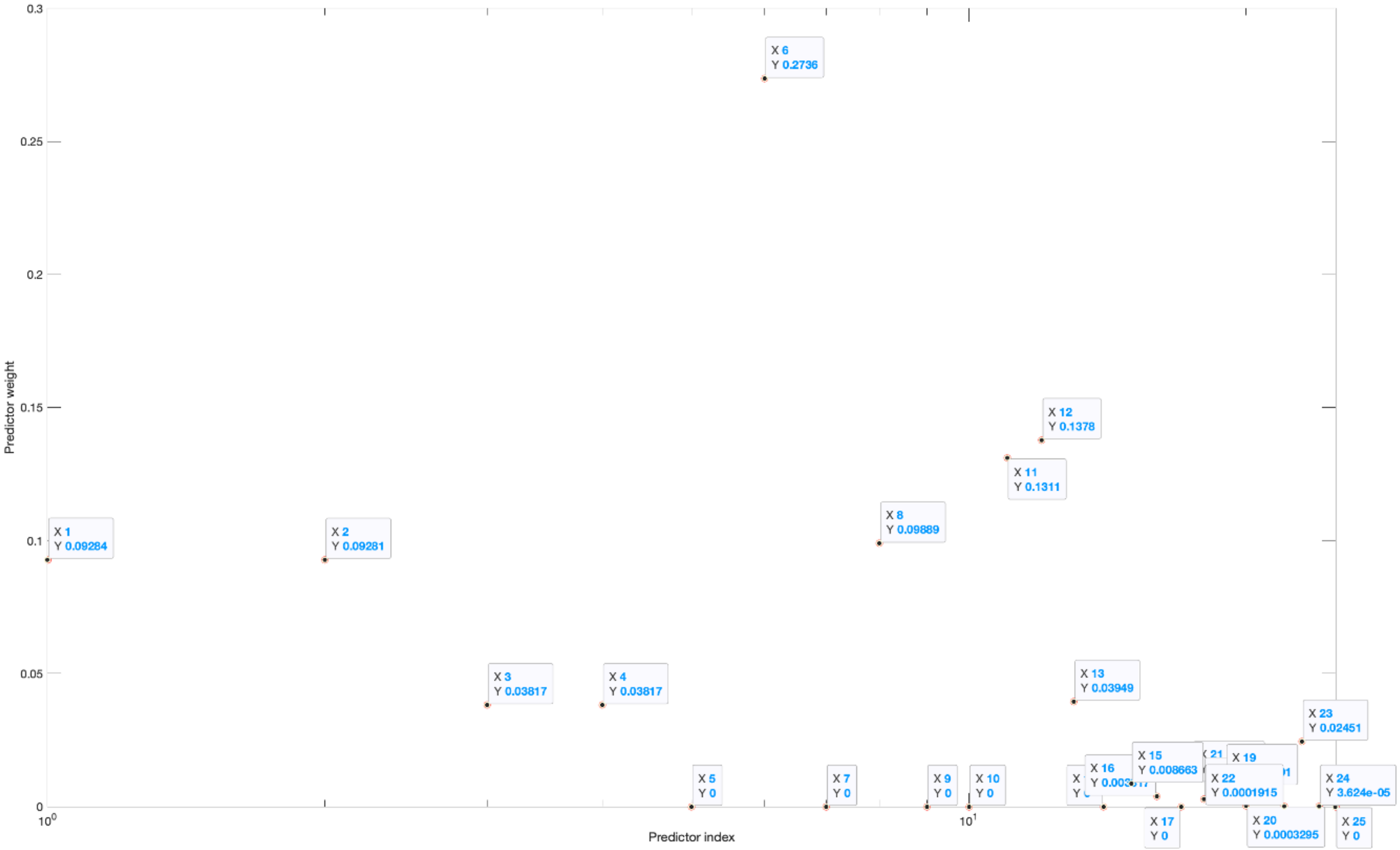


Figure 6

Plot of Normalised Weights for Features Predicting Externalising Problems (Enlarged Bottom Right Corner)

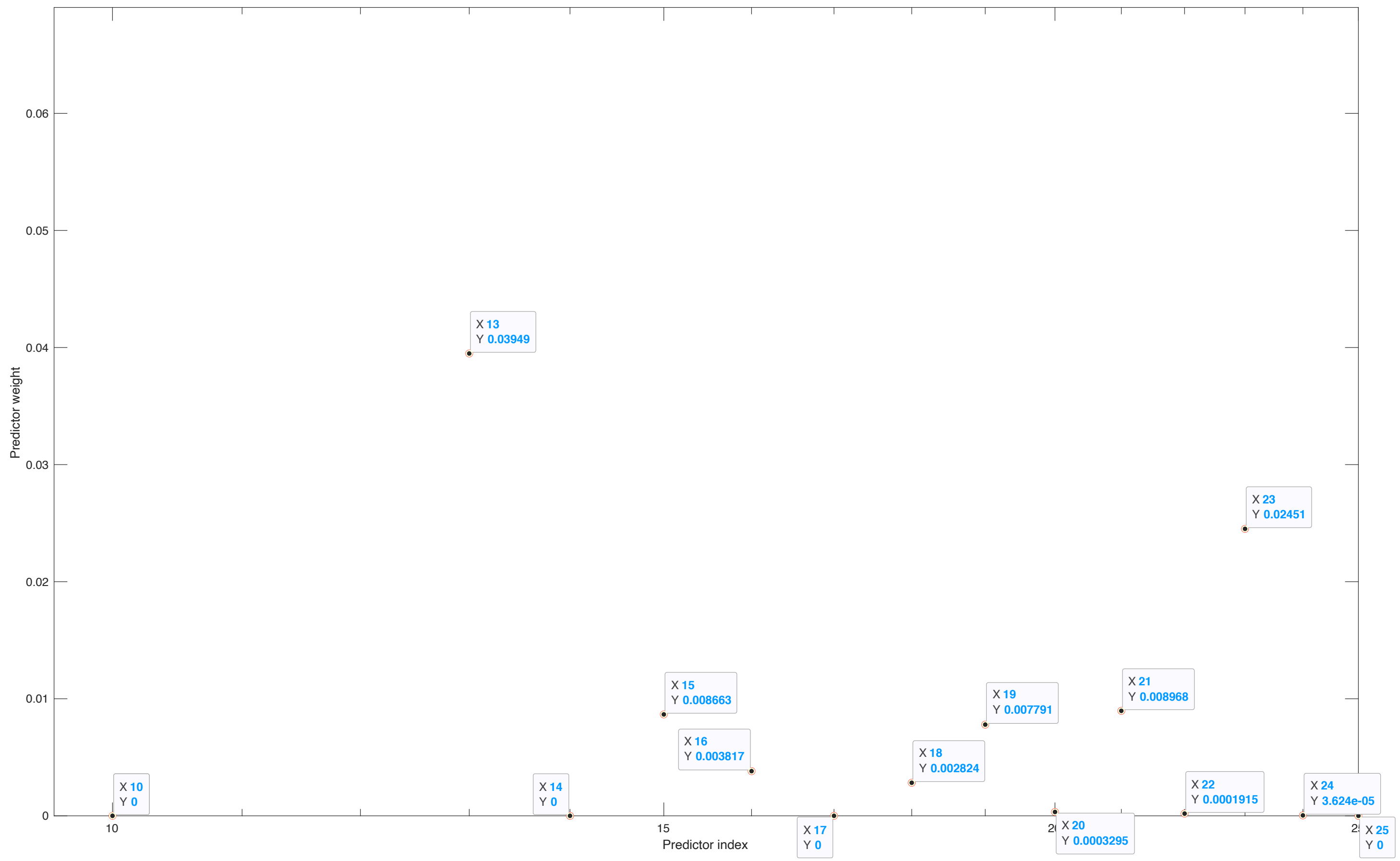


Table 5

Normalised Weights for Features Predicting Externalizing Problems

Feature Name	Predictor Index (x Label)	Normalized Weights (y Labels)
Type B 15-mo Attachment	6	0.2736
Type D 36-mo Attachment	12	0.1378
Type C 36-mo Attachment	11	0.1311
Type D 15-mo Attachment	8	0.09889
Male Child Gender	1	0.09284
Female Child Gender	2	0.09281
Parenting Stress	13	0.03949
White Ethnic Background	3	0.03817
Other Ethnic Background	4	0.03817
Non-Family Child Care Hours	23	0.02451
Social Support	21	0.008968
Maternal Age	15	0.008663
Maternal Depression	19	0.007791
Maternal Education	16	0.003817
Single Parenthood	18	0.002824
Marital/Partner Relationship Intimacy	20	0.0003295
Maternal Psychological Adjustment	22	0.0001915
Child's Temperament (mother-report)	24	3.62E-05
Type A 15-mo Attachment	5	0
Type C 15-mo Attachment	7	0
Type A 36-mo Attachment	9	0
Type B 36-mo Attachment	10	0
Early Maternal Sensitivity	14	0
Income-to-Need Ratio	17	0
24-mo Attachment Q-sort Security	25	0

Figure 7

Plot of Normalised Weights for Features Predicting Internalising Problems

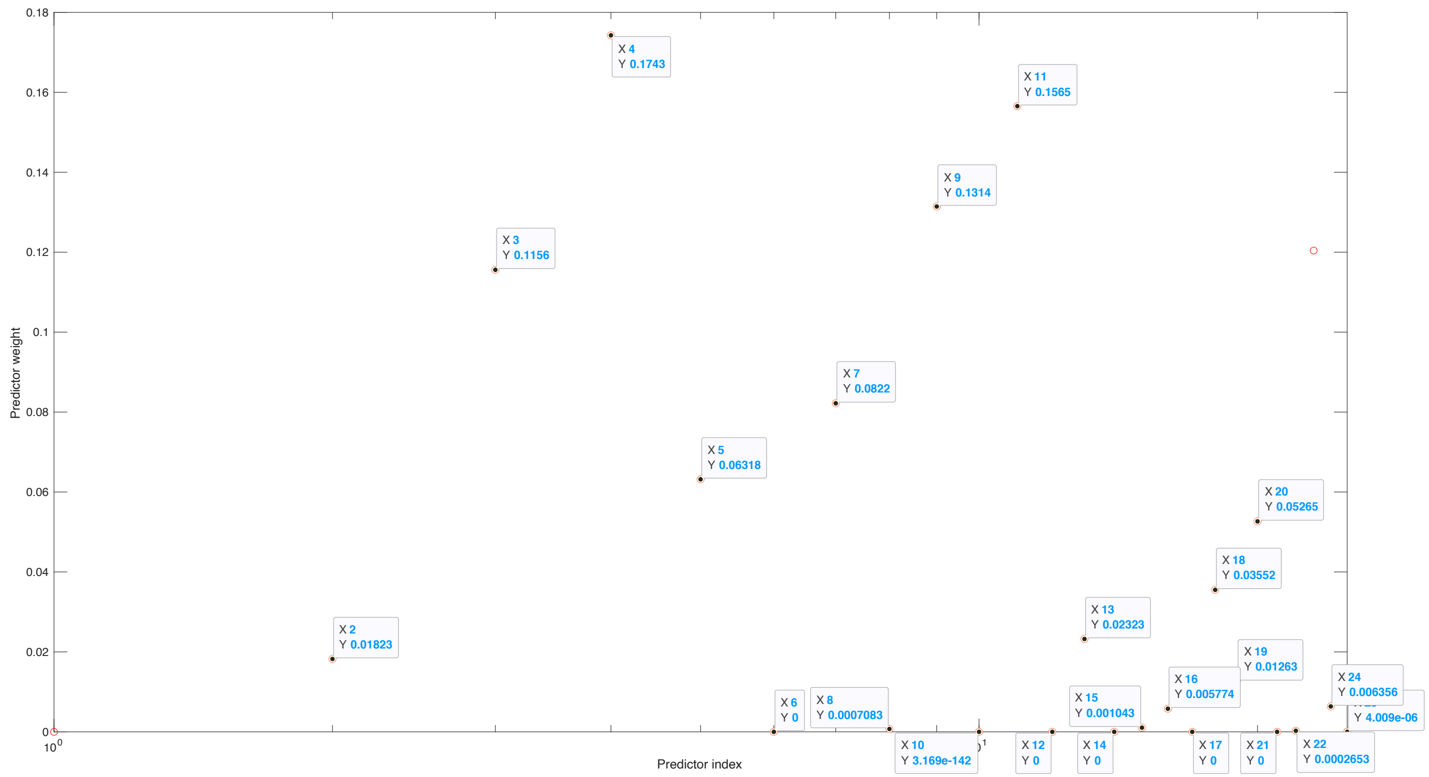


Table 6

Normalised Weights for Features Predicting Internalizing Problems

Feature Name	Predictor Index (x Label)	Normalized Weights (y Labels)
Other Ethnic Background	4	0.1743
Type C 36-mo Attachment	11	0.1565
Type A 36-mo Attachment	9	0.1314
Non-Family Child Care Hours	23	0.1204
White Ethnic Background	3	0.1156
Type C 15-mo Attachment	7	0.0822
Type A 15-mo Attachment	5	0.06318
Marital/Partner Relationship Intimacy	20	0.05265
Single Parenthood	18	0.03552
Parenting Stress	13	0.02323
Female Child Gender	2	0.01823
Maternal Depression	19	0.01263
Child's Temperament	24	0.006356
Maternal Education	16	0.005774
Maternal Age	15	0.001043
Type D 15-mo Attachment	8	0.0007083
Maternal Psychological Adjustment	22	0.0002653
24-mo Attachment Q-sort Security	25	4.01E-06
Male Child Gender	1	7.15E-64
Type B 36-mo Attachment	10	3.17E-142
Type B 15-mo Attachment	6	0
Type D 36-mo Attachment	12	0
Early Maternal Sensitivity	14	0
Income-to-Need Ratio	17	0
Social Support	21	0

Hyperparameter Optimization. Given that a GPR model was found to be the best model in our previous analyses with respect to both response variables, the optimizable model within the Gaussian Process Regression Models category in the Regression Learner app was selected to be trained using automated hyperparameter optimization. This function tests a different combinations of the model's internal parameters at each iteration and, after a pre-determined number of iterations (set at default value of 30), selects the model with the minimum Mean Squared Error (MSE). Similarly to the RMSE, the MSE is another indicator of model performance, which provides a gauge of the prediction error derived when the predicted responses are compared to the actual responses. The lower the value of the MSE, the better the performance of the model (Ciaburro, 2017; MathWorks, 2020c, 2020e, 2020f).

With respect to the prediction of externalizing problems, automated hyperparameter optimization did not yield a much-improved model. The new Gaussian Process Regression (Rational Quadratic GPR) model, as a result of the process of optimization, had slightly lower RMSE value (0.84724) and showed no increase in the R squared value (0.18).

With respect to the prediction of internalizing problems, automated hyperparameter optimization did not yield a much-improved model. The new model (Rational Quadratic GPR), as a result of the process of optimization, had slightly lower RMSE value (0.87093) and showed no increase in the R squared value (0.11).

Table 7 provides a summary of the model statistics comparing the optimizable GPR models to the previous GPR models selected prior to hyperparameter optimization. Figures 2.8 and 2.9 show the Minimum MSE Plots for both response variables, which display the various iterations that occurred as various models with different hyperparameter combinations were trained over and over again to ultimately get to the best combination of internal parameters associated with the lowest MSE estimate.

Table 7

Model Statistics Before and After Hyperparameter Optimization

	Externalizing		Internalizing	
	Before	After	Before	After
Model (Kernel Function)	GPR (Exponential)	Optimizable GPR (Rational Quadratic)	GPR (Exponential)	Optimizable GPR (Rational Quadratic)
RMSE	0.84936	0.84724	0.87275	0.87093
R-squared	0.18	0.18	0.11	0.11

Figure 8

Minimum MSE Plot for Externalizing Problems

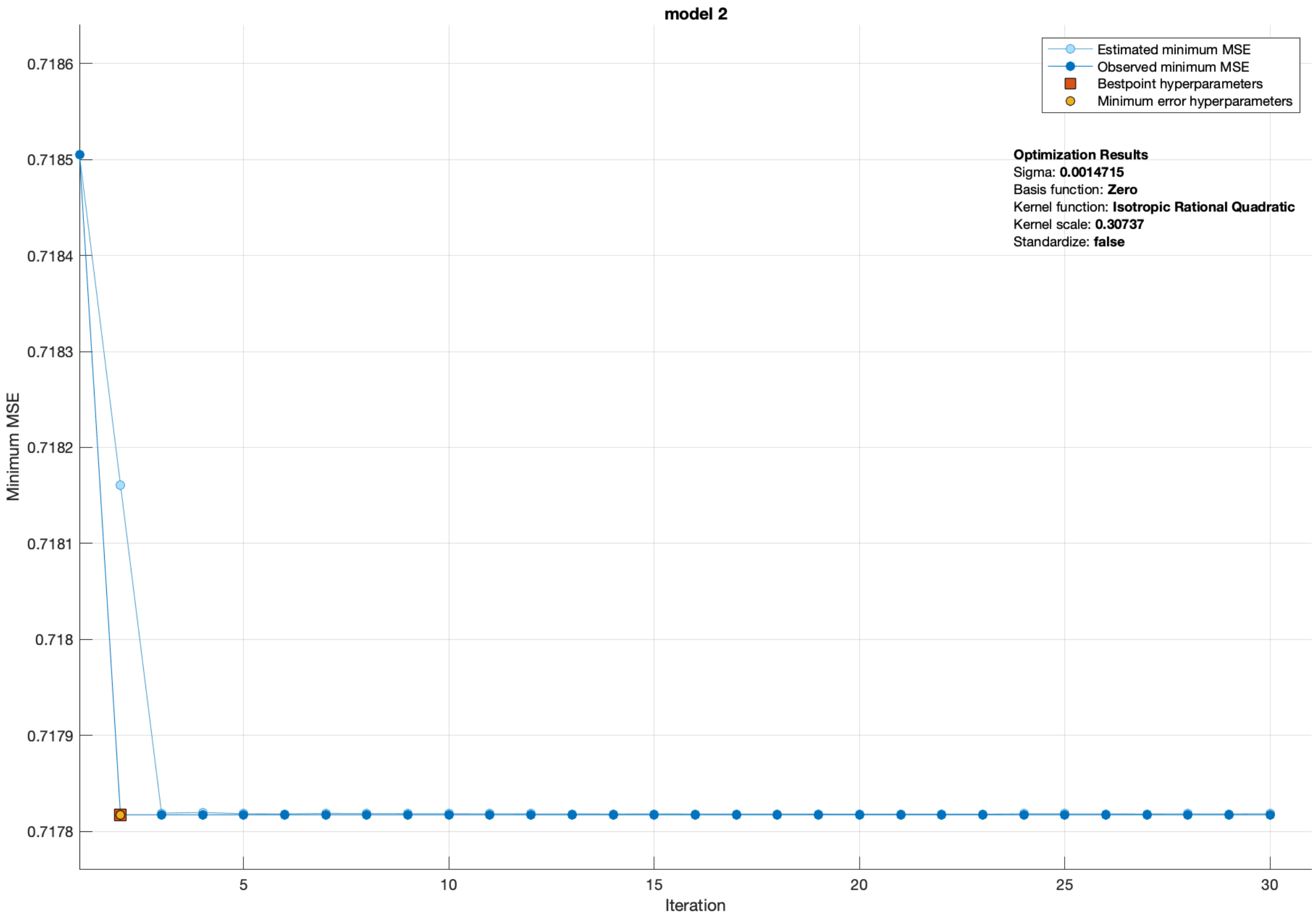
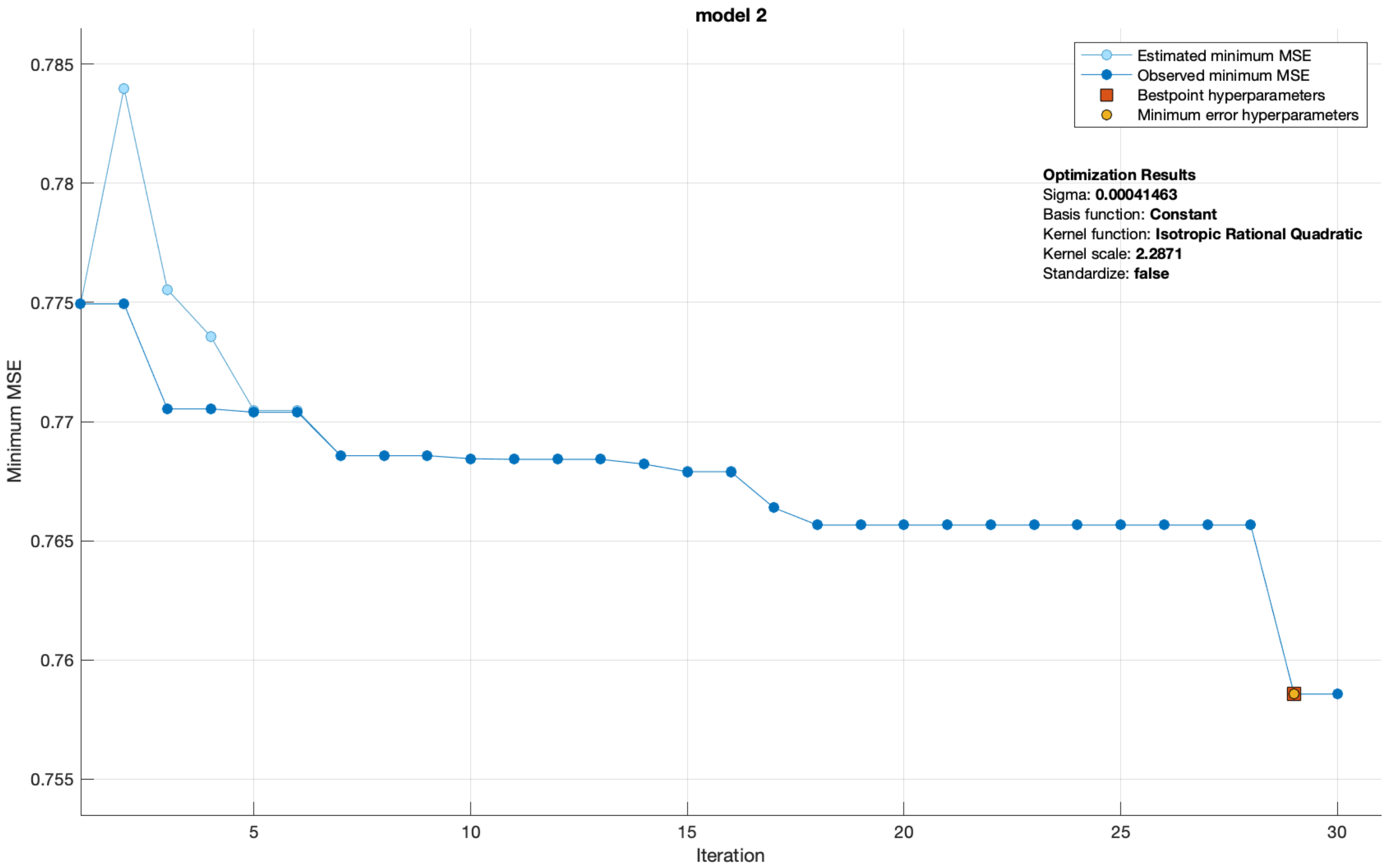


Figure 9

Minimum MSE Plot for Internalizing Problems



Machine Learning Predictive Modeling using Data Excluding Outliers

The same Machine Learning analyses described above were repeated using data excluding outliers to check for any discrepancies in findings before and after the removal of the outlying cases (refer to Appendix K for details about this set of analyses). The findings were somewhat similar. Discrepancies were found with respect to the amount of variance explained in both response (outcome) variables and estimates of predictors' relevance. Slightly lower amounts of variance (1 percent difference) in externalizing (R-squared = 0.17) and internalizing problems (R-squared = 0.10) were explained by the selected GPR (exponential GPR) models.

The results of the embedded feature selection showed more or less minor discrepancies with respect to the ranking of predictors according to their relevance, yet that is expected given that different samples were used across analyses. The only striking inconsistencies were changes that involved a reversal of feature importance across analyses from most relevant or relevant to least relevant or irrelevant, and vice versa. The features mostly affected by this shift in relevance were the 15-month attachment variables. For the prediction of externalizing problems, the 15-month attachment classifications A and C changed from being irrelevant (weight of zero) to most relevant and the 15-month attachment classification D changed from most relevant to irrelevant (weight of zero). For the prediction of internalizing problems, the 15-month attachment classifications B and D shifted respectively from irrelevant and least relevant to relevant, whereas the 15-month attachment classification A changed from relevant to irrelevant (weight of zero). Two other features, parenting stress and marital/partner relationship intimacy, also showed a striking shift from relevance to irrelevance with respect to the prediction of externalizing and internalizing problems, respectively.

Finally, the following is a brief summary of the most consistent findings across analyses (including versus excluding outliers). With respect to externalizing problems, attachment variables (15-month secure, 36-month insecure-controlling/insecure-other, and 36-month insecure-resistant) as well as gender were most relevant. Conversely, a few other attachment variables (36-month avoidant, 36-month secure, and 24-month attachment Q-sort security), maternal sensitivity, as well as temperament and income-to-need ratio were found to be least irrelevant. With respect to internalizing problems, attachment variables (36-months insecure-resistant, 36-months insecure-avoidant, and 15-month ambivalent), ethnicity (“white” and “other”), and Non-family childcare were found to be the most influential predictors. However, other attachment variables (36-month secure attachment, 36-month insecure-controlling/insecure-other, and 24-month attachment Q-sort security), maternal sensitivity, and social support were found to be the least important predictors.

Discussion

The primary goal of the current study was to examine the predictive ability of Machine Learning techniques over orthodox statistical techniques with respect to the response variables, externalizing and internalizing problems. Different types of predictive models were trained using a set of 19 supervised Machine Learning algorithms and then the one with the best performance was selected. Predictive models based on the more commonly used Linear Regression algorithm were compared to the best performing models, which were based on the less conventional Gaussian Process Regression algorithm. A secondary aim was to examine the predictive power of a wide array of characteristics (or features), related to the child and his/her environment during the first three years of life.

With regards to the first aim, the analyses before and after removing outliers, yielded similar results. Gaussian Process Regression (exponential GPR) models were selected for best performance, with respect to the prediction of both response variables. Low percentages of variance in externalizing (18 and 17% before and after removing outliers) and internalizing (11% for both sets of analyses) problems were explained by these models. The performance of the best models in terms of R-squared and RMSE values was only slightly different from the Linear Regression and Support Vector Machine (Coarse Gaussian SVM) models. Linear Regression models are parametric and thus not suitable for the current analyses considering the characteristics of the variables used, which violate the assumptions of normality and absence of collinearity (Field, 2009). The coefficient estimates of the Linear Regression models obtained are unlikely to be reliable, given that the regression design matrices for these models were rank deficient. Rank deficiency occurs when there is collinearity in the regression model, whereby “one or more of the independent variables are a linear function of the other independent variables in the model” (O’Brien, 2012).

With respect to comparing Gaussian Process Regression and Support Vector Machines models, both have been used for the detection and diagnosis of mental health problems, with different types of input data (Shatte et al., 2019). According to the no-free-lunch theorem, no Machine Learning algorithm can be suitable for every possible dataset (Wolpert, 2002). It is therefore good practice to try and compare several algorithms based on several criteria to select the best one for a particular set of input data (Gómez & Rojas, 2016; Ogundepo & Fokoué, 2019). Both types of models are nonparametric and thus suitable for nonnormally distributed data and can handle nonlinear relationships. However, the Support Vector Machine (Coarse Gaussian SVM) model is less flexible and more susceptible to overfitting when tuned to increase its flexibility, whereas the Gaussian Process Regression model's degree of flexibility is automatically set to simultaneously reduce prediction error as well as the risk of overfitting (MathWorks, 2020g). Both model types are hard to interpret, yet unlike the case with the Support Vector Machine model, embedded feature selection can be implemented while fitting the Gaussian Process Regression model using MATLAB, which enhances its usefulness and interpretability (MathWorks, 2020d). Moreover, embedded feature selection that is specific to a Gaussian Process Regression model involves the use of automatic relevance determination, which is a procedure that prunes the least influential features to yield a sparse model at a reduced risk of overfitting (Wipf & Nagarajan, 2008). One disadvantage of the Gaussian Process Regression algorithm compared to the Support Vector Machine algorithm is that it is computationally demanding and may have a slower training phase, yet this is only relevant for high-dimensional datasets, which is not the case in the current study (MathWorks, 2016b). Thus, considering all of the above advantages of the Gaussian Process Regression model and its superior performance in the current analyses, it was selected as the best model.

The small magnitude of the variance explained by the predictive models suggests that other features, which are also important to the prediction of the response variables, were not included in the modeling process. This is consistent with the many theories of developmental psychopathology suggesting a myriad of potential predictors of psychopathology, other than the ones examined in the present study, such as parenting practices and styles; parental modeling, pattern of reinforcement, and instruction; family functioning characteristics; peer relationships; disturbances in the behavioral activation and inhibition motivational systems; general life stress; specific stressors; genetic predispositions and their expression; and aspects of the cultural context (Rudolph et al., 2016). Moreover, the developmental psychopathology perspective posits that early life experiences can be predictive of later adjustment outcomes yet not fully deterministic, emphasizing the potential contribution of later experiences in the development of psychopathology (Rudolph et al., 2016). Moreover, differences in the magnitude of variance explained by antecedent variables between externalizing and internalizing variables were previously reported in the literature. For instance, meta-analytic findings examining observer-assessed early attachment as an antecedent variable did report stronger associations with externalizing compared to internalizing problems (Fearon et al., 2010; Groh et al., 2012). A similar pattern of results was also reported when parenting stress was used as an antecedent variable (Barroso et al., 2018).

Embedded feature selection was performed to primarily determine the predictive power of the features included in the Gaussian Process Regression models chosen for best performance. Embedded feature selection using Automatic Relevance Determination ultimately aims to identify the subset of features that are most relevant to the prediction of the response (outcome) variable while the model is being trained, which will then be used to build a sparse model with greater predictive accuracy. Thus, Machine Learning predictive modeling's main focus is on maximising predictive accuracy rather than providing explanations and theory building (Guyon and Elisseeff, 2003; Yarkoni & Westfall, 2017).

This raises two important issues. First, with its focus on predictive accuracy, supervised Machine Learning fails to enhance our understanding about the nature of the relationships between variables and related causal mechanisms. The normalised weights obtained in embedded feature selection provide information about the predictive power of individual features, but it cannot inform us about the direction of the association between each of the features and the response variable.

Second, the ranking of each features in embedded feature selection using Automatic Relevance Determination occurs as the model is being created and in the context of other features. Thus, the determination of relevance depends on the learning process of the model being trained and the features that were included the model. A feature found to be relevant or irrelevant due to being redundant with other features with greater predictive power, may not necessarily be unrelated to the response variable in the real world. The ranking of features merely provides us with information on how to predict the response variable most accurately and efficiently in the context of the specific subset of features used and cannot robustly inform us about the importance of these features to specific outcomes in the real world (Wipf & Nagarajan, 2008; Guyon and Elisseeff, 2003).Thus, it is crucial that one interprets the findings of the current study holding the above considerations in mind, which are inherent to Machine Learning predictive modeling due to its focus on predictive accuracy over explanation. The current discussion will address the findings of feature selection, which remained fairly consistent, across the two sets of analyses, before and after the removal of outliers.

Attachment as assessed by the Strange Situation Procedure (SSP) was found to be particularly important for the prediction of both response variables.

Out of the four 15-month SSP attachment full dummy coded variables, one variable for each response variable was found to be an important feature, including 15-month Secure attachment for externalizing problems and 15-month Resistant attachment for internalizing problems. Out of the four 36-month SSP attachment full dummy coded variables, two variables for each response variable emerged as important features. The 36-months insecure-controlling/insecure-other (or disorganized) and insecure-avoidant attachment classifications were found to be influential for externalizing and internalizing problems, respectively, and 36-month ambivalent attachment emerged as important for both response variables.

A few interesting findings regarding attachment are worth discussing. First, more 36-month attachment variables were identified as important features than 15-month attachment variables. This is consistent with meta-analytic findings showing a significant moderating effect for the age at attachment assessment (using observational measures) ($k = 59$, $b = .004$, $p = .01$), suggesting that the association between insecurity and internalizing problems was stronger at a later age (Madigan et al., 2013).

Moreover, in the current study, 15-month attachment was assessed based on the standard SSP whereas the assessment of 36-month attachment was based on the modified SSP. In a previous meta-analysis by Groh et al. (2012), the combined effect size of the association between attachment insecurity and internalizing psychopathology was found to be greater in the studies which used the modified SSP (d = 0.29) compared to those using the standard SSP (d = 0.10) (Groh et al., 2012).

Second, of these eight features, the 36-month ambivalent attachment classification emerged as one of the most important features for the prediction of both externalizing and internalizing problems. This is largely inconsistent with meta-analytic findings reporting insignificant combined effect sizes, showing consistently that this attachment classification is not associated with externalizing (Fearon et al., 2010) nor internalizing problems (Groh et al., 2012; Madigan et al., 2013). However, most of the studies included in these meta-analyses examined early observational assessments of attachment in relation to later behaviour problems in children younger than 12. Thus, these findings do not necessarily generalise to the association of early attachment to the emergence of behaviour problems in adolescence or when other types of attachment measures are used. Interestingly, another meta-analysis by Madigan et al. (2016) examining this association in a wider age group (age 3-18) and using representational or questionnaires measures of attachment reported a different finding, with ambivalent attachment being significantly associated with internalizing problems (d = .40).

Third, the findings of the current study indicated that avoidance and disorganization as assessed by observational methods were differentially related to externalizing and internalizing problems. The 36-month avoidant attachment classification was found to be an important feature for the prediction of internalizing problems yet an irrelevant for the prediction of externalizing problems. The 36-months insecure-controlling/insecure-other (or disorganized) classification was found to be an important feature for the prediction of externalizing problems, yet an irrelevant one for the prediction of internalizing problems. This is somewhat in line with the results of meta-analyses examining the association between attachment and behaviour problems (Groh et al., 2012; Fearon et al., 2010; Madigan et al. 2013). In relation to internalizing problems, consistently significant and stronger combined effect sizes were reported for avoidance compared to disorganization, for which combined effect sizes were consistently weaker and most frequently insignificant (Groh et al., 2012; Madigan et al. 2013; Madigan et al., 2016). Conversely, with respect to externalizing problems, the opposite pattern of findings was reported, with disorganization having consistently significant and stronger associations compared to avoidance, for which either insignificant or weaker associations were reported (Fearon et al., 2010; Madigan et al., 2016).

Fourth, interestingly, the 24-mo Attachment Q-Sort Security and maternal sensitivity were found to be irrelevant, with zero or near-zero predictor weights for both response variables, with data including or excluding outlying cases. This is inconsistent with previous research showing significant associations between these variables and behaviour problems. In the meta-analysis by Fearon et al. (2010), the association between attachment and externalizing problems was found to be stronger in studies using Attachment Q-Sort security (d = 0.70) compared to those using a measurement based on the standard SSP (d = 0.18). However, there was some evidence that the effect of age at attachment assessment may be confounded with type of measure used, so it is difficult to disentangle the potential contribution of these two variables (Fearon et al., 2010; Madigan et al. 2013). Interestingly, with respect to maternal sensitivity, the finding of the current secondary analysis of the NICHD SECCYD data was not consistent with prior analyses using data from the same study. Maternal sensitivity in middle childhood was found to be a significant moderator of the association between early childcare and externalizing behaviour problems at age 15 (Burchinal et al., 2014). Moreover, in a longitudinal study (N = 265) by Doan et al. (2012), maternal responsiveness significantly mediated of the effect of early contextual risk on internalizing problems in adolescence. However, these unexpected findings are not surprising considering how embedded feature selection operates by treating redundant features as unimportant and consequently setting their predictor weights to zero or near-zero. The above features identified as irrelevant were shown to be not only conceptually but also statistically related to the attachment variables which emerged as most relevant (De Wolff & Van Ijzendoorn, 1997; Van Ijzendoorn et al. 2004), and thus could have been treated as superfluous partly as a result of this shared variance.

In addition to providing evidence highlighting the importance of attachment to the prediction of socioemotional development outcomes, these findings also show that another aspect of early interpersonal experiences is also relevant to the prediction of later behaviour problems. Early non-family childcare was found to be consistently one of the most powerful predictors of internalizing problems. Although the degree of importance of early nonfamily childcare was not consistent across analyses (including versus excluding outliers) with respect to predicting externalizing problem, it still emerged as a relevant predictor. These findings were partly in line with previous research.

Prior analyses of NICHD SECCYD data used in the current study focused on examining different aspects of early childcare in relation to externalizing problems. The findings showed that more time spent in child care was associated with greater externalizing problems in early childhood (NICHD Early Child Care Research Network [NICHD ECCRN], 2003) and adolescence (Belsky et al., 2007; Burchinal et al., 2014; Vandell et al., 2010), yet not in middle childhood (NICHD ECCRN, 2005). Consistent results were reported by a longitudinal study using a nationally representative sample (N = 6000), whereby full-time childcare in preschool was associated with increased behaviour problems (Coley et al., 2013). However, these findings which are based on diverse samples in terms of socioeconomic status, do not generalize to financially disadvantaged families, as shown by previous research, whereby more hours in childcare were found to be associated with less externalizing problems in middle childhood and adolescence (Orri et al., 2019; Votruba-Drzal et al. 2004). Interestingly, the amount of early childcare was not extensively examined in relation to internalizing problems. One study by Votruba-Drzal et al. (2004) showed that the amount of early childcare was not found to be significantly associated with internalizing problems in middle childhood (Votruba-Drzal et al., 2004). To date, the current study is one of the few studies to examine the association between early non-family childcare hours and internalizing problems in adolescence and found a significant association between these two variables. However, more research is needed to further elucidate the nature of this relationship.

Of the contextual risk variables, ethnicity variables ("white ethnic background" and "other ethnic background") were found to be some of the most important predictors of internalizing problems. These features were also found to be relevant predictors for externalizing problems, yet not as important as they were for the prediction of internalizing problems. These findings were in line with research showing ethnicity as a significant moderator of the association between various early environmental factors and later behaviour problems. Meta-analytic findings showed that the combined effect sizes for the associations between attachment and each of the socioemotional development indicators, externalizing and internalizing problems, were larger in white children (Madigan et al., 2016). Similar results were also reported for other early environmental characteristics (e.g., family turbulence and physical discipline) in relation to both behaviour problems domains, with these antecedent variables being significant predictors only in children of white ethnic background (Lansford et al., 2004; Womack et al., 2019). The apparent increased risk of behaviour problems in White children raises questions regarding the extent to which the effect of early environmental characteristics is universal or culture specific. More research is needed to further elucidate the potential role of culture in the development of psychopathology.

Most of the remaining contextual risk variables fell below attachment variables in terms of importance yet were still relevant for the prediction of both outcome variables. These findings suggest that contextual risk variables were not redundant and independently explained additional variance in the response variables. This is somewhat in line with prior analyses of data from the NICHD SECCYD, which also provided evidence for the contribution of these two sets of variables in the prediction of behaviour problems. Early cumulative contextual risk was found to have a significant (unadjusted) direct effect on total behaviour problems at the age of 3 (Belsky & Fearon, 2002a).

Moreover, the findings of other research unrelated to the NICHD SECCYD also showed the independent effects of contextual risk in predicting behaviour problems. In a longitudinal study of 566 children followed overtime throughout middle childhood (from 5 to 10 years), the set of contextual risk variables and that of parenting/caregiving characteristics were found to contribute 1 to 4 % and 2 to 6 % to the unique variance in externalizing problems, respectively (Deater-Deckard et al.,1998). In another longitudinal study, which followed 206 males from early adolescence into adulthood over a period of 10 years, contextual variables were significantly associated with variations in depressive symptoms and parenting practices did not mediate these effects (Kim et al., 2003).

Of the two child characteristics, temperament and gender, only the latter was found to be consistently important for the prediction of externalizing problems across analyses, before and after removing outliers. Initially, both male and female gender emerged as highly relevant predictors. However, after removing the outlying cases, female gender remained a highly important predictor, but male gender emerged as less influential but still relevant. This is partly consistent with previous research showing gender differences in the risk of developing behaviour problems (Zahn-Waxler et al., 2008). Yet, the current finding is unexpected given that female gender stood out as a more important predictor of externalizing problems than male gender in the analysis excluding outliers. This stands in contrast with prior research reporting evidence showing that male gender is associated with a higher risk of developing externalizing problems (Ara, 2016; Barroso et al., 2018; Fearon et al., 2010; Careneiro et a., 2016). However, a review by Ara (2016) showed evidence suggesting that adolescent girls were more likely to have comorbid externalizing and internalizing problems compared to adolescent boys. This could possibly account for the current finding showing female gender as possibly a superior predictor of externalizing problems at age 15 compared to male gender. However, considering that the analyses for the prediction of externalizing problems, before and after removing outliers, did not yield consistent results in relation to the individual contribution of male and female gender, these findings should be interpreted with caution. Further research is needed to examine whether co-occurring externalizing and internalizing problems differs from the pure presentations in terms of risk profile and aetiology. However, with respect to predicting internalizing problems, the findings of the current study were in line with previous research (Carneiro et a., 2016; Wang et al., 2018), with female gender emerging as a more influential predictor than male gender, before and after the removal of outlying cases.

Although the findings were generally consistent, discrepancies were found with respect to the ranking of predictors in terms of relevance across analyses, before and after removing outliers. The most striking inconsistencies mainly affected the 15-month attachment variables, showing that the embedded feature selection may not have been robust to the presence of outliers. This is expected since outlying cases introduce bias in the model's learning process potentially compromising its generalizability (Mueller & Massaron, 2016). However, this does not mean that the findings of the second set of analyses excluding the outliers are less biased since the removal of outlying cases further reduced the analytic sample, which was already not big enough for the implementation of Machine Learning. Inadequately sized sample sizes for model training are not ideal for building a generalizable model (Cearns et al., 2019; MathWorks, 2016b).

The current study extended previous research, in that it implemented a novel analytic approach to the prediction of long-term mental health outcomes, including a large number of predictors, which standard statistical techniques can't accommodate. Machine learning using MATLAB allowed for the efficient and simultaneous training and direct comparison of a wide set of parametric and nonparametric models. Moreover, the NICHD SECCYD from which the data was driven, is known for its intricate design, sample diversity, long follow-up period (from one month to the age of 15), extensive data collections procedures, and use of observational assessment tools of childcare experiences (NICHD ECCRN, 2005a).

However, the study had several limitations that are important to discuss and eventually address in future research. The most obvious limitations related to the design of the NICHD SECCYD were (a) the inability to make causal inferences, (b) the limited generalizability of the findings to clinical samples, and (c) the nonrepresentative nature of the sample which did not include high-risk populations (e.g., those who resided in dangerous neighbourhoods). Moreover, several issues may have compromised the statistical validity of the current analyses. First, Machine Learning algorithms are robust to violations of normality but not to the presence of outliers. It is hard to distinguish erroneous data from data that represents the real world and make a confident assessment of the validity of the results before and after the removal of outlying cases (Field, 2009; Mueller & Massaron, 2016). Second, there is no perfect solution for addressing missing values, especially if the missingness is not completely at random (MCAR), which was the case in the current dataset (Tabachnick & Fidell, 2013). Although kNN as a nonparametric Machine Learning algorithm is robust to violations of normality, its ability to perform well on data that is missing at random (MAR) or not missing at random (NMAR) is questionable. Moreover, kNN imputation is likely to yield a well-behaved dataset, since it relies on predicting missing values on particular features based on available values of other features. Thus, the imputed values are more likely to be within the range of expected values in the dataset and are less likely to show noise than the real values (Batista & Monard, 2002). Third, the validity of the 5-fold cross-validation process is questionable due to the presence of data leakage. Data leakage occurred because values in the testing set were influenced by values in the training sets as a result of implementing standardization and missing values imputation using the entire dataset. The testing sets have to be completely independent from the training sets, and when this is breached the cross-validation process can yield overly optimistic estimates of the quality of the model performance (Cearns et al., 2019). Moreover, although the study's sample size is considered large enough for traditional statistical techniques, this is not the case for Machine learning predictive modeling, which is more suitable for high-dimensional data. Models trained on large dataset are more likely to perform well on new data (Cearns et al., 2019; MathWorks, 2016b). Thus, the generalizability of the model obtained in the current study may be compromised.

Lastly, with respect to the incremental utility of the findings, it is important to note the limited ability of Machine Learning predictive modeling to provide adequate explanations of psychological phenomena for theory building. Moreover, although it is clear that features with zero normalised weights are to be excluded, the case for the features with near zero normalized weights is less straightforward. So far, no clear criteria have been made available providing a threshold for feature inclusion/exclusion, which makes the process of determining feature relevance less systematic and more arbitrary, further compromising the interpretability and hence utility of the findings.

A few recommendations are worth considering in future research regarding the use of supervised Machine Learning for the prediction of externalizing and internalizing problems and other mental health outcomes, First, given that the amount of explained variance was low for internalizing and to a lesser extent externalizing problems, future endeavours may benefit from using different types of measurements for these outcomes variables and samples from clinical settings. Groh et al. (2012) suggested that the reporting of internalizing problems may be less straightforward than that of externalizing problems, which may have compromised the construct validity of the measurement used. It was suggested that future research may benefit from using observer- or clinician- assessed behaviour problems as outcome variables. It may also be useful to use assessment of behaviour problems completed with respondents other than the mother, given that the association between attachment and these outcomes was found to vary depending on the informant (Madigan et al., 2016). Moreover, given that the sample used in the NICHD SECCYD study was community-based, it may be worthwhile to examine whether the explanatory power of the predictive model used in the current study may be greater in clinical populations. A larger combined effect size was reported for the association of attachment and externalizing problems in studies using clinical samples compared to those using nonclinical samples (Fearon et al., 2010). Interestingly, most research examining the role of early interpersonal experiences on later behavioural problems have focused on the mother as the primary caregiver. A greater focus on examining the contribution of father-child interactions may shed more light on the role of the quality of parenting in later development. Finally, with respect to model generalizability, the use of external samples, which are completely independent from the training sets for model testing, is necessary to reduce the risk of overfitting. This can be achieved by using the hold-out or nested cross-validation schemes for model testing (refer to part III for a more thorough discussion). Some of the findings of the current study were not consistent with the literature. These inconsistent findings may represent genuine relationships between variables that were not previously properly captured by traditional statistical techniques, that are limited in their ability to work with non-linear relationships compared to Machine Learning Gaussian Process Regression models. However, before drawing any meaningful conclusions about the validity of these results and the added value of using Machine Learning, it is crucial to test the models obtained in the current study on previously unseen data, to assess for generalizability and predictive accuracy.

The findings of the current study showed that early interpersonal experiences as well as mother and child characteristics account for only a small portion of the variance in behaviour problems in adolescence, especially in the case of internalizing problems. Of these early antecedent variables, early attachment emerged as the most influential, suggesting that the quality of the relationship with the mother early on may have an enduring effect on mental health later in life. The results of feature selection were generally in line with previous research, providing evidence for the potential utility of supervised Machine Learning techniques in predicting future mental health outcomes and the identification of important risk factors which can inform prevention and early intervention efforts. Future research should further examine the utility of this novel analytic approach in the field of psychology and psychiatry and focus on making the implementation of these techniques more grounded in theory and systematic and the reporting of the results more transparent and informative. The use of external samples for model testing should be prioritized, as the applicability of the trained models to real-world data can’t be determined otherwise.

References

Abidin, R. R. (1983). Parenting Stress Index Manual. Charlottesville, VA: Pediatric Psychology Press.

Abidin, R. R. (2012). *Parenting stress index* (4th ed.). Lutz, FL: PAR.

Achenbach, T. M. (1991). Manual for the Child Behavior Checklist/4-18 and 1991 profile. *University of Vermont, Department of Psychiatry*.

Ainsworth, M. D. S., & Bell, S. M. (1969). Some contemporary patterns in the feeding situation. In A. Ambrose (Ed.), Stimulation in early infancy (pp. 133–170). London: Academic Press.

Ainsworth, M. D., & Wittig, B. A. (1969). Attachment and exploratory behavior of one-year-olds in a strange situation. In B. M. Foss (Ed.), Determinants of infant behavior (Vol. 4, pp. 113-136). London: Methuen.

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. (1974). Infant–mother attachment and social development: “Socialization” as a product of reciprocal responsiveness to signals. In P. M. Richards (Ed.), The integration of a child into a social world (pp. 99–135). Cambridge: Cambridge University Press.

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). Patterns of attachment: Patterns of attachment: A psychological study of the Strange Situation . Hillsdale, N: Erlbaum.

American Psychological Association. (2011). *Parenting Stress Index*. Retrieved from https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/parenting-stress

Ara, E. (2016). Internalizing and externalizing problems in adolescents analyzing the gender difference. International Journal of Research in Social Sciences, 6(1), 328-337.

Ashford, J., Smit, F., Van Lier, P. A., Cuijpers, P., & Koot, H. M. (2008). Early risk indicators of internalizing problems in late childhood: a 9-year longitudinal study. *Journal of Child Psychology and Psychiatry*, *49*(7), 774-780.

Bali, R., Sarkar, D., Lantz, B., & Lesmeister, C. (2016). R: Unleash Machine Learning Techniques. Packt Publishing Ltd.

Barnett, R.C. & Marshall, N.L. (1991). The relationship between women's work and family roles and their subjective well-being and psychological distress. In M. Frankenhaeuser, U. Lundberg, & M. Chesney (Eds.), Women, work and health: Stress and opportunities. (pp. 111-136). New York: Plenum.

Barnett, R.C., Marshall, N.L., & Pleck, J. H. (1992). Men's multiple roles and their relationship to men's psychological distress. Journal of Marriage and the Family, 54, 358-367

Barroso, N. E., Mendez, L., Graziano, P. A., & Bagner, D. M. (2018). Parenting stress through the lens of different clinical groups: A systematic review & meta-analysis. Journal of abnormal child psychology, 46(3), 449-461.

Batista, G. E., & Monard, M. C. (2002). A Study of K-Nearest Neighbor as an Imputation Method. *His*, *87*(251-260), 48.

Bayer, J. K., Ukoumunne, O. C., Mathers, M., Wake, M., Abdi, N., & Hiscock, H. (2012). Development of children's internalising and externalising problems from infancy to five years of age. *Australian & New Zealand Journal of Psychiatry*, *46*(7), 659-668.

Belsky, J., & Fearon, R.M.P. (2002a). Infant–mother attachment security, contextual risk, and early development: A moderational analysis. Development and Psychopathology, 14, 293–310.

Belsky, J., & Fearon, R.M.P. (2002b). Early attachment security, subsequent maternal sensitivity, and later child development: Does continuity in development depend on continuity in caregiving? Attachment & Human Development, 3, 361–387.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care?. *Child development*, *78*(2), 681-701.

Bowlby, J. (1969). Attachment and loss: Vol. 1. Attachment. New York: Ba- sic Books.

Bowlby, J. (1973). Attachment and loss: Vol. 2. Separation: Anxiety and an- ger. New York: Basic Books.

Bowlby, J. (1980). Attachment and loss: Vol. 3. Loss. New York: Basic Books

Broerman, R. (2020). Diathesis-Stress Model. *Encyclopedia of Personality and Individual Differences*, 1107-1109.

Burchinal, M. R., Lowe Vandell, D., & Belsky, J. (2014). Is the prediction of adolescent outcomes from early child care moderated by later maternal sensitivity? Results from the nichd study of early child care and youth development. *Developmental Psychology*, *50*(2), 542.

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of Machine Learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of affective disorders*, *245*, 869-884.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the infant temperament questionnaire. *Pediatrics*, *61*(5), 735-739.

Carneiro, Alexandra & Dias, Pedro & Soares, Isabel. (2016). Risk Factors for Internalizing and Externalizing Problems in the Preschool Years: Systematic Literature Review Based on the Child Behavior Checklist 1½–5. Journal of Child and Family Studies. 25. 10.1007/s10826-016-0456-z.

Cassidy, J., & Marvin, R. S., with the MacArthur Attachment Working Group. (1992a). *Attachment organization in preschool children: Coding guidelines* (4th ed.). Unpublished manuscript, University of Virginia.

Cassidy, J., Parke, R.D., Butkovsky, L., Braungart, J.M. (1992b). Family-peer connections: the roles of emotional expressiveness within the family and children's understanding of emotions. Child Development, 63(3), 603-18.

Cearns, M., Hahn, T., & Baune, B. T. (2019). Recommendations and future directions for supervised Machine Learning in psychiatry. *Translational psychiatry*, *9*(1), 1-12.

Ciaburro, G. (2017). MATLAB for Machine Learning. Packt Publishing Ltd. Packt Publishing Ltd.

Coley, R. L., Votruba-Drzal, E., Miller, P. L., & Koury, A. (2013). Timing, extent, and type of child care and children's behavioral functioning in kindergarten. *Developmental Psychology, 49*(10), 1859–1873. https://doi.org/10.1037/a0031251

Connell, A. M., & Goodman, S. H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: A meta-analysis. *Psychological Bulletin, 128*(5), 746–773. https://doi.org/10.1037/0033-2909.128.5.746

Cordeiro, C., Machás, A., & Neves, M. M. (2010). A case study of a customer satisfaction problem: Bootstrap and imputation techniques. In *Handbook of partial least squares* (pp. 279-287). Springer, Berlin, Heidelberg.

Costa, P.T., Jr., & McCrae, R.R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psy- chological Assessment Resources.

Costa, P.T., Jr., & McCrae, R.R. (1989). *The NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Costa, P.T., Jr., & McCrae, R.R. (2008). The revised NEO personality inventory (NEO-PI-R). The SAGE Handbook of Personality Theory and Assessment. 2. 179-198. 10.4135/9781849200479.n9.

Cox, M. J., Mills-Koonce, R., Propper, C., & Gariépy, J. L. (2010). Systems theory and cascades in developmental psychopathology. *Development and Psychopathology*, *22*(3), 497-506.

Deans, C. L. (2020). Maternal sensitivity, its relationship with child outcomes, and interventions that address it: a systematic literature review. *Early Child Development and Care*, *190*(2), 252-275.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: group and individual differences. *Development and psychopathology*, *10*(3), 469–493. https://doi.org/10.1017/s0954579498001709

De Wolff, M. S., & Van Ijzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. Child development, 68(4), 571-591.

Doan, S. N., Fuller-Rowell, T. E., & Evans, G. W. (2012). Cumulative risk and adolescent's internalizing and externalizing problems: The mediating roles of maternal responsiveness and self-regulation. *Developmental Psychology, 48*(6), 1529–1539. https://doi.org/10.1037/a0027815

Egeland, B., & Heister, M. (1993). Teaching task rating scales. Institute of Child Development, University of Minnesota.

Eskelson, Bianca N. I., Hailemariam Temesgen, Valerie Lemay, Tara M. Barrett, Nicholas L. Crookston, and Andrew T. Hudak. 2009. “The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases.”

Scandinavian Journal of Forest Research 24 (3). Taylor & Francis: 235–46. doi:10.1080/02827580902870490.

Fearon, R.P., Bakermans-Kranenburg, M.J., van Ijzendoorn, M.H., Lapsley, A.M., Roisman, G.I. (2010).The significance of insecure attachment and disorganization in the development of children's externalizing behavior: a meta-analytic study. Child Development, 81(2), 435-56.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd edition). Sage Publications Ltd., London.

Fraley, R. C., & Roisman, G. I. (2015). Do early caregiving experiences leave an enduring or transient mark on developmental adaptation?. *Current Opinion in Psychology*, *1*, 101-106.

Gómez, D., & Rojas, A. (2016). An empirical overview of the no free lunch theorem and its effect on real-world Machine Learning classification. *Neural computation*, *28*(1), 216-228.

Groh, A.M., Roisman, G.I., van Ijzendoorn, M.H., Bakermans-Kranenburg, M.J., & Fearon, R.P. (2012). The significance of insecure and disorganized attachment for children's internalizing symptoms: a meta-analytic study. Child Development, 83(2), 591-610.

Halkola, R. (2020). Predicting mental health problems in adolescence using Machine Learning techniques. *PloS one*, *15*(4), e0230389.

Haltigan, J.D., Roisman, G.I., & Fraley, R.C. (2013). The predictive significance of early caregiving experiences for symptoms of psychopathology through midadolescence: enduring or transient effects? Developmental Psychopathology, 25(1), 209-21.

Heberle, A. E., Krill, S. C., Briggs-Gowan, M. J., & Carter, A. S. (2015). Predicting externalizing and internalizing behavior in kindergarten: Examining the buffering role of early social support. *Journal of Clinical Child and Adolescent Psychology, 44*(4), 640–654. https://doi.org/10.1080/15374416.2014.886254

Hofstra, M. B., & Verhulst, F. C. (2000). Continuity and change of psychopathology from childhood into adulthood: a 14-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(7), 850-858.

Hofstra, M. B., Van Der Ende, J. A. N., & Verhulst, F. C. (2001). Adolescents' self-reported problems as predictors of psychopathology in adulthood: 10-year follow-up study. *The British Journal of Psychiatry*, *179*(3), 203-209.

Hofstra, M. B., Van der Ende, J., & Verhulst, F. C. (2002). Pathways of self-reported problem behaviors from adolescence into adulthood. *American Journal of Psychiatry*, *159*(3), 401-407.

Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of Performance of Data Imputation Methods for Numeric Dataset. *Applied Artificial Intelligence*, *33*(10), 913-933.

Khaire, U. M., & Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. Journal of King Saud University-Computer and Information Sciences.

Kjeldsen, A., Janson, H., Stoolmiller, M., Torgersen, L., & Mathiesen, K. S. (2014). Externalising behaviour from infancy to mid-adolescence: Latent profiles and early predictors. *Journal of Applied Developmental Psychology*, *35*(1), 25-34.

Kim, H. K., Capaldi, D. M., & Stoolmiller, M. (2003). Depressive symptoms across adolescence and young adulthood in men: Predictions from parental and contextual risk factors. *Development and Psychopathology*, *15*(2), 469-495.

Kirchebner, J., Sonnweber, M., Nater, U. M., Günther, M., & Lau, S. (2020). Stress, Schizophrenia, and Violence: A Machine Learning Approach. *Journal of interpersonal violence*, 0886260520913641.

Kok, R., Linting, M., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Jaddoe, V. W., Hofman, A., ... & Tiemeier, H. (2013). Maternal sensitivity and internalizing problems: Evidence from two longitudinal studies in early childhood. *Child Psychiatry & Human Development*, *44*(6), 751-765.

Kontopantelis, E., White, I. R., Sperrin, M., & Buchan, I. (2017). Outcome-sensitive multiple imputation: A simulation study. BMC Medical Research Methodology, 17(1). https://doi.org/10.1186/s12874-016-0281-5

Kowarik, A. & Templ, M.(2016). Imputation with R package VIM. *Journal of Statistical Software*, *74*(7), 1-16.

Lansford, J. E., Malone, P. S., Tapanya, S., Tirado, L. M. U., Zelli, A., Alampay, L. P., ... & Deater-Deckard, K. (2019). Household income predicts trajectories of child internalizing and externalizing behavior in high-, middle-, and low-income countries. *International journal of behavioral development*, *43*(1), 74-79.

Lansford, J. E., Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Ethnic differences in the link between physical discipline and later adolescent externalizing behaviors. *Journal of Child Psychology and Psychiatry*, *45*(4), 801-812.

Leve, L. D., Kim, H. K., & Pears, K. C. (2005). Childhood temperament and family environment as predictors of internalizing and externalizing trajectories from ages 5 to 17. *Journal of abnormal child psychology*, *33*(5), 505-520.

Lorber, M.F. & Egeland, B. (2009). Infancy Parenting and Externalizing Psychopathology from Childhood through Adulthood: Developmental Trends. Developmental Psychology, 45(4), 909-912.

Madigan, S., Atkinson, L., Laurin, K., & Benoit, D. (2013). Attachment and internalizing behavior in early childhood: A meta-analysis. *Developmental Psychology, 49*(4), 672–689. https://doi.org/10.1037/a0028793

Madigan, S., Brumariu, L. E., Villani, V., Atkinson, L., & Lyons-Ruth, K. (2016). Representational and questionnaire measures of attachment: A meta-analysis of relations to child internalizing and externalizing problems. *Psychological Bulletin, 142*(4), 367–399. https://doi.org/10.1037/bul0000029.

Marshall, N.L., & Barnett, R.C. (1993). Work-family strains and gains among two- earner couples. Journal of Community Psychology, 21, 64-78.

MathWorks. (2016a). *Introducing Machine Learning.* The MathWorks, Inc. E-books.

MathWorks. (2016b). *Applying Supervised Learning*. The MathWorks, Inc. E-books.

MathWorks. (2016c). *Getting Started with Machine Learning*. The MathWorks, Inc. E-books.

MathWorks. (2020a). *Deep Learning and Traditional Machine Learning: Choosing the Right Approach*. Retrieved from https://explore.mathworks.com/machine-learning-vs-deep-learning

MathWorks. (2020b). *Select data and validation for regression problem*. Retrieved from https://www.mathworks.com/help/stats/select-data-and-validation-for-regression-problem.html

MathWorks. (2020c). *Assess Model Performance in Regression Learner*. Retrieved from https://www.mathworks.com/help/stats/assess-model-performance-in-regression-learner.html

MathWorks. (2020d). *Introduction to Feature Selection*. Retrieved from https://www.mathworks.com/help/stats/feature-selection.html

MathWorks. (2020e). *Train Regression Model using Hyperparameter Optimization in Regression Learner App*. Retrieved from https://www.mathworks.com/help/stats/train-regression-model-using-hyperparameter-optimization-in-regression-learner-app.html

MathWorks. (2020f). Hyperparameter Optimization in Regression Learner App. Retrieved from https://www.mathworks.com/help/stats/hyperparameter-optimization-in-regression-learner-app.html

MathWorks. (2020g). Choose Regression Model Options. Retrieved from https://www.mathworks.com/help/stats/choose-regression-model-options.html

MathWorks. (2020h). *Dummy Variables*. Retrieved from

https://www.mathworks.com/help/stats/dummy-indicator-variables.html

MathWorks. (2020i). *fitrgp*. Retrieved from https://www.mathworks.com/help/stats/fitrgp.html#mw_3bed4be1-f064-43a9-9a8f-4f7abd6089df

MathWorks. (2020j). Understanding Support Vector Machine Regression. Retrieved from https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html

MathWorks. (2020k). Gaussian Process Regression Models. Retrieved from https://www.mathworks.com/help/stats/gaussian-process-regression-models.html

MathWorks. (2020l). Train Regression Models in Regression Learner app. Retrieved from https://www.mathworks.com/help/stats/train-regression-models-in-regression-learner-app.html

McCartney, K., Owen, M.T., Booth, C.L., Clarke-Stewart, A., & Vandell, D.L. (2004). Testing a maternal attachment model of behavior problems in early childhood. Journal of Child Psychology and Psychiatry, 45, 765–778.

McCarty, C. A., & McMahon, R. J. (2003). Mediators of the relation between maternal depressive symptoms and child internalizing and disruptive behavior disorders. *Journal of family psychology*, *17*(4), 545.

McLaughlin, K. A., Hilt, L. M., & Nolen-Hoeksema, S. (2007). Racial/ethnic differences in internalizing and externalizing symptoms in adolescents. *Journal of abnormal child psychology*, *35*(5), 801-816.

McLoyd, V.C. (1998). Socioeconomic disadvantage and child development. Am Psychol., 53(2),185-204.

Meng, X.-L. (1994). Multiple-Imputation Inferences with Uncongenial Sources of Input. Statistical Science, 9(4), 538–558.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Kindergarten Predictors of Recurring Externalizing and Internalizing Psychopathology in the Third and Fifth Grades. *Journal of Emotional and Behavioral Disorders*, *17*(2), 67–79. https://doi.org/10.1177/1063426608324724

Mueller, J. P., Massaron, L. (2016). *Machine Learning for Dummies*. Hoboken, New Jersey: John Wiley & Sons Inc.

Na, K. S., Geem, Z. W., & Cho, S. E. (2020). Machine learning-based prediction of persistent oppositional defiant behavior for 5 years. *Nordic Journal of Psychiatry*, 1-6.

Narusyte, J., Ropponen, A., Alexanderson, K., & Svedberg, P. (2017). Internalizing and externalizing problems in childhood and adolescence as predictors of work incapacity in young adulthood. *Social psychiatry and psychiatric epidemiology*, *52*(9), 1159-1168.

Newman, L., Sivaratnam, C., & Komiti, A. (2015). Attachment and early brain development–neuroprotective interventions in infant–caregiver therapy. *Translational Developmental Psychiatry*, *3*(1), 28647.

NICHD Early Child Care Research Network (2001). Nonmaternal care and family factors in early development: An overview of the NICHD study of early child care. Journal of Applied Developmental Psychology, 22, 457-492.

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten?. *Child development*, *74*(4), 976-1005.

NICHD Early Child Care Research Network. (2005a). Childcare and child development: results from the NICHD study of early childcare and youth development. Guilford press.

NICHD Early Child Care Research Network. (2005b). Early childcare and children's development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *42*(3), 537-570.

O'Brien R. M. (2012). Visualizing rank deficient models: a row equation geometry of rank deficient matrices and constrained-regression. PloS one, 7(6), e38923. https://doi.org/10.1371/journal.pone.0038923

O'Connor, E. E., Collins, B. A., & Supplee, L. (2012). Behavior problems in late childhood: The roles of early maternal attachment and teacher–child relationship trajectories. *Attachment & Human Development*, *14*(3), 265-288.

Ogundepo, E. A., & Fokoué, E. (2019). An empirical demonstration of the no free lunch theorem. *Math. Appl*, 8, 173-188.

Orme, J.G., Reis, J., & Herz, E.J. (1986). Factorial and discriminant validity of the Center for Epidemiological Studies Depression (CES-D) Scale. Journal of Clinical Psychology, 42, 28-33.

Orri, M., Tremblay, R. E., Japel, C., Boivin, M., Vitaro, F., Losier, T., ... & Côté, S. M. (2019). Early childhood child care and disruptive behavior problems during adolescence: a 17-year population-based propensity score study. Journal of child psychology and psychiatry, 60(11), 1174-1182.

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology*, *10*, 2970.

Palmer, F. B., Graff, J. C., Jones, T. L., Murphy, L. E., Keisling, B. L., Whitaker, T. M., Wang, L., & Tylavsky, F. A. (2018). Socio-demographic, maternal, and child indicators of socioemotional problems in 2-year-old children: A cohort study. *Medicine*, *97*(28), e11468.

Pluess, M., & Belsky, J. (2009). Differential susceptibility to rearing experience: The case of childcare. Journal of Child Psychology and Psychiatry and Allied Disciplines, 50(4), 396–404.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1, 385-401.

Rasmussen, C. E. and Williams, C. K. (2006). Gaussian processes for Machine Learning, volume 1. MIT press Cambridge.

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: results of a 24-year longitudinal study. *Social psychiatry and psychiatric epidemiology*, *46*(12), 1233-1241.

Rudolph, K. D., Lansford, J. E., & Rodkin, P. C. (2016). Interpersonal theories of developmental psychopathology. *Developmental psychopathology*, 1-69.

Roberts, R.E. (1980). Reliability of the CES-D scale in different ethnic contexts. Psychiatry Research, 2, 125-134.

Roberts, R.E., & Vernon, S.W. (1983). The Center for Epidemiological Studies Depression Scale: Its use in a community sample. American Journal of Psychiatry, 140, 41- 46.

Rumsey, D. J. (2016). *Statistics for Dummies (2nd Ed.)*. John Wiley & Sons, Inc., New York.

Schaefer, M.T. & Olson, D.H. (1981). Assessing Intimacy: The Pair Inventory. Journal of Marital and Family Therapy, 7(l), 640-653.

Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. Psychological medicine, 49(9), 1426-1448.

Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using Machine Learning techniques. PloS one, 15(4), e0230389.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics: International edition. *Pearson2012*.

Vaez, E., Indran, R., Abdollahi, A., Juhari, R., & Mansor, M. (2015). How marital relations affect child behavior: review of recent research. Vulnerable Children and Youth Studies, 10(4), 321-336.

Van Ijzendoorn, M. H., Vereijken, C. M., Bakermans-Kranenburg, M. J., & Marianne Riksen-Walraven, J. (2004). Assessing attachment security with the attachment Q sort: Meta-analytic evidence for the validity of the observer AQS. *Child development*, *75*(4), 1188-1213.

Vandell, D.L., Belsky, J., Burchinal, M., Vandergrift, N. & Steinberg, L. (2010). Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development. Child Development, 81(3), 737–756.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. Child development, 81(3), 737-756.

Vandell, D.L., Burchinal, M., & Pierce, K.M. (2016). Early Child Care and Adolescent Functioning at the End of High School: Results from the NICHD Study of Early Child Care and Youth Development. Developmental Psychology, 52(10), 1634–1645.

Votruba-Drzal, E., Levine Coley, R., & Lindsay Chase-Lansdale, P. (2004). Child care and low-income children's development: Direct and moderated effects. Child development, 75(1), 296-312.

Wang, F. L., Eisenberg, N., Valiente, C., & Spinrad, T. L. (2016). Role of temperament in early adolescent pure and co-occurring internalizing and externalizing problems using a bifactor model: Moderation by parenting and gender. *Development and psychopathology*, *28*(4 Pt 2), 1487.

Wang, C., Williams, K. E., Shahaeian, A., & Harrison, L. J. (2018). Early predictors of escalating internalizing problems across middle childhood. *School Psychology Quarterly, 33*(2), 200–212. https://doi.org/10.1037/spq0000218

Watamura, S.E., Phillips, D.A., Morrissey, T.W., McCartney, K., Bub, K. (2011). Double jeopardy: poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child-care environments that confer risk. Child Dev., 82(1), 48-65.

Waters, E., & Deane, K. E. (1985). Defining and assessing individual differences in attachment relationships: Q-methodology and the organization of behavior in infancy and early childhood. *Monographs of the Society for Research in Child Development*, *50*(1/2), 41–65.

Wipf, D. P., & Nagarajan, S. S. (2008). A new view of automatic relevance determination. In *Advances in neural information processing systems* (pp. 1625-1632).

Wolpert DH. The supervised learning no-free-lunch theorems. Soft computing and industry: Springer; 2002. p. 25–42.

Womack, S. R., Taraban, L., Shaw, D. S., Wilson, M. N., & Dishion, T. J. (2019). Family turbulence and child internalizing and externalising behaviors: Moderation of effects by race. *Child development*, *90*(6), e729-e744.

World Health Organization. (23 October 2019). Adolescent Mental Health. Retrieved from https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of child- hood and adolescence: Gender and psychopathology. Annual Review of Clinical Psychology, 4, 275–303.

Part 3: Critical Appraisal

Introduction

In this critical appraisal, I will be sharing my reflections on the work that I have undertaken for my empirical paper. I will start with explaining the theoretical perspectives and personal preferences that influenced my choice of topic. I will proceed with reflecting on how the research challenged my expectations as a researcher and the learning I have achieved as a result of it. I will proceed with discussing the most important issues and challenges with respect to the use of a Machine Learning approach to analysing data and the recommendations that were made to improve its implementation. I will conclude by sharing my views with respect to the application of Machine Learning to the field of mental health.

My empirical project consisted of implementing a secondary analysis based on Machine Learning using data from the NICHD SECCYD study, which examined early environmental (interpersonal and contextual) as well as child characteristics in relation to various developmental outcomes longitudinally.

Choice of Topic

I was drawn to this topic as a result of my interest in psychodynamic theory. From a psychodynamic perspective, early parent-child (or alternative primary caregiver-child) interactions influence the development of emotional regulation strategies and ultimately later mental health. One of the most important developments in the literature for psychodynamic theory is the advent of attachment theory, which enabled the empirical examination of the role of early caregiving in psychopathology. I was particularly interested in examining the contribution of early interpersonal experiences to the development of psychopathology, specifically the role of early attachment and maternal sensitivity. I believed that shedding light on the nature of the relationship between these early experiences and later mental health, is valuable as it can inform the development of early intervention and prevention strategies.

I was also drawn to this topic because it involved learning a novel technique for analysing data. At the time, I was under the impression that this unorthodox analytic approach is superior to more traditional approaches, due to its ability to examine a large number of predictors simultaneously and its primary focus on predictive accuracy (Yarkoni & Westfall, 2017).

Finally, being able to analyse data from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD) study seemed to me as a valuable opportunity, considering how difficult it is to collect and have access to good-quality longitudinal data.

Learned Insights

The work on this research challenged some of the assumptions I had regarding the research topic and the analytic method used. When I first embarked on this project, I assumed that early interpersonal experiences, particularly early experiences with parents (or other primary caregivers) would be much more influential on the development of later psychopathology.

However, the findings that I obtained challenged my expectations and brought to my attention issues that I so far was not fully aware of.

The predictive ability of the trained models was low, which was slightly surprising for me. However, this finding was not inconsistent with the literature highlighting the role of later experiences in the emergence of mental health problems. I realized that my interest in psychodynamic theory may have biased my views and made me initially disregard literature highlighting the influence of later experiences on mental health outcomes. Furthermore, some of the trained models showed almost similar performance, which made me less confident about interpreting the results and selecting the best model. My attention was drawn to the potential dangers of implementing Machine Learning techniques in an uninformed way, making interpretations by relying crudely on indicators of model performance, such as the Root Mean Square Error (RMSE).

Strengths, Weakness and Future Recommendations

Readily available and widely acknowledged guidelines about how to implement Machine Learning in a way that is theory-driven and systematic are lacking. Cearns et al. (2019) suggested a set of criteria for evaluating the readiness of a Machine Learning model to be translated clinically or its maturity, encompassing (a) generalization, (b) model scope, and (c) incremental utility. Generalization refers to the extent to which the chosen model will generalize to new datasets. Model scope refers to the type of population that the model can make predictions about. This is important because the model's utility can only be estimated for samples similar to those on which the model was initially trained (and tested). Incremental utility refers to the extent to which a Machine Learning model improves current practice if clinically implemented.

I will use most of the remaining section of the critical appraisal to evaluate the soundness of the Machine Learning analysis that I have undertaken in my empirical paper according to the above-mentioned criteria proposed by Cearns et al. (2019). With respect to the first criterion, the generalizability of the model obtained is questionable due to the limited sample size used and potential data leakage (Cearns et al., 2019; Dwyer et al., 2018). Small samples are more likely to be homogeneous and, if used for model training and testing, there is the risk that the generalizability of the model to more heterogeneous real-world data is compromised.

Data leakage occurs when the test data become contaminated with information from the training data (Cearns et al., 2019).

Data leakage may have occurred due to the implementation of variable standardization and missing values imputation using the entire dataset, whereby the data to be included in the testing sets is no longer independent from the data selected for training. Data leakage compromises the soundness of the validation/testing process and may lead to overly optimistic estimation of the generalizability of the model to new data. One way to avoid data leakage is to use a nested cross-validation scheme, whereby certain data pre-processing procedures (e.g., standardizations, imputation, transformations, hyper-parameter optimization, etc.) are restricted to training samples. This is useful when the training and testing sets used are not adequately large, as it warns the creator of the model of the risk of overfitting. However, it's not always reliable, as there is the risk of having testing partitions that are similar to the training ones especially in the case of small samples, which may bias the cross-validation metrics and fail to warn against overfitting (Cearns et al., 2019).

Ideally, the best way to enhance generalizability and avoid data leakage is to use large samples, which is not always easy to achieve (Bzdok & Meyer-Lindenberg, 2018). Unlike small samples, large samples are heterogeneous enough, and are more suitable for model building, as it enhances the likelihood that the built model will be generalizable to new data. Moreover, big samples allow for the use of the hold-out validation scheme, which involves leaving a portion of the dataset out for testing, making sure that the training and testing procedures are based on completely independent partitions of the data. It is even good practice to hold the testing set inaccessible during the training phase to avoid any possible data leakage (Cearns et al., 2019; Dwyer et al,. 2018; Yarkoni & Westfall, 2017). However, with respect to my project, it was not possible to use the nested cross-validation scheme and the sample size was not large enough for the use of the hold-out validation scheme. This brought my attention to the dangers of relying too much on automated Machine Learning applications, which may give the impression that there is a one-size-fits all protocol for implementing Machine Learning for analysing data. Careful considerations of the characteristics of the available data is needed to determine the most suitable way of performing the analysis, which may require manual coding. However, one can't undermine the value of using the automated Machine Learning app in MATLAB, as it enabled the training of a wide array of models simultaneously in less than a few minutes, which is very valuable considering that manual coding may be very challenging and time consuming especially to the novice programmer.

Regularization is another way of minimizing the risk of overfitting and as a result enhancing generalizability, other than using large samples (Yarkoni & Westfall, 2017). During the learning process, regularization prevents the model from becoming unnecessarily complex, by pruning away redundant features. In the current study, it was not possible to collect more data to increase the sample size and ultimately the generalizability of the trained model. However, the use of Automatic Relevance Determination (ARD), which is a regularization method may, have provided some protection against overfitting, which I consider as a strength.

With respect to model scope, although the data used from the NICHD SECCYD study is not nationally representative, it is known for its diversity in terms of various characteristics such as demographics, socioeconomic status, ethnicity, family structure, place of resident, child-care arrangements, etc. Thus, I am moderately confident that this model could generalize to a group of Americans recruited from the community. Yet, its ability to make predictions about Americans with mental health disorders recruited from clinical/hospital settings, and certainly to contexts outside America is questionable (Cearns et al., 2019).

Regarding incremental utility, the current study extends previous research on the prediction of behaviour problems by enabling the simultaneous examination of a large number of variables as potential predictors. Moreover, the use of embedded feature selection to identify the most influential predictors enhanced the interpretability of the selected model and contributed theoretical insight into the potential early antecedents of mental health difficulties later in life. In addition, this study provided some empirical support for the use of Machine Learning to make sound predictions about mental health outcomes. Such proof-of-concept studies are needed initially to pave the way for research aimed at examining the utility of the Machine Learning models clinically (Cearns et al., 2019).

Conclusion

Although the use of Machine Learning in mental health poses its own challenges, it is feasible and seems be a promising development in this field. Reading about the potential uses of Machine Learning models in psychology and psychiatry has further spurred my interest in this approach. Yarkoni & Westfall (2017) strongly argued for the benefits of psychology becoming more focused on making accurate predictions about behaviour and not only explaining it. Machine Learning predictive modeling's primary focus on making accurate predictions about behaviour was deemed as a valuable addition to the field of mental health, especially that the generalizability of the findings from experimental research has been questionable due to the well-documented problem of replication (Bzdok & Meyer-Lindenberg, 2018; Shatte et al., 2019; Yarkoni & Westfall, 2017). I am a firm believer in eclecticism, and I think that predicting behaviour accurately as well as understanding its causes are crucial for theory building and can only be achieved by using Machine Learning as well as traditional statistical techniques for analysing data. I also believe that the use of Machine Learning in mental health is very promising, particularly with respect to enhancing the accuracy of predictions made about the emergence of psychopathology and the response to treatment, independently of diagnostic labels that to some extent lack empirical support.

References

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 223-230.

Cearns, M., Hahn, T., & Baune, B. T. (2019). Recommendations and future directions for supervised Machine Learning in psychiatry. *Translational psychiatry*, *9*(1), 1-12.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual review of clinical psychology*, *14*, 91-118.

Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, *49*(9), 1426-1448.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.

Appendix A

Search Strategy

Web of Science

Search 1: TS=((mother NEAR/1 interaction*) or (sensitive NEAR/1 parenting) or (sensitive NEAR/1 care*) or "sensitivity of maternal care" or "sensitivity of mother's care" or (sensitive NEAR/1 interaction*) or (sensitive NEAR/1 behavio$r) or "sensitive parenting" or "parental sensitivity" or "sensitive child care*" or "sensitivity of child care" or "sensitive interaction*" or "sensitive behavio$r" or "maternal sensitivity" or "mother's sensitivity")

Databases= WOS, BCI, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC Timespan=All years

Search language=Auto

Search 2: TS=("cogniti* develop*" or "mental develop*" or "intellectual develop*" or intelligen* or IQ or "cogniti* abilit*" or "mental abilit*" or "intellectual abilit*" or "woodcock-johnson" or peabody or mullen or leiter or "humanics national infant-toddler assessment" or "early learning accomplishment profile" or "developmental profile" or "developmental observation checklist" or griffiths or battelle or wechsler or "differential ability scales" or kaufman or "wide range intelligence" or bayley or "stanford-binet" or mccarthy or raven or fagan)

Databases= WOS, BCI, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC Timespan=All years

Search language=Auto

Search 3: TS=(study or longitudinal or prospective or cohort or follow-up or experiment* or trial*).

Databases= WOS, BCI, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC Timespan=All years

Search language=Auto

Search 4: #3 AND #2 AND #1

Databases= WOS, BCI, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC Timespan=All years

Search language=Auto

Medline

1. ("mother- ADJ1 interaction*" or "sensitive ADJ1 parenting" or "sensitive ADJ1 care*" or "sensitivity of maternal care" or "sensitivity of mother's care" or "sensitive ADJ1 interaction*" or "sensitive ADJ1 behavio?r" or "sensitive parenting" or "parental sensitivity" or "sensitive child care*" or "sensitivity of child care" or "sensitive interaction*" or "sensitive behavio?r" or "maternal sensitivity" or "mother's sensitivity").mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
2. ("cogniti* develop*" or "mental develop*" or "intellectual develop*" or intelligen* or IQ or "cogniti* abilit*" or "mental abilit*" or "intellectual abilit*" or "woodcock-johnson" or peabody or mullen or leiter or "humanics national infant-toddler assessment" or "early learning accomplishment profile" or "developmental profile" or "developmental observation checklist" or griffiths or battelle or wechsler or "differential ability scales" or kaufman or "wide range intelligence" or bayley or "stanford-binet" or mccarthy or raven or fagan).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word,

keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

3. (study or longitudinal or prospective or cohort or follow-up or experiment* or trial*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

4. exp maternal behavior/ or exp mother-child relations/

5. exp Mental Competency/

6. exp Intelligence Tests/ or exp Intelligence/

7. exp Cohort Studies/

8. exp follow-up studies/ or exp longitudinal studies/ or exp prospective studies/

9. exp Randomized Controlled Trial/ or exp Clinical Trial/

10. 1 or 4

11. 2 or 5 or 6

12. 3 or 7 or 9

13. 10 and 11 and 12

14. limit 13 to animals

15. 13 not 14

16. limit 15 to pharmacologic actions

17. 15 not 16

PsycINFO

1. ("mother- ADJ1 interaction*" or "sensitive ADJ1 parenting" or "sensitive ADJ1 care*" or "sensitivity of maternal care" or "sensitivity of mother's care" or "sensitive ADJ1 interaction*" or "sensitive ADJ1 behavio?r" or "sensitive parenting" or "parental sensitivity" or "sensitive child care*" or "sensitivity of child care" or "sensitive interaction*" or "sensitive behavio?r" or "maternal sensitivity" or "mother's sensitivity").mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]

2. ("cogniti* develop*" or "mental develop*" or "intellectual develop*" or intelligen* or IQ or "cogniti* abilit*" or "mental abilit*" or "intellectual abilit*" or "woodcock-johnson" or peabody or mullen or leiter or "humanics national infant-toddler assessment" or "early learning accomplishment profile" or "developmental profile" or "developmental observation checklist" or griffiths or battelle or wechsler or "differential ability scales" or kaufman or "wide range intelligence" or bayley or "stanford-binet" or mccarthy or raven or fagan).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]

3. (study or longitudinal or prospective or cohort or follow-up or experiment* or trial*).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]

4. exp mother child relations/ or exp mother child communication/

5. exp cognitive development/ or exp intellectual development/

6. exp cognitive ability/

7. exp cognitive assessment/ or exp intelligence/ or exp intelligence quotient/

8. 2 or 5 or 6 or 7

9. 1 or 4

10. exp longitudinal studies/ or exp experimental design/ or exp prospective studies/ or exp followup studies/

11. exp experimental methods/

12. 3 or 10 or 11

13. 8 and 9 and 12

14. limit 13 to ("0120 non-peer-reviewed journal" or "0200 book" or "0240 authored book" or "0280 edited book" or "0300 encyclopedia" or "0400 dissertation abstract")

15. 13 not 14

16. limit 15 to animal

17. 15 not 16

Appendix B

National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Criterion	Description	Yes	No	Other (CD, NR, or NA)
Research question	1. Was the research question or objective in this paper clearly stated?	1	0	0
Study population	2. Was the study population clearly specified and defined?	1	0	0
Participation rate	3. Was the participation rate of eligible persons at least 50%?	1	0	0
Recruitment from same population and uniform eligibility criteria	4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	1	0	0
Sample size justification	5. Was a sample size justification, power description, or variance and effect estimates provided?	1	0	0

Exposure assessed prior to outcome measurement	6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	1	0	0
Sufficient timeframe to see an effect	7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	1	0	0
Different levels of the exposure of interest	8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?	1	0	0
Exposure measures and assessment	9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	1	0	0

Repeated exposure assessment	10. Was the exposure(s) assessed more than once over time?	1	0	0
Outcome measures	11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	1	0	0
Blinding of outcome assessors	12. Were the outcome assessors blinded to the exposure status of participants?	1	0	0
Followup rate	13. Was loss to follow-up after baseline 20% or less?	1	0	0
Statistical analyses	14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?	1	0	0

Note. CD, cannot determine; NA, not applicable; NR, not reported

Appendix C

R Code for the Quantitative Analysis

<u>Code for outlier identification</u>

```
# Install required packages

> install.packages(c("robumeta", "metafor", "dplyr", "foreign"))

# Load required packages

> library("robumeta")

> library("metafor")

> library("dplyr")

> library("foreign")

# Import excel dataset

> library(readxl)

> metadatanoout <- read_excel("metadatanoout.xlsx",

+                     col_types = c("numeric", "text", "text",

+                           "text", "numeric", "numeric", "numeric",

+                           "text", "numeric", "text", "numeric",

+                           "numeric"))

> View(metadatanoout)

# Fisher's z-transformation of the correlation coefficients

> metadatanoout <- escalc(measure="ZCOR", ri=r, ni=n, data=metadatanoout)

# Identify outliers with effect sizes that are 3 standard deviations from the mean of
effect sizes

> mean(metadatanoout$yi, na.rm = TRUE)

> sd(metadatanoout$yi, na.rm = TRUE)
```

```
> scale(metadatanoout$yi, center = TRUE, scale = TRUE)
```

Remove the outlier(s) manually from the dataset and use the new dataset for subsequent analyses.

<u>Code for meta-analysis</u>

```
# Import new excel dataset with the outlier(s) removed
> library(readxl)
> metadatanoout <- read_excel("metadatanoout.xlsx",
+                        col_types = c("numeric", "text", "text",
+                                 "text", "numeric", "numeric", "numeric",
+                                 "text", "numeric", "text", "numeric",
+                                 "numeric"))
> View(metadatanoout)
# Define each categorical variable as a factor and dummy code it
> metadatanoout$`MS Stimulation` <- as.factor(metadatanoout$`MS Stimulation`)
> MS.Stimulationdummies.matrix = model.matrix(~factor(metadatanoout$`MS
Stimulation`))
> metadatanoout$MS.Stimulationdummies = MS.Stimulationdummies.matrix
> MS.Stimulationdummies.frame = data.frame(MS.Stimulationdummies.matrix)
> metadatanoout$SES <- as.factor(metadatanoout$SES)
> SESdummies.matrix = model.matrix(~factor(metadatanoout$SES))
> metadatanoout$SESdummies = SESdummies.matrix
> SESdummies.frame = data.frame(SESdummies.matrix)
> metadatanoout = cbind(metadatanoout,MS.Stimulationdummies.frame)
> metadatanoout = cbind(metadatanoout,SESdummies.frame)
# Fisher’s z-transformation of the correlation coefficients
```

```
> metadatanoout <- escalc(measure="ZCOR", ri=r, ni=n, data=metadatanoout)
# Fit the intercept-only model
> metadatanoout_intercept <- robu(formula = yi ~ 1, data = metadatanoout,
studynum = StudyID, var.eff.size = vi, rho = .8, small = TRUE)
# Print (view) output of intercept-only model
> print(metadatanoout_intercept)
# Sensitivity analysis
> sensitivity(metadatanoout_intercept)
# Forest plot
> forest.robu(metadatanoout_intercept, es.lab = "es.lab", study.lab = "Author",
"Effect Size" = effect.size,"Weight" = r.weights)
# Fit a Random-Effect Model to perform Egger test
> res <- rma(yi, vi, data=metadatanoout)
# Classical Egger test
> regtest(res, model="lm")
# Random/mixed effects version of the Egger test
> regtest(res)
# Egger test using the sample size (or a transformation thereof) as predictor
> regtest(metadatanoout$yi, metadatanoout$vi, predictor="ni")
# Get the funnel Plot
> funnel(res, xlab = "Correlation coefficient")
# Fit a RVE Correlated Effects model with Small-Sample Corrections with age at
maternal sensitivity assessment as the only covariate (moderator)
> CorrModC1 <- robu(formula = yi ~ Age.MS.Ax, data = metadatanoout, studynum
= StudyID, var.eff.size = vi, modelweights = "CORR", rho = 0.8, small = TRUE)
```

```
> print(CorrModC1)
# Fit a RVE Correlated Effects model with Small-Sample Corrections with maternal sensitivity construct as the only covariate (moderator)
> CorrModC2 <- robu(formula = yi ~ MS.Stimulation, data = metadatanoout, studynum = StudyID, var.eff.size = vi, modelweights = "CORR", rho = 0.8, small = TRUE)
> print(CorrModC2)
# Fit a RVE Correlated Effects model with Small-Sample Corrections with SES as a covariate (moderator)
> CorrModC3 <- robu(formula = yi ~ SES, data = metadatanoout, studynum = StudyID, var.eff.size = vi, modelweights = "CORR", rho = 0.8, small = TRUE)
>  print(CorrModC3)
# Fit a RVE Correlated Effects meta-regression model with Small-Sample Corrections including all covariates (moderators)
> CorrModAll<- robu(formula = yi ~ Age.MS.Ax + MS.Stimulation + SES, data = metadatanoout, studynum = StudyID, var.eff.size = vi, modelweights = c("CORR"), rho = 0.8, small = TRUE)
> print(CorrModAll)
# Convert Fisher’s z into correlation coefficient r
> install.packages("esc")
> library("esc")
> convert_z2r(.313)
```

Appendix D

Comparison between Analysis and Attrition Samples in terms of Continuous Variables

Group Statistics

	CBCL	N	Mean	Std. Deviation	Std. Error Mean
Mother's Age	Missing	391	26.94	5.63	0.285
	Nonmissing	973	28.58	5.57	0.179
Mother's Education	Missing	390	13.69	2.582	0.131
	Nonmissing	973	14.45	2.452	0.079
Income-to-Need ratio	Missing	384	2.9015	2.81078	0.14344
	Nonmissing	971	3.5186	2.62366	0.0842
Single Parenthood Status	Missing	391	2.6905	1.56857	0.07933
	Nonmissing	973	3.408	1.2575	0.04031
Maternal Depression	Missing	390	10.1943	7.55058	0.38234
	Nonmissing	973	9.72	6.58484	0.2111
Marital/Partner Relationship Intimacy	Missing	357	5.0478	1.02724	0.05437
	Nonmissing	931	4.7734	0.93235	0.03056
Social Support	Missing	390	4.9935	0.69575	0.03523
	Nonmissing	973	4.9774	0.57636	0.01848
Psychological Adjustment	Missing	326	58.9565	14.73215	0.81594
	Nonmissing	946	59.0164	13.68022	0.44478

Parenting Stress	Missing	390	-0.0645	0.86458	0.04378
	Nonmissing	973	0.028	0.76434	0.0245
Maternal Sensitivity	Missing	339	-0.1548	0.81272	0.04414
	Nonmissing	967	0.0305	0.7526	0.0242
Non-family Child Care Hours	Missing	254	29.5437	14.03957	0.88092
	Nonmissing	781	27.959	14.83812	0.53095
Child Temperament (mother-report)	Missing	327	3.21	0.396	0.022
	Nonmissing	952	3.17	0.407	0.013
24-mo Attachment Q-Sort Security	Missing	270	0.27	0.207	0.013
	Nonmissing	927	0.3	0.206	0.007

Independent Sample Tests

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Mother's Age	Equal variances assumed	1.1	0.294	-4.898	1362	0	-1.638	0.335	-2.295	-0.982
	Equal variances not assumed			-4.875	712.832	0	-1.638	0.336	-2.298	-0.979
Mother's Education	Equal variances assumed	0.032	0.858	-5.11	1361	0	-0.762	0.149	-1.055	-0.47
	Equal variances not assumed			-4.998	685.166	0	-0.762	0.153	-1.062	-0.463
Income-to-Need ratio	Equal variances assumed	0.007	0.936	-3.823	1353	0	-0.61715	0.16144	-0.93384	-0.30046
	Equal variances not assumed			-3.711	661.406	0	-0.61715	0.16632	-0.94373	-0.29056
Single Parenthood Status	Equal variances assumed	97.674	0	-8.85	1362	0	-0.71748	0.08107	-0.87651	-0.55845
	Equal variances not assumed			-8.063	601.369	0	-0.71748	0.08898	-0.89223	-0.54273
Maternal Depression	Equal variances assumed	8.81	0.003	1.151	1361	0.25	0.47434	0.41202	-0.33392	1.28259
	Equal variances not assumed			1.086	638.572	0.278	0.47434	0.43674	-0.38329	1.33197
Marital/Partner Relationship Intimacy	Equal variances assumed	3.39	0.066	4.595	1286	0	0.27445	0.05973	0.15726	0.39164
	Equal variances not assumed			4.401	593.756	0	0.27445	0.06237	0.15197	0.39694
Social Support	Equal variances assumed	12.159	0.001	0.438	1361	0.661	0.0161	0.03673	-0.05596	0.08815
	Equal variances not assumed			0.405	613.844	0.686	0.0161	0.03978	-0.06203	0.09422
Psychological Adjustment	Equal variances assumed	0.825	0.364	-0.067	1270	0.947	-0.05988	0.89636	-1.81838	1.69862
	Equal variances not assumed			-0.064	530.73	0.949	-0.05988	0.92929	-1.88542	1.76567
Parenting Stress	Equal variances assumed	8.708	0.003	-1.943	1361	0.052	-0.09248	0.0476	-0.18586	0.00091
	Equal variances not assumed			-1.843	645.543	0.066	-0.09248	0.05017	-0.19099	0.00604
Maternal Sensitivity	Equal variances assumed	1.052	0.305	-3.819	1304	0	-0.18528	0.04852	-0.28046	-0.09011
	Equal variances not assumed			-3.681	554.24	0	-0.18528	0.05034	-0.28417	-0.0864

Non-family Child Care Hours	Equal variances assumed	4.98	0.026	1.498	1033	0.134	1.58475	1.05795	-0.49122	3.66072
	Equal variances not assumed			1.541	450.903	0.124	1.58475	1.02856	-0.43661	3.60611
Child Temperament (mother-report)	Equal variances assumed	1.65	0.199	1.602	1277	0.109	0.041	0.026	-0.009	0.092
	Equal variances not assumed			1.624	579.849	0.105	0.041	0.026	-0.009	0.092
24-mo Attachment Q-Sort Security	Equal variances assumed	0.006	0.937	-2.216	1195	0.027	-0.032	0.014	-0.059	-0.004
	Equal variances not assumed			-2.21	435.897	0.028	-0.032	0.014	-0.06	-0.003

Appendix E

Comparison between Analysis and Attrition Samples in terms of Categorical Variables

Cross-tabulations for Gender

			Missing CBCL	Not Missing CBCL	Total
Child's Gender	Male	Count	219	486	705
		Expected Count	202.1	502.9	705
	Female	Count	172	487	659
		Expected Count	188.9	470.1	659
Total		Count	391	973	1364
		Expected Count	391	973	1364

Chi-Square test for Gender

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.104[a]	1	0.043		
Continuity Correction[b]	3.865	1	0.049		
Likelihood Ratio	4.113	1	0.043		
Fisher's Exact Test				0.048	0.025
Linear-by-Linear Association	4.101	1	0.043		
N of Valid Cases	1364				

[a]0 cells (.0%) have expected count less than 5. The minimum expected count is 188.91.
[b]Computed only for a 2x2 table

Cross-tabulations for Child's Race

			Missing CBCL	Not Missing CBCL	Total
Child's Race	White	Count	303	794	1097
		Expected Count	314.5	782.5	1097
	Other	Count	88	179	267
		Expected Count	76.5	190.5	267
Total		Count	391	973	1364
		Expected Count	391	973	1364

Chi-Square for Child's Race

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.992[a]	1	0.084		
Continuity Correction[b]	2.737	1	0.098		
Likelihood Ratio	2.931	1	0.087		
Fisher's Exact Test				0.097	0.05
Linear-by-Linear Association	2.99	1	0.084		
N of Valid Cases	1364				

[a]0 cells (.0%) have expected count less than 5. The minimum expected count is 76.54.

[b]Computed only for a 2x2 table

Cross-tabulation for 15-mo Attachment

			Missing CBCL[a]	Not Missing CBCL[b]	Total
15-mo Attachment	A	Count	32	128	160
		Expected Count	36.3	123.7	160
	B	Count	158	552	710
		Expected Count	161	549	710
	C	Count	27	75	102
		Expected Count	23.1	78.9	102
	D & U	Count	53	166	219
		Expected Count	49.6	169.4	219
Total		Count	270	921	1191
		Expected Count	270	921	1191

Note. CBCL = Child Behavior Checklist

[a] Participants had missing scores on both externalizing and internalizing problems

[b] Participants had no missing scores on either externalizing or internalizing problems

Chi-Square for 15-mo Attachment

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1.854[a]	3	0.603
Likelihood Ratio	1.839	3	0.607
Linear-by-Linear Association	1.237	1	0.266
N of Valid Cases	1191		

[a]0 cells (.0%) have expected count less than 5. The minimum expected count is 23.12.

Cross-tabulations for 36-mo Attachment

			Missing CBCL[a]	Not Missing CBCL[b]	Total
36-mo Attachment	A - Avoidant	Count	15	40	55
		Expected Count	11.5	43.5	55
	B - Secure	Count	139	562	701
		Expected Count	147	554	701
	C - Ambivalent	Count	48	149	197
		Expected Count	41.3	155.7	197
	D - Insecure other/ Controlling	Count	37	150	187
		Expected Count	39.2	147.8	187
Total		Count	239	901	1140
		Expected Count	239	901	1140

Note. CBCL = Child Behavior Checklist
[a] Participants had missing scores on both externalizing and internalizing problems
[b] Participants had no missing scores on either externalizing or internalizing problems

Chi-Square of 36-mo Attachment

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.398[a]	3	0.334
Likelihood Ratio	3.274	3	0.351
Linear-by-Linear Association	0.011	1	0.917
N of Valid Cases	1140		

[a]0 cells (.0%) have expected count less than 5. The minimum expected count is 11.53.

Appendix F

Skewness and Kurtosis Statistics

	N		Skewness	Std. Error of Skewness	Z-score for Skewness	Kurtosis	Std. Error of Kurtosis	Z-score for Kurtosis
	Valid	Missing						
Mother's Age	1364.00	0.00	0.06	0.07	0.85	-0.63	0.13	-4.80
Mother's Education	1363.00	1.00	0.15	0.07	2.24	-0.03	0.13	-0.24
Income-to-Need Ratio	1355.00	9.00	2.02	0.07	30.58	5.90	0.13	44.35
Single Parenthood	1364.00	0.00	-1.38	0.07	-20.91	0.44	0.13	3.31
Maternal Depression	1363.00	1.00	1.25	0.07	18.89	1.81	0.13	13.68
Intimacy	1288.00	76.00	-0.28	0.07	-4.04	-0.03	0.14	-0.23
Social Support	1363.00	1.00	-0.98	0.07	-14.85	2.19	0.13	16.55
Psychological Adjustment	1272.00	92.00	-0.12	0.07	-1.67	-0.02	0.14	-0.18
Parenting Stress	1363.00	1.00	0.35	0.07	5.24	0.03	0.13	0.22
Maternal Sensitivity	1306.00	58.00	-0.85	0.07	-12.49	0.97	0.14	7.19
Non-Family Child Care Hours	1035.00	329.00	-0.22	0.08	-2.92	-1.14	0.15	-7.52
Child Temperament (mother's report)	1279.00	85.00	-0.13	0.07	-1.93	0.48	0.14	3.50
24-mo Attachment Q-Sort Security	1197.00	167.00	-0.55	0.07	-7.75	0.09	0.14	0.65
Internalizing Problem (mother/alternate caregiver's report)	973.00	391.00	0.33	0.08	4.28	-0.20	0.16	-1.28

Externalizing Problem (mother/alternate caregiver's report)	973.00	391.00	0.61	0.08	7.85	-0.01	0.16	-0.06

Appendix G

List of Outliers

Outliers	Case numbers
1	12
2	15
3	20
4	22
5	26
6	28
7	37
8	48
9	58
10	61
11	81
12	85
13	93
14	96
15	111
16	113
17	122
18	132
19	148
20	208
21	209
22	217
23	244
24	256
25	260
26	266
27	274
28	307
29	317
30	318
31	341
32	377
33	399

34	416
35	419
36	441
37	446
38	464
39	468
40	474
41	483
42	492
43	515
44	525
45	526
46	529
47	542
48	551
49	552
50	597
51	618
52	633
53	642
54	670
55	682
56	700
57	722
58	732
59	734
60	743
61	749
62	754
63	760
64	763
65	771
66	782
67	809
68	811
69	814
70	823
71	833
72	836
73	849

74	857
75	863
76	868
77	904
78	907
79	919
80	933
81	941
82	942
83	964
84	976
85	981
86	984
87	1007
88	1035
89	1056
90	1057
91	1068
92	1074
93	1076
94	1077
95	1081
96	1100
97	1102
98	1133
99	1146
100	1152
101	1171
102	1178
103	1190
104	1220
105	1227
106	1234
107	1239
108	1240
109	1256
110	1270
111	1273
112	1274
113	1280

114	1284
115	1285
116	1288
117	1317
118	1336
119	1341
120	1346
121	1348
122	1349

Appendix H

Little's MCAR test (outliers included)

EM Correlations[a]

	MAGE	MEDUC	INCND	SGPAR	MDEP	INTI	SOC	PSYADJ	PARSTR	MSENS	CHCA	TEMP	QSET	MBIN	MBEX
MAGE	1														
MEDUC	0.545	1													
INCND	0.472	0.537	1												
SGPAR	0.367	0.327	0.371	1											
MDEP	-0.278	-0.318	-0.285	-0.285	1										
INTI	-0.091	0.02	0.045	-0.183	-0.308	1									
SOC	0.032	0.101	0.133	0.147	-0.504	0.443	1								
PSYADJ	0.221	0.26	0.284	0.244	-0.622	0.28	0.489	1							
PARSTR	-0.022	-0.061	-0.113	-0.079	0.53	-0.365	-0.494	-0.535	1						
MSENS	0.415	0.501	0.416	0.388	-0.355	0.052	0.167	0.347	-0.136	1					
CHCA	-0.117	-0.115	-0.061	-0.198	0.038	0.009	-0.046	-0.063	-0.064	-0.21	1				
TEMP	-0.178	-0.141	-0.146	-0.138	0.253	-0.041	-0.16	-0.25	0.236	-0.21	-0.029	1			
QSET	0.149	0.172	0.14	0.164	-0.189	-0.002	0.123	0.187	-0.129	0.291	-0.054	-0.127	1		
MBIN	-0.123	-0.087	-0.065	-0.061	0.279	-0.136	-0.179	-0.244	0.219	-0.038	-0.048	0.132	-0.093	1	
MBEX	-0.236	-0.237	-0.183	-0.125	0.293	-0.196	-0.257	-0.269	0.231	-0.159	0.04	0.142	-0.108	0.6	1

Note. MAGE = Maternal Age; MEDUC = Maternal Education; INCND = Income-to-Need Ratio; SGPAR = Single Parenthood Status; MDEP = Maternal Depression; INTI = Marital/Partner Relationship Intimacy; SOC = Social Support; PSYACJ = Maternal Psychological Adjustment; PARSTR = Parenting Stress; MSENS = Maternal Sensitivity; CHCA = Nonfamily Child Care Hours; TEMP = Child's Temperament (mother-report); QSET = 24-month Attachment Q-Sort Security; MBIN = Internalizing Problems (mother-report); MBEX = Externalizing Problems (mother-report).

[a] Little's MCAR test: Chi-Square = 1385.896, DF = 474, Sig. = .000

Little’s MCAR Test (excluding outliers)

	PARSTR	MSENS	ZMAGE	ZMEDUC	ZINCND	ZSGPAR	ZMDEP	ZINTI	ZSOC	ZPSYADJ	ZCHCA	ZTEMP	ZQSET	ZMBIN	ZMBEX
PARSTR	1														
MSENS	-0.105	1													
ZMAGE	0.004	0.384	1												
ZMEDUC	-0.039	0.499	0.526	1											
ZINCND	-0.083	0.435	0.465	0.511	1										
ZSGPAR	-0.052	0.352	0.344	0.299	0.374	1									
ZMDEP	0.52	-0.324	-0.256	-0.296	-0.269	-0.262	1								
ZINTI	-0.329	0.054	-0.099	0.017	0.042	-0.204	-0.3	1							
ZSOC	-0.488	0.134	0.011	0.077	0.118	0.108	-0.471	0.435	1						
ZPSYADJ	-0.517	0.331	0.204	0.242	0.271	0.203	-0.592	0.274	0.459	1					
ZCHCA	-0.06	-0.21	-0.122	-0.129	-0.082	-0.211	0.049	0.004	-0.041	-0.068	1				
ZTEMP	0.247	-0.184	-0.147	-0.116	-0.134	-0.096	0.243	-0.089	-0.182	-0.276	-0.031	1			
ZQSET	-0.131	0.285	0.129	0.171	0.148	0.145	-0.188	-0.003	0.106	0.182	-0.066	-0.12	1		
ZMBIN	0.208	-0.049	-0.107	-0.065	-0.05	-0.031	0.269	-0.137	-0.171	-0.225	-0.048	0.138	-0.12	1	
ZMBEX	0.205	-0.16	-0.236	-0.226	-0.186	-0.099	0.279	-0.169	-0.23	-0.258	0.035	0.144	-0.124	0.579	1

Note. PARSTR = Parenting Stress; MSENS = Maternal Sensitivity; MAGE = Maternal Age; MEDUC = Maternal Education; INCND = Income-to-Need Ratio; SGPAR = Single Parenthood Status; MDEP = Maternal Depression; INTI = Marital/Partner Relationship Intimacy; SOC = Social Support; PSYACJ = Maternal Psychological Adjustment; CHCA = Nonfamily Child Care Hours; TEMP = Child’s Temperament (mother-report); QSET = 24-month Attachment Q-Sort Security; MBIN = Internalizing Problems (mother-report); MBEX = Externalizing Problems (mother-report).

[a] Little's MCAR test: Chi-Square = 1301.811, DF = 462, Sig. = .000

Appendix I

R Code for k-Nearest Neighbor Imputation

<u>R Code for missing values imputation for the dataset including the outliers</u>

```
# Import the comma-separated values (CSV) data file using the "Import Dataset" dropdown from the "Environment" pane
# Install VIM R package
> install.packages("VIM")

# Load the VIM R package
> library(VIM)

# Define each categorical variable as a factor
> DatasetIDR$ATT36 <- as.factor(DatasetIDR$ATT36)
> DatasetIDR$ATT15 <- as.factor(DatasetIDR$ATT15)
> DatasetIDR$CRACE <- as.factor(DatasetIDR$CRACE)
> DatasetIDR$SEX <- as.factor(DatasetIDR$SEX)

# Impute missing values using the k-nearest neighbor algorithm
> DatasetIMP <- kNN(DatasetIDR)

# Get a summary of the descriptives of the imputed data
> summary(DatasetIMP)

# Save the imputed dataset as a CSV file in the working directory
> write.table(DatasetIMP, file = "expfile.csv", sep=",")
> write.csv(DatasetIMP,'DatasetIMP.csv')
```

<u>R Code for missing values imputation for the dataset excluding the outliers</u>

```
# Import the comma-separated values (CSV) data file using the "Import Dataset" dropdown from the "Environment" pane
# Install VIM R package
> install.packages("VIM")

# Load the VIM R package
> library(VIM)

# Define each categorical variable as a factor
> DatasetIMPO$ATT36 <- as.factor(DatasetIMPO$ATT36)
> DatasetIMPO$ATT15 <- as.factor(DatasetIMPO$ATT15)
> DatasetIMPO$CRACE <- as.factor(DatasetIMPO$CRACE)
> DatasetIMPO$SEX <- as.factor(DatasetIMPO$SEX)

# Impute missing values using the k-nearest neighbor algorithm
> DatasetIMPOR <- kNN(DatasetIMPO)
```

```
# Get a summary of the descriptives of the imputed data
> summary(DatasetIMPOR)

# Save the imputed dataset as a CSV file in the working directory
> write.table(DatasetIMPOR, file = "expfile.csv", sep=",")
> write.csv(DatasetIMPOR,' DatasetIMPOR.csv')
```

Appendix J

Embedded Feature Selection MATLAB Code

For the prediction of Externalizing Problems (including outliers)

```
% Extract predictors and response
% This code processes the data into the right shape for training the model.
inputTable = DatasetIMP;
predictorNames = {'SEX', 'CRACE', 'ATT15', 'ATT36', 'PARSTR', 'MSENS',
'ZMAGE', 'ZMEDUC', 'ZINCND', 'ZSGPAR', 'ZMDEP', 'ZINTI', 'ZSOC',
'ZPSYADJ', 'ZCHCA', 'ZTEMP', 'ZQSET'};
predictors = inputTable(:, predictorNames);
response = inputTable.ZMBEX;
isCategoricalPredictor = [true, true, true, true, false, false, false, false, false, false,
false, false, false, false, false, false, false];
% Train a regression model
% This code specifies all the model options and trains the model.
regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'ardexponential', ...
    'Standardize', true);
% Find the predictor weights by taking the exponential of the negative learned length
scales. Normalize the weights.
>> sigmaL = regressionGP.KernelInformation.KernelParameters(1:end-1);
weights = exp(-sigmaL);
weights = weights/sum(weights);
% Plot the normalized predictor weights
>> figure;
semilogx(weights,'ro');
xlabel('Predictor index');
ylabel('Predictor weight');
```

For the prediction of Externalizing Problems (excluding outliers)

```
% Extract predictors and response
% This code processes the data into the right shape for training the model.
inputTable = DatasetIMPOR;
predictorNames = {'SEX', 'CRACE', 'ATT15', 'ATT36', 'PARSTR', 'MSENS',
'ZMAGE', 'ZMEDUC', 'ZINCND', 'ZSGPAR', 'ZMDEP', 'ZINTI', 'ZSOC',
'ZPSYADJ', 'ZCHCA', 'ZTEMP', 'ZQSET'};
predictors = inputTable(:, predictorNames);
response = inputTable.ZMBEX;
isCategoricalPredictor = [true, true, true, true, false, false, false, false, false, false,
false, false, false, false, false, false, false];
```

```
% Train a regression model
% This code specifies all the model options and trains the model.
regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'ardexponential', ...
    'Standardize', true);
% Find the predictor weights by taking the exponential of the negative learned length
scales. Normalize the weights.
>> sigmaL = regressionGP.KernelInformation.KernelParameters(1:end-1);
weights = exp(-sigmaL);
weights = weights/sum(weights);
% Plot the normalized predictor weights
>> figure;
semilogx(weights,'ro');
xlabel('Predictor index');
ylabel('Predictor weight');
```

For the prediction of Internalizing Problems (including outliers)

```
% Extract predictors and response. This code processes the data into the right shape
for training the model.
>> inputTable = DatasetIMP;
predictorNames = {'SEX', 'CRACE', 'ATT15', 'ATT36', 'PARSTR', 'MSENS',
'ZMAGE', 'ZMEDUC', 'ZINCND', 'ZSGPAR', 'ZMDEP', 'ZINTI', 'ZSOC',
'ZPSYADJ', 'ZCHCA', 'ZTEMP', 'ZQSET'};
predictors = inputTable(:, predictorNames);
response = inputTable.ZMBIN;
isCategoricalPredictor = [true, true, true, true, false, false, false, false, false, false,
false, false, false, false, false, false, false];
% Train a regression model. This code specifies all the model options and trains the
model.
>> regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'ardexponential', ...
     'Standardize', true);
% Find the predictor weights by taking the exponential of the negative learned length
scales. Normalize the weights.
>> sigmaL = regressionGP.KernelInformation.KernelParameters(1:end-1);
weights = exp(-sigmaL);
weights = weights/sum(weights);
Unable to resolve the name regressionGP.KernelInformation.KernelParameters.
 % Plot the normalized predictor weights
>> figure;
semilogx(weights,'ro');
xlabel('Predictor index');
ylabel('Predictor weight');
```

For the prediction of Internalizing Problems (excluding outliers)

```
% Extract predictors and response. This code processes the data into the right shape for training the model.
>> inputTable = DatasetIMPOR;
predictorNames = {'SEX', 'CRACE', 'ATT15', 'ATT36', 'PARSTR', 'MSENS', 'ZMAGE', 'ZMEDUC', 'ZINCND', 'ZSGPAR', 'ZMDEP', 'ZINTI', 'ZSOC', 'ZPSYADJ', 'ZCHCA', 'ZTEMP', 'ZQSET'};
predictors = inputTable(:, predictorNames);
response = inputTable.ZMBIN;
isCategoricalPredictor = [true, true, true, true, false, false, false, false, false, false, false, false, false, false, false, false, false];
% Train a regression model. This code specifies all the model options and trains the model.
>> regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'ardexponential', ...
    'Standardize', true);
% Find the predictor weights by taking the exponential of the negative learned length scales. Normalize the weights.
>> sigmaL = regressionGP.KernelInformation.KernelParameters(1:end-1);
weights = exp(-sigmaL);
weights = weights/sum(weights);
% Plot the normalized predictor weights
>> figure;
semilogx(weights,'ro');
xlabel('Predictor index');
ylabel('Predictor weight');
```

Appendix K

Machine Learning Predictive Modeling using Data excluding Outliers

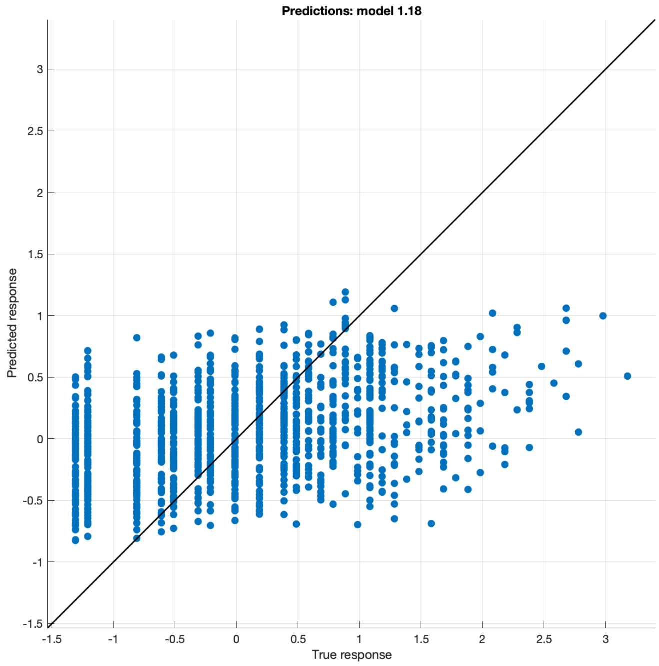
Model Statistics for the Prediction of Externalizing Problems

Model Type (Kernel Function)	RMSE	R-squared
Gaussian Process Regression (Exponential GPR)	0.85623	0.17
Gaussian Process Regression (Rational Quadratic GPR)	0.85903	0.16
Gaussian Process Regression (Matern 5/2 GPR)	0.86014	0.16
Gaussian Process Regression (Squared Exponential GPR)	0.86113	0.16
Linear Regression (Linear)	0.86301	0.16
Support Vector Machines (Coarse Gaussian SVM)	0.86431	0.15
Linear Regression (Robust Linear)	0.86479	0.15
Support Vector Machines (Linear SVM)	0.86998	0.14
Support Vector Machines (Medium Gaussian SVM)	0.87119	0.14
Ensemble (Bagged Trees)	0.88221	0.12
Ensemble (Boosted Trees)	0.88467	0.11
Linear Regression (Stepwise Linear)	0.88868	0.11
Support Vector Machines (Quadratic SVM)	0.90035	0.08
Tree (Coarse Tree)	0.91142	0.06
Support Vector Machines (Fine Gaussian SVM)	0.93777	0
Support Vector Machines (Cubic SVM)	0.96465	-0.05
Linear Regression (Interactions Linear)	0.97279	-0.07
Tree (Medium Tree)	1.0049	-0.14
Tree (Fine Tree)	1.1082	-0.39

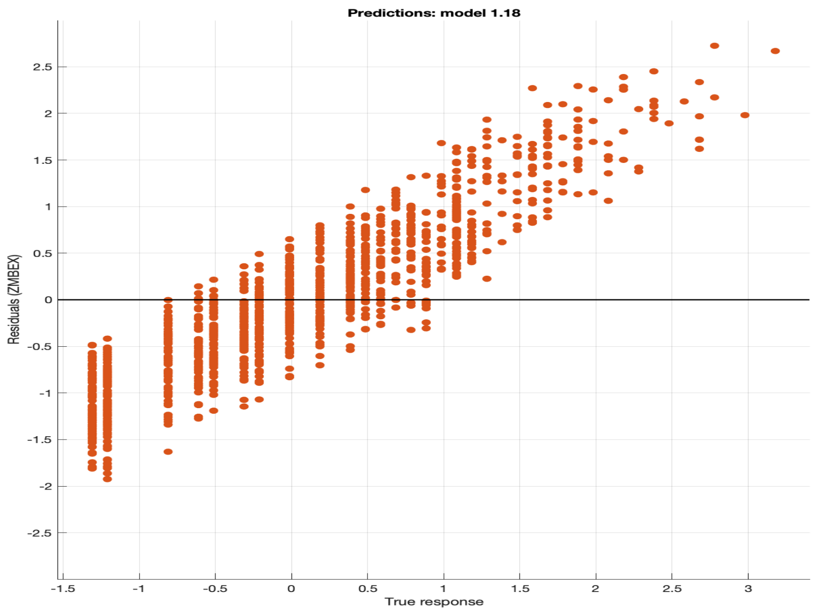
Model Statistics for the Prediction of Internalizing Problems

	RMSE	R-squared
Gaussian Process Regression (Exponential GPR)	0.8995	0.1
Gaussian Process Regression (Matern 5/2 GPR)	0.90241	0.1
Gaussian Process Regression (Squared Exponential GPR)	0.90244	0.1
Gaussian Process Regression (Rational Quadratic)	0.90381	0.1
Support Vector Machines (Coarse Gaussian SVM)	0.90499	0.1
Linear Regression (Linear)	0.90532	0.1
Linear Regression (Robust Linear)	0.9056	0.1
Support Vector Machines (Linear SVM)	0.90589	0.1
Ensemble (Bagged Trees)	0.91016	0.09
Ensemble (Boosted Trees)	0.91123	0.08
Support Vector Machines (Medium Gaussian SVM)	0.92109	0.06
Linear Regression (Stepwise Linear)	0.93966	0.03
Support Vector Machines (Quadratic SVM)	0.94622	0.01
Support Vector Machines (Fine Gaussian SVM)	0.95139	0
Tree (Coarse Tree)	0.96568	-0.03
Support Vector Machines (Cubic SVM)	1.0136	-0.13
Linear Regression (Interactions Linear)	1.0228	-0.15
Tree (Medium Tree)	1.0545	-0.23
Tree (Fine Tree)	1.1465	-0.45

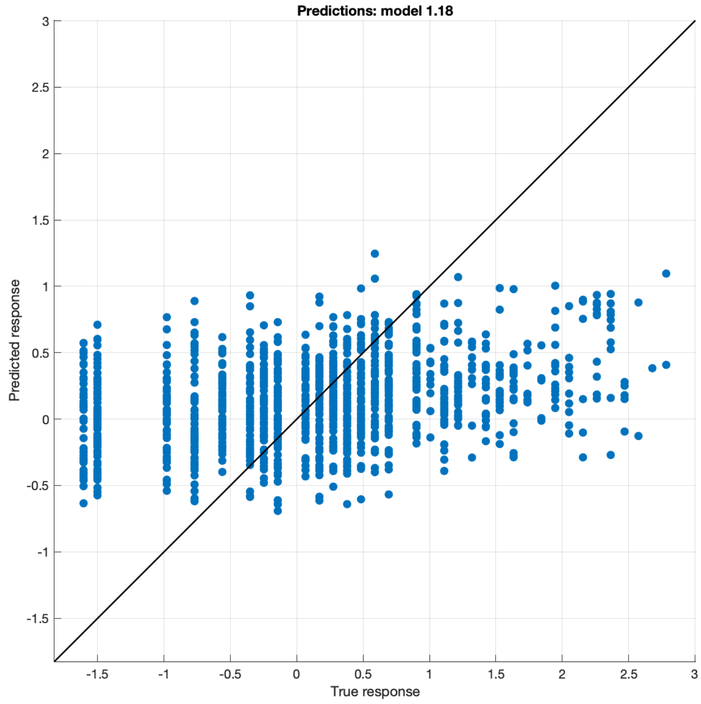
Plot of Predicted vs. Actual Response for Externalizing Problems



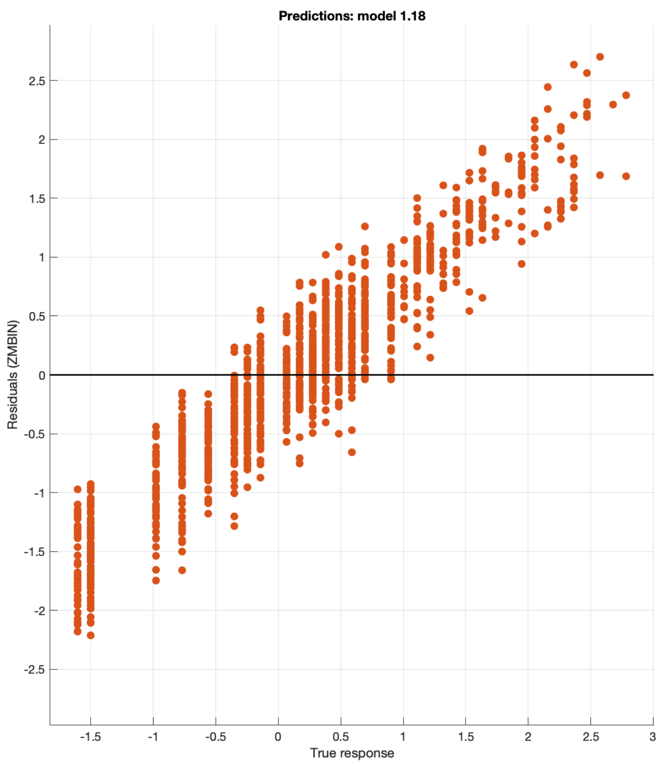
Plot for Residuals Plot for Externalizing Problems



Plot of Predicted vs. Actual Response for Internalizing Problems



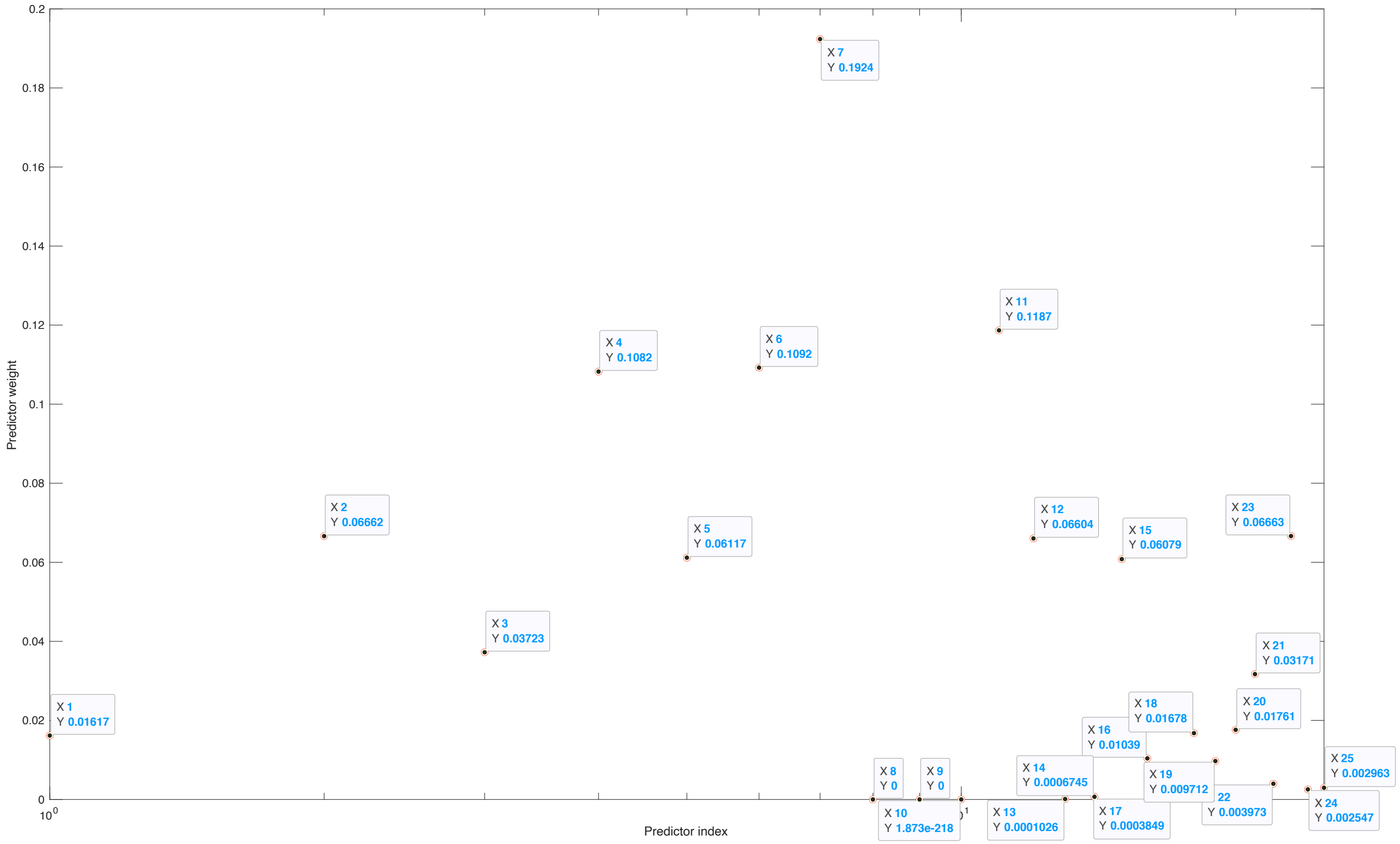
Plot for Residuals Plot for Internalizing Problems



Normalized Weights for Features Predicting Externalizing Problems

Feature Name	Predictor Index (x Label)	
Type C 15-mo Attachment	7	0.19235133
Type C 36-mo Attachment	11	0.11867031
Type B 15-mo Attachment	6	0.10924468
Other Ethnic Background	4	0.10823366
Non-Family Child Care Hours	23	0.06663231
Female Gender	2	0.06661884
Type D 36-mo Attachment	12	0.0660365
Type A 15-mo Attachment	5	0.06117073
Maternal Age	15	0.06079368
‘White’ Ethnic Background	3	0.03723455
Social Support	21	0.03170639
Intimacy	20	0.01761485
Single Parenthood	18	0.01677623
Male Gender	1	0.01616643
Maternal Education	16	0.01039214
Maternal Depression	19	0.0097118
Maternal Psychological Adjustment	22	0.00397273
24-mo Attachment Q-Sort Security	25	0.00296348
Child's Temperament (mother-report)	24	0.00254749
Early Maternal Sensitivity	14	0.00067446
Income-to-Need Ratio	17	0.00038485
Parenting Stress	13	0.00010256
Type B 36-mo Attachment	10	1.87E-218
Type D 15-mo Attachment	8	0
Type A 36-mo Attachment	9	0

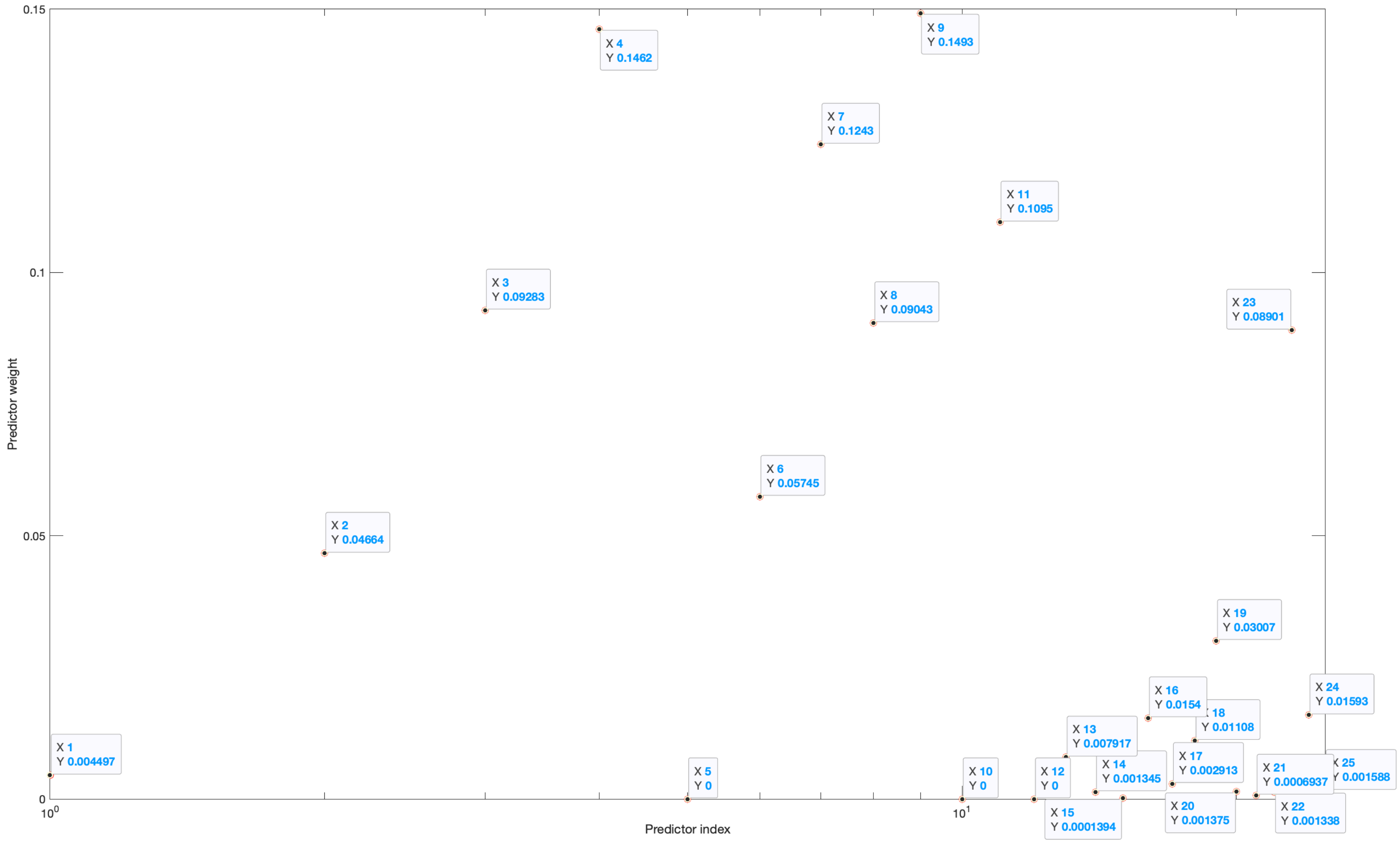
Plot of Normalised Weights for Features Predicting Externalising Problems



Normalised Weights for Features Predicting Internalizing Problems

Feature Name	Predictor Index (x Label)	Normalized Weights (y Labels)
Type A 36-mo Attachment	9	0.14926553
Other Ethnic Background	4	0.14622479
Type C 15-mo Attachment	7	0.1243326
Type C 36-mo Attachment	11	0.10953503
‘White’ Ethnic Background	3	0.09282994
Type D 15-mo Attachment	8	0.09043297
Non-Family Child Care Hours	23	0.08901047
Type B 15-mo Attachment	6	0.05744969
Female Gender	2	0.04663659
Maternal Depression	19	0.03007105
Maternal Rating of Child's Temperament	24	0.01592898
Maternal Education	16	0.01540078
Single Parenthood	18	0.01107572
Parenting Stress	13	0.00791739
Male Gender	1	0.00449691
Income-to-Need Ratio	17	0.00291344
24-mo Attachment Q-Sort Security	25	0.00158774
Intimacy	20	0.00137456
Early Maternal Sensitivity	14	0.0013445
Maternal Psychological Adjustment	22	0.00133824
Social Support	21	0.00069371
Maternal Age	15	0.00013936
Type A 15-mo Attachment	5	0
Type B 36-mo Attachment	10	0
Type D 36-mo Attachment	12	0

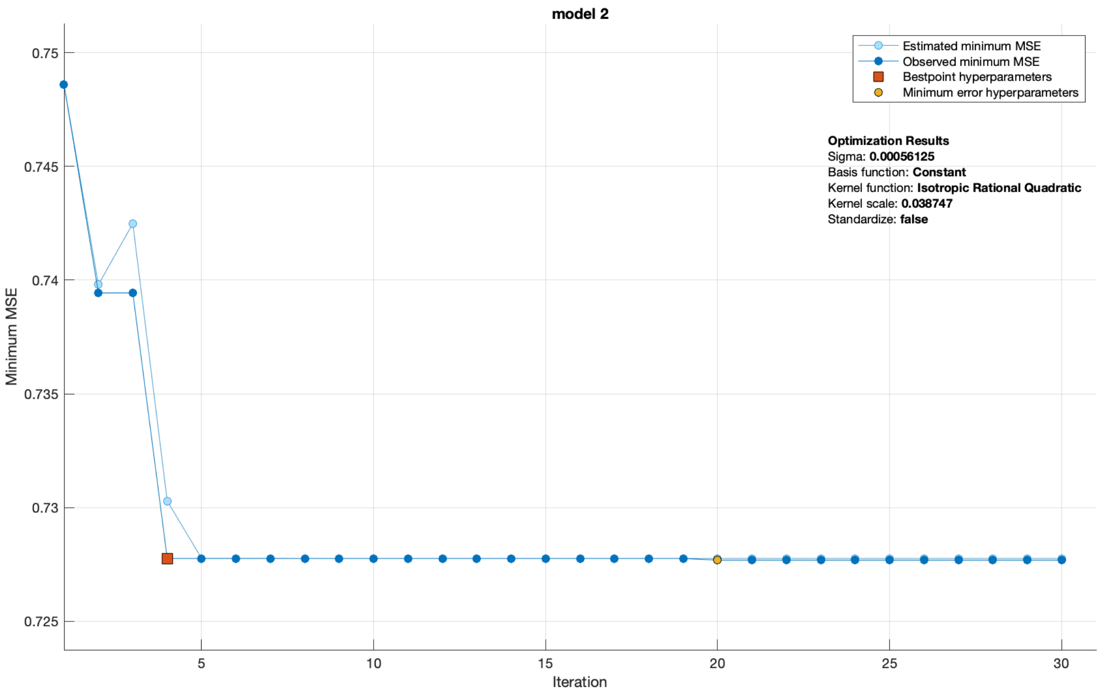
Plot of Normalised Weights for Features Predicting Internalizing Problems



Model Statistics Before and After Hyperparameter Optimization

	Externalizing		Internalizing	
	Before	After	Before	After
Model (Kernel Function)	GPR (Exponential)	Optimizable GPR (Rational Quadratic)	GPR (Exponential)	Optimizable GPR (Rational Quadratic)
RMSE	0.85623	0.85309	0.8995	0.8972
R-squared	0.17	0.18	0.1	0.11

Minimum MSE Plot for Externalizing Problems



Minimum MSE Plot for Internalizing Problems