



Research paper

# Predicting children's emotional and behavioral difficulties at age five using pregnancy and newborn risk factors: Evidence from the UK Household Longitudinal Study

Xu Zong<sup>a,b,\*</sup>, Yan Li<sup>c,\*\*</sup>, Can Liu<sup>d,e,f</sup>, Edith Aguirre<sup>g</sup>

<sup>a</sup> Helsinki Institute for Demography and Population Health, Faculty of Social Sciences, University of Helsinki, Unioninkatu 33, Helsinki 00710, Finland

<sup>b</sup> Max Planck — University of Helsinki Center for Social Inequalities in Population Health, Unioninkatu 33, Helsinki, 00710, Finland

<sup>c</sup> Faculty of Medicine, Dept. of Psychology and Logopedics, University of Helsinki, Haartmaninkatu 3, Helsinki 00290, Finland

<sup>d</sup> Department of Public Health Sciences, Stockholm University, Stockholm 11419, Sweden

<sup>e</sup> Centre for Health Equity Studies (CHESS), Stockholm University/Karolinska Institutet, Stockholm 11419, Sweden

<sup>f</sup> Clinical Epidemiology Division, Department of Medicine, Karolinska Institutet, Solna, 17165, Stockholm, Sweden

<sup>g</sup> Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom



## ARTICLE INFO

## Keywords:

Machine learning  
Emotional and behavioral difficulties  
Early childhood  
Risk factors  
Gender differences

## ABSTRACT

Childhood emotional and behavioral difficulties have a profound impact on later life outcomes, making it crucial to identify early-life risk factors that predict emotional and behavioral difficulties. However, much of the existing research has concentrated on diagnosing, rather than predicting, emotional and behavioral difficulties, and has often focused on adolescents rather than younger children. This study employs machine learning (ML) techniques to construct an interpretable predictive model using data from the UK Household Longitudinal Study, aiming to identify key risk factors that influence children's emotional and behavioral difficulties during childhood. We examined maternal habits during pregnancy and parent-reported data on birth, breastfeeding and regulatory problems during the newborn stage. Our findings highlighted lack of breastfeeding, low birthweight and maternal smoking during pregnancy as the three most significant predictors of emotional behavioral difficulties. Other important factors were related to infant regulatory problems. Heterogeneity analysis revealed gender differences in predictive power, with maternal smoking during pregnancy being a stronger predictor for boys, and the amount of fussing in infancy having a greater impact on girls. This study highlights the importance of comprehensive prenatal and postnatal care, advocates for early screening of emotional and behavioral difficulties, and calls for gender-specific approaches in assessing and addressing emotional and behavioral difficulties in children.

## 1. Introduction

Childhood emotional and behavioral difficulties in early life are associated with long-term adverse outcomes, including persistent physical and mental health problems, as well as impaired academic performance (Agnafors et al., 2021; D'Souza et al., 2019; Hofstra et al., 2002). Exposure to early life risks during critical periods of development has been linked to structural, physiological, and metabolic changes in the fetus and infant, increasing the likelihood of psychological and developmental disturbances in childhood (Cioni et al., 2016; Sellers

et al., 2015; Stein et al., 2014). Research suggests that cumulative risk factors, such as maternal smoking, and perceived stress during pregnancy, significantly elevate the probability of emotional and behavioral difficulties in childhood (D'Souza et al., 2019; Theunissen et al., 2022; Wallander et al., 2019). Beyond risk factors during pregnancy, factors during the newborn stage—including low birthweight and lack of breastfeeding—have also been associated with an increased risk of childhood emotional and behavioral difficulties (Oddy et al., 2010; Rice et al., 2007). Furthermore, severe regulatory problems including excessive crying, feeding-eating and sleeping, have been found

\* Correspondence to: X. Zong, Room 308, Unioninkatu 33, Helsinki 00710, Finland.

\*\* Correspondence to: Y. Li, Haartmaninkatu 3, Helsinki 00290, Finland.

E-mail addresses: [xu.zong@helsinki.fi](mailto:xu.zong@helsinki.fi) (X. Zong), [yan.z.li@helsinki.fi](mailto:yan.z.li@helsinki.fi) (Y. Li), [can.liu@su.se](mailto:can.liu@su.se) (C. Liu), [edith.aguirre@essex.ac.uk](mailto:edith.aguirre@essex.ac.uk) (E. Aguirre).

<sup>1</sup> Xu Zong and Yan Li contributed equally to this work. Therefore, both authors can list themselves as the first author.

associated with emotional and behavioral disorders in school ages (Asmussen et al., 2023; Pant et al., 2024). The impact of regulatory problems on children's behavioral difficulties can be long-lasting, extending into adulthood, particularly for infants who experience multiple regulatory challenges (Bilgin et al., 2020; Jaekel et al., 2021). Thus, there is broad consensus regarding the influence of early life risks on childhood emotional and behavioral difficulties, underscoring the critical need for early identification and intervention strategies to mitigate these challenges (Lopez et al., 2000; Tennant et al., 2007).

The deficit model suggests that an accumulation of risk factors leads to poorer outcomes, with children from high-risk backgrounds experiencing developmental disadvantages (Evans et al., 2013; McGrath et al., 2020). These adversities can disrupt neurodevelopmental, physiological, and socio-emotional processes (Doom et al., 2023; McGrath et al., 2020; Nelson et al., 2020), increasing the likelihood of long-term difficulties in behavioral and emotional functioning. In line with this framework, previous research has identified early-life risks including parental mental health, maternal smoking during pregnancy, low birthweight, lack of breastfeeding, and early regulatory problems as key contributors to emotional and behavioral difficulties in childhood (Asmussen et al., 2023; Huizink and Mulder, 2006; Oddy et al., 2010; Ramchandani et al., 2008; Theunissen et al., 2022).

Despite significant progress in understanding these risk factors, much of the existing research has relied on traditional statistical methods that may not fully capture the complex and nonlinear interactions between multiple early-life risks and children's emotional and behavioral difficulties. Machine learning (ML) approaches offer a powerful alternative, enabling the analysis of high-dimensional data to uncover hidden patterns that conventional methods may overlook (Abdolali and Gillis, 2021; Aghaabbasi and Chalermpong, 2023). While ML has been extensively applied in diagnosing mental health conditions in children, such as identifying symptoms of depression (Gaus et al., 2023; Haque et al., 2021; Khare et al., 2023; Vaishnavi et al., 2022), its potential for predicting emotional and behavioral difficulties based on early-life risk factors remains underexplored. Moreover, the existing research on predicting children's emotional and behavioral problems has focused on adolescents rather than younger children (Crowley et al., 2024; Rothenberg et al., 2023; Tate et al., 2020). Additionally, regulatory problems in children have been largely overlooked in previous ML studies.

Furthermore, gender differences in emotional and behavioral difficulties are well-documented, with girls exhibiting higher rates of internalizing disorders, such as mood and anxiety disorders, and boys showing higher prevalence of externalizing disorders, such as disruptive behavior (Giarelli et al., 2010; Liu et al., 1999; Vater and Schröder-Abé, 2015; Zahn-Waxler et al., 2008). For instance, boys between the ages of five and eight exhibit greater temperamental difficulty and lower persistence scores (Prior et al., 1993). Regarding sensitivity to early life stress, adolescent studies suggest that girls are more likely than boys to develop depression and anxiety symptoms when exposed to similar levels of early life stress (Bouma et al., 2008; Jiang et al., 2022). In younger children, research on those aged 4–7 found that maternal exposure to intimate partner violence was associated with a higher prevalence of behavioral and inattention/hyperactivity problems in young girls compared to young boys (Sonego et al., 2018). However, perinatal, neonatal, and early childhood factors appear to place young boys at an inherently greater baseline risk, particularly for neurodevelopmental challenges (Hintz et al., 2006; O'Driscoll et al., 2018). Therefore, while younger girls may be more vulnerable to certain early-life stressors, boys' heightened baseline risk for neurodevelopmental challenges suggests that susceptibility to early adversities may manifest differently across genders, depending on the specific risk factors.

Given the profound influence of childhood emotional and behavioral difficulties on subsequent life stages (Baker et al., 2003; Thompson et al., 2021), it is crucial to detect early-life predictors of childhood emotional and behavioral difficulties. By integrating insights from the deficit

model, this study aims to construct an interpretable predictive model using population-based data to identify prenatal and postnatal risk factors associated with emotional and behavioral difficulties in children at the age of five, and to explore gender differences in these predictions. These factors include maternal habits during pregnancy and parent-reported data on birth, breastfeeding, and regulatory problems during the newborn stage.

## 2. Data and method

### 2.1. Participants

Participants in this study were drawn from the Pregnancy and Early Childhood (PEACH) data, which is part of Understanding Society, the UK Household Longitudinal Study (UKHLS), a nationally representative survey initiated in the UK in 2009 (University of Essex, Institute for Social and Economic Research, 2023, 2024). PEACH includes rich and detailed information of 19,327 children and their families up to 10 years, reported by their parents, focusing on key developmental points before and after childbirth, with the latest release including children reported up to wave 13, the most recent wave released by Understanding Society. Moreover, all this information can be linked to UKHLS to gather data about parents and other family members, as well as household characteristics, making it a valuable resource. When children are recorded as newborns in Understanding Society, information on pregnancy and newborn is gathered. However, the Strengths and Difficulties Questionnaire (SDQ) is administered when the child reaches five years of age, five years after the initial newborn data collection. The data collection was approved by the University of Essex Ethics Committee. Detailed information has been published (University of Essex, Institute for Social and Economic Research, 2024). The analysis included data from 5882 children who had complete information on the SDQ Total Difficulties Score.

### 2.2. Measures

Children's emotional and behavioral difficulties are measured using the Strength and Difficulties Questionnaire (SDQ), a scale that has been widely used to assess mental health in epidemiological, psychological, and educational studies (Arman, 2012; Hobbs et al., 2007; Ortuño-Sierra et al., 2018). In Understanding Society, SDQs are administered when children are 5 and 8 years old, with responses provided by parents. For this study, we are focusing on SDQs collected at age 5, as this timing closely aligns with the pregnancy and newborn information gathered, allowing us to better examine how early-life risk factors influence children's emotional and behavioral outcomes. In Understanding Society, SDQ measures are available to assess positive and negative behaviors of children, including five subscales: emotional symptoms, conduct problems, hyperactivity, peer relationship problems and prosocial behavior. Following a prior study (Aguirre et al., 2024), we used the SDQ Total Difficulties Score as the measure of children's emotional and behavioral difficulties. This score is calculated by summing the scores from the SDQ subscales, excluding prosocial behavior, with higher scores indicating more emotional and behavioral difficulties.

To evaluate the importance of early-life risk factors during pregnancy and newborn stages in predicting emotional and behavioral difficulties at the age of five, we incorporated five risk factors related to pregnancy and nine risk factors related to the newborn stage. These include type of delivery (normal delivery or c-section), multiple births, smoked during pregnancy, frequency of alcohol during pregnancy, fertility treatments, birth timing, birthweight, ever breastfed, problems calming child, amount of fussing, time to sleep, number of wakes, refuse to eat, no appetite statement. Detailed information on these factors can be found in Appendix A. Considering the gender disparity in emotional and behavioral difficulties, gender was included as covariate.

To understand the relationships between different pregnancy and

newborn risk factors, we calculated the Pearson correlation matrix for all predictors. This correlation matrix shows the potential multicollinearity issue and how various risk factors are related to each other. As shown in Fig. 1, hierarchical clustering dendrogram reveals features with similar correlation patterns, such as ‘Birthweight’ and ‘Birth timing’, indicating high similarity. Fig. 2 illustrates the intercorrelations among the predictors included in the predictive model. Notably, ‘Multiple births’ and ‘Number of wakes’, ‘Birthweight’ and ‘Birth timing’ show moderate bivariate correlations, while other features mainly demonstrate a low level of correlation.

### 2.3. Machine learning model

In this study, we utilized CatBoost regression models to assess the importance of 14 risks factors related to pregnancy and newborn stages in predicting emotional and behavioral difficulties among children at the age of five in the UK. CatBoost, a powerful machine learning model for classification and regression tasks, has been applied to identify risk factors of mental health (Rezpour and Hansen, 2022; Singh et al., 2022; Zhang et al., 2021). This model has several advantages beyond conventional linear regression models. CatBoost model can capture nonlinear associations between predictors and outcomes, deal with multicollinearity and uncover novel patterns due to its non-parametric nature. In this study, the CatBoost model was developed using the Python package ‘catboost’.

We divided the dataset into training (70 %) and testing (30 %) subsets to ensure generalization capability of the CatBoost model. To tune hyperparameter, we used random search and 5-fold cross-validation. To enhance the interpretability of CatBoost, we employed

Shapley Additive Explanations (SHAP), an extension of game theory (Baptista et al., 2022) to measure the contribution of each predictor in CatBoost models.

We used two indicators to assess the predictive performance of the CatBoost model, including mean squared error (MSE) and Root Mean Squared Error (RMSE).

We first trained the CatBoost model using the full sample to assess the overall predictive power of early-life risk factors. Then, we conducted separate analyses for boys and girls to examine potential gender differences in the influence of these risk factors on emotional and behavioral difficulties at age five.

### 3. Result

In this study, the MSE value of CatBoost model is 31.18 and RMSE value is 5.58. Fig. 3 shows the importance ranking of the pregnancy and newborn predictors of emotional and behavioral difficulties at the age of five, as well as the nonlinear relationship between some predictors and emotional and behavioral difficulties. We found that ‘Ever breastfed’ is the most crucial predictor among early-life risk factors, indicating not being breastfed has a large influence on more emotional and behavioral difficulties at the age of five.

‘Birthweight’ is the second most important predictor among early-life risk factors, with lower birthweights generally associated with more emotional and behavioral difficulties at age five. ‘Birthweight’ also shows a distinct nonlinear relationship with children’s emotional and behavioral difficulties. To illustrate this, we created a partial dependence plot (Fig. 4), which reveals the ranges where the predictor’s SHAP values are positive, neutral, or negative. Our analysis identified a

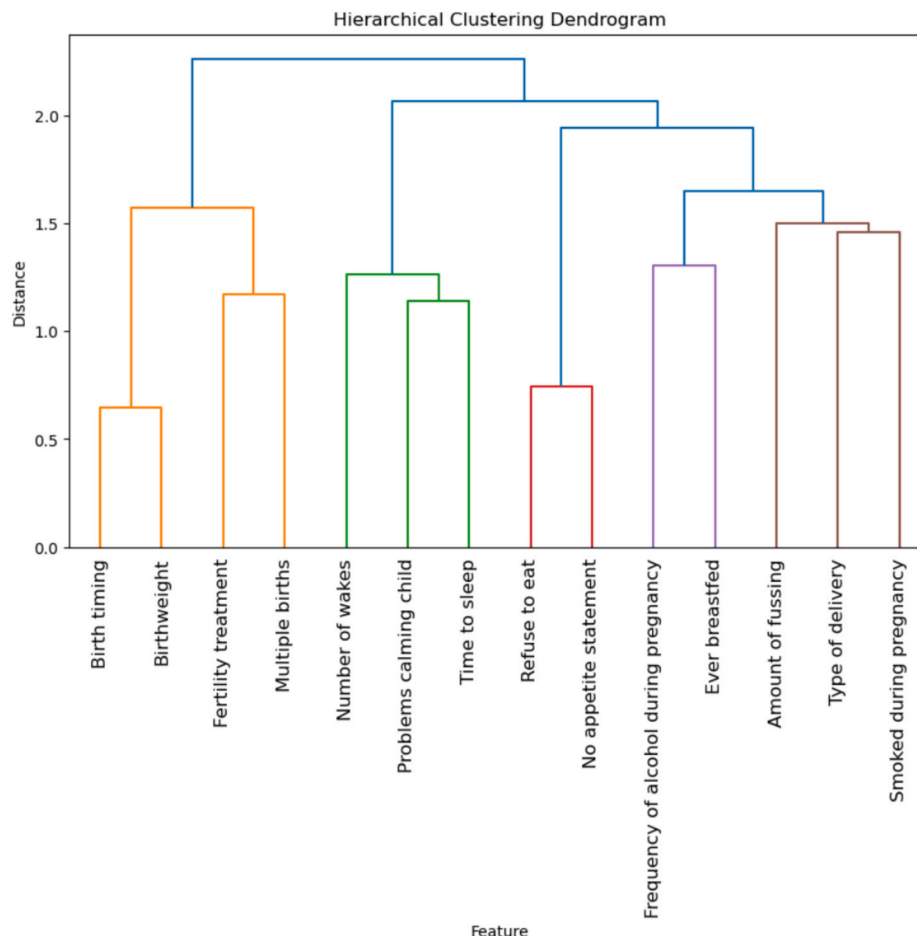


Fig. 1. Hierarchical clustering dendrogram.

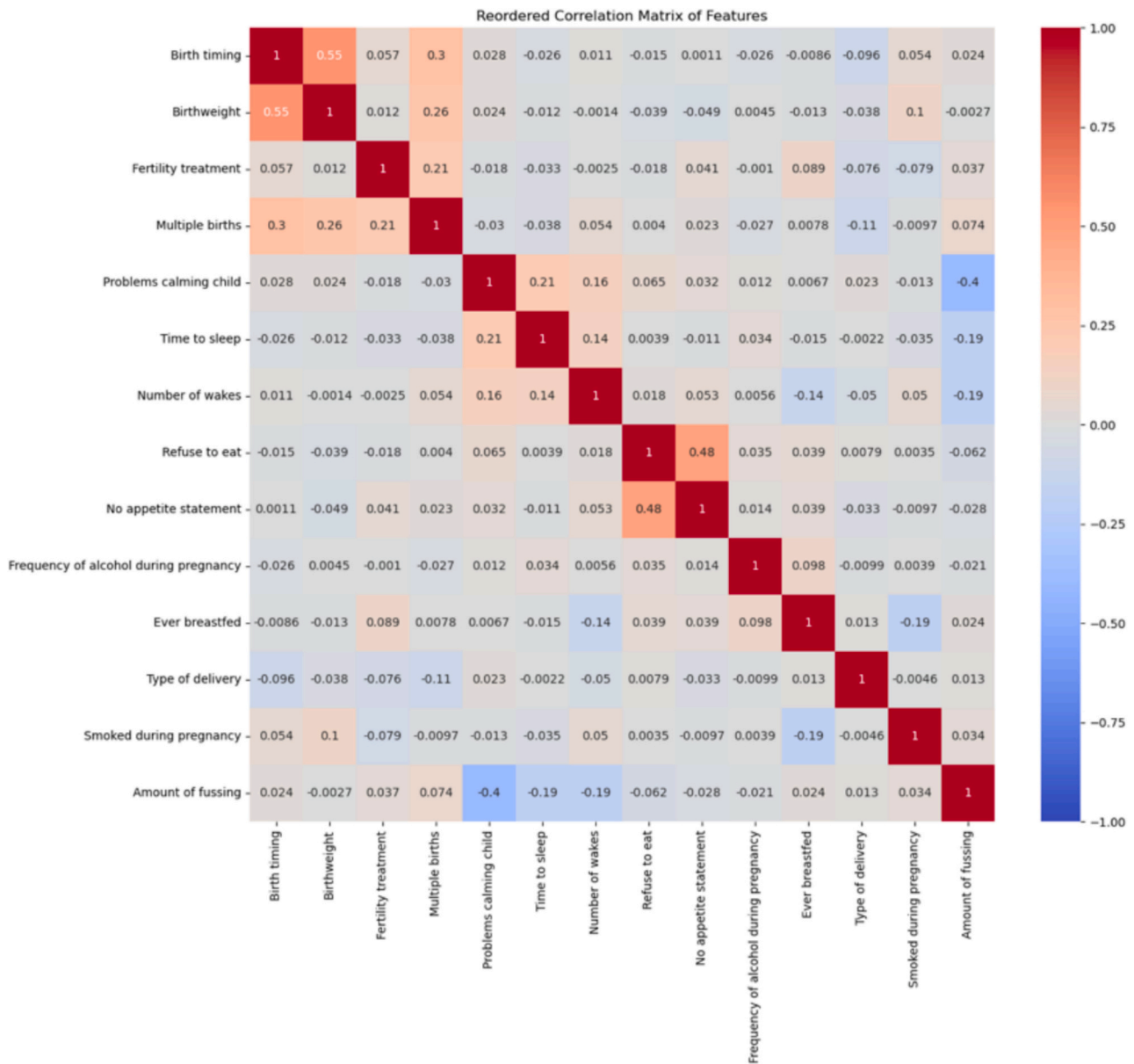


Fig. 2. Correlation matrix of the features included in the predictive model.

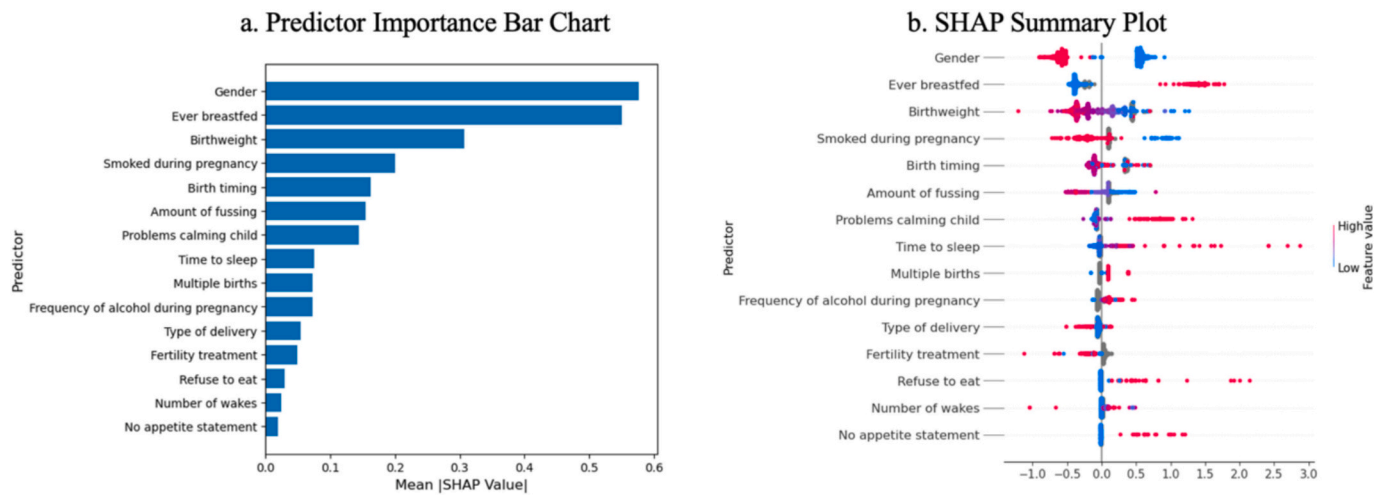
threshold around three kilograms of birthweight: positive SHAP values, i.e. higher degree of emotional and behavioral difficulties at age five, were observed below this point, with a sharp decline in SHAP values between two and three kilograms. Additionally, maternal smoking during pregnancy emerged as a significant predictor, underscoring the harmful impact of prenatal exposure to smoking on children’s emotional and behavioral difficulties. Other important predictors include ‘Problems calming child’, ‘Amount of fussing’ and ‘time to sleep’, which are crucial indicators of future emotional and behavioral difficulties and may be early signs of underlying psychological issues. Additional important predictors like ‘Birth timing’, ‘Type of delivery’ and ‘Refuse to eat’ have distinctive contributions to predicting emotional and behavioral difficulties at the age of five. Notably, ‘Fertility treatment’ and ‘Number of wakes’ have relatively low importance in predicting emotional and behavioral difficulties, and their positive or negative influence is not clear, showing the nonlinear associations.

To discern the gender disparity of the pregnancy and newborn predictor on children’s emotional and behavioral difficulties, we trained separate CatBoost models for boys and girls using their respective

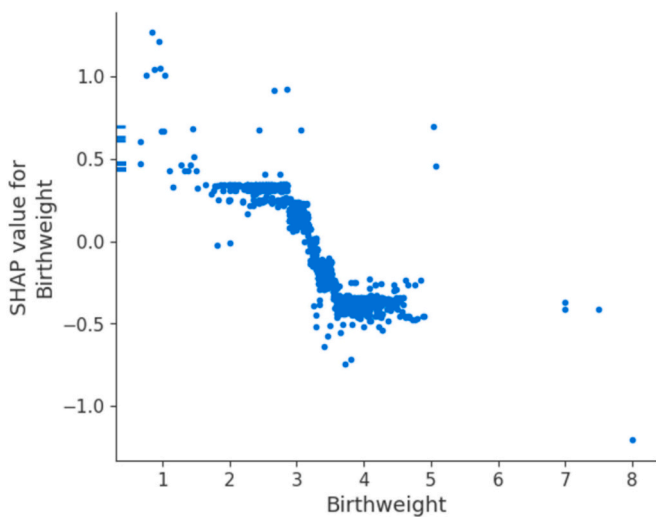
datasets and then tested each model within the corresponding gender-specific test set. The results are shown in Fig. 5. Male children and female children share some common top predictors, such as ‘ever breastfed’, ‘birthweight’ and ‘smoked during pregnancy’. However, there are different predictive powers of certain pregnancy and newborn predictors. For example, based on the importance ranking, ‘smoked during pregnancy’ is a more important predictor of emotional and behavioral difficulties for boys, while ‘amount of fussing’ is more predictive for girls.

#### 4. Discussion

The current study used ML to evaluate how the early-life risks during prenatal and postnatal periods contribute to child emotional and behavioral difficulties at five years old. It revealed that lack of breastfeeding is the most important predictor of children’s emotional and behavioral difficulties, followed by lower birthweight, maternal smoking during pregnancy and regulatory problems (i.e., difficulties in calming the child, prolonged fussing). Notably, there was a gender



**Fig. 3.** SHAP values of pregnancy and newborn predictors for emotional and behavioral difficulties at the age of five. Note: Predictor Importance Bar Chart shows the ranked importance of predictors based on the mean absolute SHAP value, indicating their overall impact on emotional and behavioral difficulties at the age of five. The top-ranked predictors play a more critical role in the predicted outcome. SHAP Summary Plot demonstrates the contribution of each predictor to emotional and behavioral difficulties at the age of five, interpreting both the direction and magnitude of each predictor’s impact. In this plot, red dots represent higher values of the predictors, while blue dots indicate lower values. Predictors positioned to the left of the origin have a negative impact on the outcome, whereas those on the right exert a positive influence. A more detailed explanation of each predictor’s value can be found in [Appendix A](#). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Partial dependence plots for birthweight on the children’s emotional and behavioral difficulties at the age 5. Note: Each point means the SHAP value of birthweight for each child. Birthweight values (measured in kilograms) are plotted on the x-axis. A higher SHAP value on the y-axis indicated a higher degree of emotional and behavioral difficulties for a given individual.

difference between the risk factors during pregnancy and the newborn period in predicting children’s emotional and behavioral difficulties. For instance, maternal smoking during pregnancy is a more important predictor for boys’ emotional and behavioral difficulties, while ‘amount of fussing’ has a stronger effect on girls than boys.

Our findings emphasize the role of parental investment, both in terms of time and money, in mitigating emotional and behavioral difficulties associated with children’s development, replicating previous research ([Francesconi and Heckman, 2016](#); [Jiang and Shi, 2024](#)). Specifically, it first highlights the central role of lack of breastfeeding as a key predictor of emotional and behavioral difficulties at age five, which has not been given sufficient attention in previous studies using ML to explore risk factors for children’s mental health ([Crowley et al., 2024](#);

[Rothenberg et al., 2023](#); [Tate et al., 2020](#)). Children not breastfed showed more emotional and behavioral difficulties, supporting research that identifies breastfeeding’s nutritional, psychological, and developmental benefits ([Modak et al., 2023](#)). The release of oxytocin during breastfeeding fosters an essential emotional connection, trust, and attachment between the mother and infant, which are critical for the long-term mental health and resilience of children ([Motataianu, 2015](#); [Smith et al., 2016](#); [Winston and Chicot, 2016](#)). Consequently, infants who experience limited or no breastfeeding may lack early bonding with their mothers, potentially impeding their mental health recovery and resilience ([Winston and Chicot, 2016](#)). Thus, in clinical practice, it is important to emphasize that breastfeeding offers more than just nutritional benefits; it also plays a critical role in fostering emotional bonds and supporting long-term emotional and behavioral development, as well as resilience in children.

Similarly, lower birthweight is linked to greater emotional and behavioral difficulties at the age of five, possibly due to physiological immaturity at birth affecting brain development and increasing vulnerability to cognitive and emotional disorders ([Nomura et al., 2007](#); [Rice et al., 2007](#); [Upadhyay et al., 2019](#)). Our findings further highlight a nonlinear relationship between birthweight and emotional and behavioral difficulties, with a critical threshold around 3 kg, below which the predicted risk increases sharply. The association between lower birthweight and increased emotional and behavioral difficulties underscores the need for vigilant monitoring and early intervention for low birthweight infants. Additionally, maternal smoking during pregnancy predicts children’s emotional and behavioral difficulties, with neurodevelopmental impacts from nicotine exposure influencing the development of attention and behavioral regulation ([Baler et al., 2008](#); [Gatzke-Kopp and Rice, 2023](#)). The identification of maternal smoking during pregnancy as a major predictor of emotional and behavioral difficulties underscores the critical need for targeted smoking cessation programs for expectant mothers to mitigate neurodevelopmental risks for their children.

At the same time, regulatory problems (e.g., difficulties in calming the child) during infancy also predict emotional and behavioral difficulties, consistent with previous studies that the co-occurrence of two or more regulatory problems marks an early vulnerability to dysregulation, which can lead to mental disorders in early childhood ([Asmussen et al., 2023](#); [Pant et al., 2024](#)). Therefore, regulatory issues in during infancy

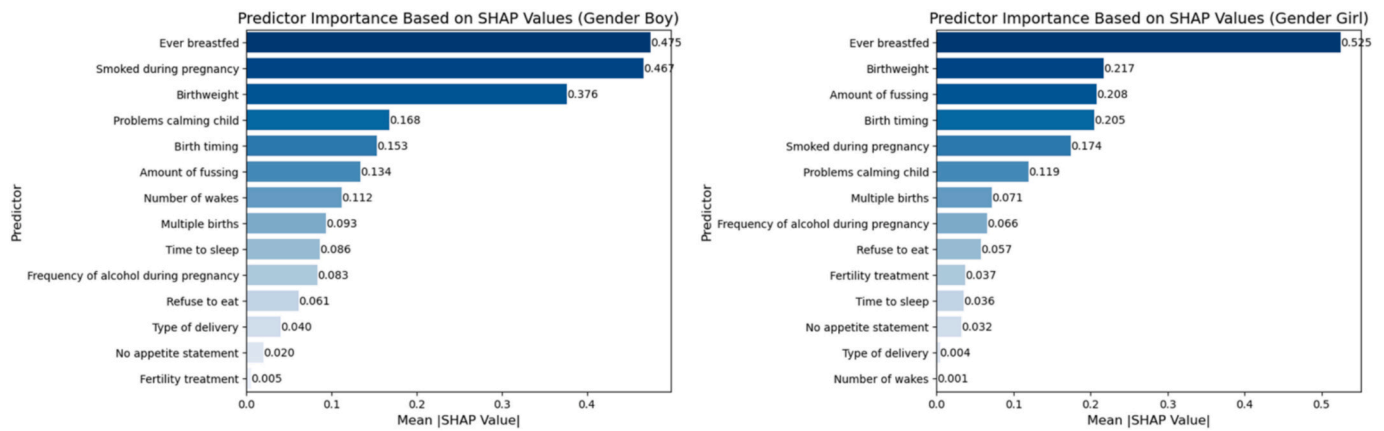


Fig. 5. SHAP values of pregnancy and newborn predictors across gender.

Note: The two bar charts show the importance ranking of predictors based on the mean absolute SHAP value, alongside the corresponding value for each predictor across gender. A higher ranking indicates a greater impact of the predictor on emotional and behavioral difficulties at the age of five.

should be considered early indicators of potential emotional and behavioral challenges, prompting pediatricians and caregivers to monitor these behaviors closely and implement early interventions as needed.

Furthermore, our analysis identified gender differences in how risk factors during pregnancy and the newborn stage influence children's emotional and behavioral difficulties. Specifically, we found that maternal smoking during pregnancy was a stronger predictor of these difficulties in boys, whereas the amount of fussing in infancy was more predictive for girls. The greater impact of maternal smoking during pregnancy on boys' emotional and behavioral difficulties may be attributed to inherent neurodevelopmental vulnerabilities (Hintz et al., 2006). Male fetuses develop more slowly than females, particularly in brain regions responsible for emotional regulation and impulse control, such as the prefrontal cortex (Wheelock et al., 2019). This slower maturation makes boys more susceptible to prenatal disruptions, including nicotine exposure, which affects the development of attention and behavioral regulation (Gatzke-Kopp and Rice, 2023). In contrast, girls' greater susceptibility to fussing-related distress may reflect differences in emotional processing and regulatory capacities, as previous research suggests that girls exhibit greater sensitivity to social and environmental stressors early in life, e.g., intimate partner violence (Sonego et al., 2018). These findings highlight the need for gender-specific approaches in early screening and intervention programs. For boys, efforts to reduce prenatal smoking exposure may be particularly important, while for girls, early interventions that promote emotional regulation and caregiver responsiveness to distress could help mitigate long-term behavioral difficulties. Future research should further explore the underlying mechanisms driving these gender differences, particularly the interplay between biological sensitivity and early caregiving environments.

#### 4.1. Advantages and limitations

This study has several notable advantages. First, this study utilized large-sample data from a representative national survey in UK, offering more statistical power than small-sample or regional data. Second, we included 14 risk factors during pregnancy and newborn periods, such as regulatory problems, which have been neglected in previous ML studies and provided a comprehensive understanding of the importance of each factor. Additionally, we accounted for gender-specific disparities in the influence of pregnancy and newborn risk factors, providing further insights. Third, we developed a machine learning model to examine the influence of pregnancy and newborn risk factors on children's emotional and behavioral difficulties at the age of five. This model outperformed

conventional statistical methods, such as linear regression, in handling multicollinearity and capturing non-linear relationships (e.g., non-linear prediction of birthweight), thereby deepening our understanding of the adverse effects of pregnancy and newborn risk factors.

However, there are some limitations to this study. First, due to limitations in data availability, some pregnancy and newborn risk factors, such as parental mental health during pregnancy, which may influence children's emotional and behavioral difficulties, were not included in this study. Future study should incorporate a wider range of pregnancy and newborn risk factors, including parental mental health, enhancing the predictive accuracy of future models. Second, the primary focus of the study is to evaluate the importance of pregnancy and newborn risk factors and their associations with children's emotional and behavioral difficulties, rather than to establish causal inferences. Future research could focus on establishing causal relationships between pregnancy and newborn risk factors and emotional and behavioral difficulties, which will help to confirm and extend the associations identified in this study.

## 5. Conclusion

This study demonstrates the value of using ML to identify early-life risk factors that significantly predict children's emotional and behavioral difficulties at age five. It uncovers the contributions of pregnancy and newborn risk factors, such as lower birthweight, but also the novel predictors neglected by other ML studies, such as lack of breastfeeding, to children's emotional and behavioral difficulties at age five. Clinically, these findings pave the way for screening for vulnerability to emotional and behavioral difficulties in children, highlighting the importance of comprehensive prenatal and postnatal care. The study also underscores the necessity for gender-specific approaches in assessing and mitigating emotional and behavioral risks. Future research should aim to validate these findings, explore causal relationships, and consider a wider range of risk factors to enhance our understanding and prevention of early emotional and behavioral issues.

#### CRediT authorship contribution statement

**Xu Zong:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yan Li:** Writing – review & editing, Writing – original draft, Conceptualization. **Can Liu:** Writing – review & editing. **Edith Aguirre:** Writing – review & editing.

## Funding

Xu Zong was supported by China Scholarship Council (award number 202106140038). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

## Acknowledgements

We extend our heartfelt gratitude to the COORDINATE Summer School on Longitudinal Analysis of Child and Adolescent Data Featuring Understanding Society: The UK Household Panel Study, held from June 20 to 24, 2022, at the University of Essex. This event, organized as part of the COORDINATE project (Grant Agreement Number: 101008589), provided invaluable training and support for working with the data used in this study.

## Appendix A

Domain	Risk factor	Meaning
Pregnancy	Type of delivery	Did the pregnancy result in a live birth with a normal delivery or by caesarean section? 1 Normal delivery 2 Caesarean section
	Multiple births	Did the pregnancy result in multiple births such as twins or triplets? 1 Yes, twins 2 Yes, triplets 4 No, it was a single birth
	Smoked during pregnancy	Did you smoke at all during the pregnancy, including before you were aware that you were pregnant? 1 Yes 2 No
	Frequency of alcohol during pregnancy	Which of these best describes how often you usually drank alcohol during this pregnancy? 1 Every day 2 5–6 times per week 3 3–4 times per week 4 1–2 times per week 5 1–2 times per week 6 Less than once a week 7 Never
	Fertility treatment	Did you receive any form of fertility treatment before becoming pregnant? 1 Yes 2 No
	Newborn	Birth timing
Birthweight		How much did child weigh when they were born? (in kg)
Ever breastfed		Did you breastfeed child, even if only for a short time? 1 Yes 2 No 3 Currently breastfeeding
Problems calming child		How easy or difficult is it for you to calm or soothe child when s/he is upset? 1 Very easy 2 Somewhat easy 3 About average 4 Somewhat difficult 5 Very difficult
Amount of fussing		How often does child usually fuss and cry during a typical 24 h period (e.g. yesterday)? 1 Most of time 2 Quite a bit of the time 3 Some of the time 4 Not very often 5 Rarely if at all
Time to sleep		How long does it usually take you to settle child for his or her night time sleep one you have started the process? (time in minutes)
Number of wakes		How many times does child usually wake him/herself up at night?
Refuse to eat		How true is the following statement about child? Child refuses to eat. 1 Not true 2 Somewhat true 3 Certainly true
No appetite statement		How true is the following statement about child? Child seem to have no appetite. 1 Not true 2 Somewhat true 3 Certainly true

## References

- Abdolali, M., Gillis, N., 2021. Beyond linear subspace clustering: a comparative study of nonlinear manifold clustering algorithms. *Comput Sci Rev* 42, 100435. <https://doi.org/10.1016/j.cosrev.2021.100435>.
- Aghaabbasi, M., Chalermpong, S., 2023. Machine learning techniques for evaluating the nonlinear link between built-environment characteristics and travel behaviors: a systematic review. *Travel Behav. Soc.* 33, 100640. <https://doi.org/10.1016/j.tbs.2023.100640>.
- Agnafors, S., Barmark, M., Sydsjö, G., 2021. Mental health and academic performance: a study on selection and causation effects from childhood to early adulthood. *Soc. Psychiatry Psychiatr. Epidemiol.* 56 (5), 857–866. <https://doi.org/10.1007/s00127-020-01934-5>.
- Aguirre, E., Benzeval, M., Murray, A., 2024. Parental gender attitudes and children's mental health: evidence from the UK household longitudinal study. *Soc. Sci. Med.* 344, 116632. <https://doi.org/10.1016/j.socscimed.2024.116632>.
- Arman, S., 2012. Epidemiological study of youth mental health using strengths and difficulties questionnaire (SDQ). *Neuropsychiatr. Enfance Adolesc.* 60 (5), S225. <https://doi.org/10.1016/j.neurenf.2012.04.518>.
- Asmusen, J., Davidsen, K.A., Olsen, A.L., Skovgaard, A.M., Bilenberg, N., 2023. The longitudinal association of combined regulatory problems in infancy and mental health outcome in early childhood: a systematic review. *Eur. Child Adolesc. Psychiatry*. <https://doi.org/10.1007/s00787-023-02262-0>.
- Baker, B.L., McIntyre, L.L., Blacher, J., Crnic, K., Edelbrock, C., Low, C., 2003. Pre-school children with and without developmental delay: behaviour problems and parenting stress over time. *J. Intellect. Disabil.* 47 (4–5), 217–230. <https://doi.org/10.1046/j.1365-2788.2003.00484.x>.
- Baler, R.D., Volkow, N.D., Fowle, J.S., Benveniste, H., 2008. Is fetal brain monoamine oxidase inhibition the missing link between maternal smoking and conduct disorders? *J. Psychiatry Neurosci.* 33 (3), 187–195.
- Baptista, M.L., Goebel, K., Henriques, E.M.P., 2022. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artif. Intell.* 306, 103667. <https://doi.org/10.1016/j.artint.2022.103667>.
- Bilgin, A., Baumann, N., Jaekel, J., Breeman, L.D., Bartmann, P., Bäuml, J.G., Avram, M., Sorg, C., Wolke, D., 2020. Early crying, sleeping, and feeding problems and trajectories of attention problems from childhood to adulthood. *Child Dev.* 91 (1). <https://doi.org/10.1111/cdev.13155>.
- Bouma, E.M.C., Ormel, J., Verhulst, F.C., Oldehinkel, A.J., 2008. Stressful life events and depressive problems in early adolescent boys and girls: the influence of parental depression, temperament and family environment. *J. Affect. Disord.* 105 (1), 185–193. <https://doi.org/10.1016/j.jad.2007.05.007>.
- Cioni, G., Inguaggiato, E., Sgandurra, G., 2016. Early intervention in neurodevelopmental disorders: underlying neural mechanisms. *Dev. Med. Child Neurol.* 58 (S4), 61–66. <https://doi.org/10.1111/dmnc.13050>.
- Crowley, R., Parkin, K., Rocheteau, E., Massou, E., Friedmann, Y., Sippy, R., Liò, P., Moore, A., 2024. Machine Learning for Prediction of Childhood Mental Health Problems in Social Care, pp. 1–18. <https://doi.org/10.1101/2024.05.03.24306756>.
- Doom, J.R., Young, E.S., Farrell, A.K., Roisman, G.I., Simpson, J.A., 2023. Behavioral, cognitive, and socioemotional pathways from early childhood adversity to BMI: evidence from two prospective, longitudinal studies. *Dev. Psychopathol.* 35 (2), 749–765. <https://doi.org/10.1017/S0954579421001887>.
- D'Souza, S., Waldie, K.E., Peterson, E.R., Underwood, L., Morton, S.M.B., 2019. Antenatal and postnatal determinants of behavioural difficulties in early childhood: evidence from growing up in New Zealand. *Child Psychiatry Hum. Dev.* 50 (1), 45–60. <https://doi.org/10.1007/s10578-018-0816-6>.
- Evans, G.W., Li, D., Whipple, S.S., 2013. Cumulative risk and child development. *Psychol. Bull.* 139 (6), 1342–1396. <https://doi.org/10.1037/a0031808>.
- Francesconi, M., Heckman, J.J., 2016. Child development and parental investment: introduction. *Econ. J.* 126 (596), F1–F27. <https://doi.org/10.1111/econj.12388>.
- Gatzke-Kopp, L.M., Rice, D.R., 2023. Nicotine as an environmental toxin: implications for children's health. *Policy Insights Behav. Brain Sci.* 10 (2), 125–132. <https://doi.org/10.1177/23727322231190588>.
- Gaus, R., Pölsterl, S., Greimel, E., Schulte-Körne, G., Wachinger, C., 2023. Can we diagnose mental disorders in children? A large-scale assessment of machine learning on structural neuroimaging of 6916 children in the adolescent brain cognitive development study. *JCPP Adv.* 3 (4). <https://doi.org/10.1002/jcv2.12184>.
- Giarelli, E., Wiggins, L.D., Rice, C.E., Levy, S.E., Kirby, R.S., Pinto-Martin, J., Mandell, D., 2010. Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disabil. Health J.* 3 (2), 107–116. <https://doi.org/10.1016/j.dhjo.2009.07.001>.
- Haque, U.M., Kabir, E., Khanam, R., 2021. Detection of child depression using machine learning methods. *PLoS One* 16 (12), e0261131. <https://doi.org/10.1371/journal.pone.0261131>.
- Hintz, S.R., Kendrick, D.E., Vohr, B.R., Poole, W.K., Higgins, R.D., 2006. Gender differences in neurodevelopmental outcomes among extremely preterm, extremely low-birthweight infants. *Acta Paediatr.* 95 (10), 1239–1248. <https://doi.org/10.1080/08035250600599727>.
- Hobbs, T., Little, M., Kaoukji, D., 2007. Using the Strengths and Difficulties Questionnaire (SDQ) to measure the behavior and emotional health of children in schools in the United Kingdom. *Int. J. Child Fam. Welf.* 10 (3–4), 150–164. <https://uqgp.rug.nl/IJCFW/article/view/37774>.
- Hofstra, M.B., Van der Ende, J., Verhulst, F.C., 2002. Child and adolescent problems predict DSM-IV disorders in adulthood: a 14-year follow-up of a Dutch epidemiological sample. *J. Am. Acad. Child Adolesc. Psychiatry* 41 (2), 182–189. <https://doi.org/10.1097/00004583-200202000-00012>.
- Huizink, A.C., Mulder, E.J.H., 2006. Maternal smoking, drinking or cannabis use during pregnancy and neurobehavioral and cognitive functioning in human offspring. *Neurosci. Biobehav. Rev.* 30 (1), 24–41. <https://doi.org/10.1016/j.neubiorev.2005.04.005>.
- Jaekel, J., Sorg, C., Breeman, L., Baumann, N., Bilgin, A., Bäuml, J.G., Wolke, D., 2021. Early regulatory problems and parenting: life-long risk, vulnerability or susceptibility for attention, internalizing and externalizing outcomes? *Eur. Child Adolesc. Psychiatry* 30 (10), 1523–1531. <https://doi.org/10.1007/s00787-020-01632-2>.
- Jiang, C., Shi, J., 2024. Money or time? The association between parental investment, school engagement and adolescent behavioral problem. *Child Youth Serv. Rev.* 157, 107391. <https://doi.org/10.1016/j.childyouth.2023.107391>.
- Jiang, W., Ji, M., Chi, X., Sun, X., 2022. Relationship between adverse childhood experiences and mental health in Chinese adolescents: differences among girls and boys. *Children* 9 (5). <https://doi.org/10.3390/children9050689>.
- Khare, S.K., March, S., Barua, P.D., Gadre, V.M., Acharya, U.R., 2023. Application of data fusion for automated detection of children with developmental and mental disorders: a systematic review of the last decade. *Inform. Fusion* 99, 101898. <https://doi.org/10.1016/j.inffus.2023.101898>.
- Liu, X., Kurita, H., Guo, C., Miyake, Y., Ze, J., Cao, H., 1999. Prevalence and risk factors of behavioral and emotional problems among Chinese children aged 6 through 11 years. *J. Am. Acad. Child Adolesc. Psychiatry* 38 (6), 708–715. <https://doi.org/10.1097/00004583-199906000-00018>.
- Lopez, M.L., Tarullo, L.B., Forness, S.R., Boyce, C.A., 2000. Early identification and intervention: head start's response to mental health challenges. *Early Educ. Dev.* 11 (3), 265–282. [https://doi.org/10.1207/s15566935eed1103\\_3](https://doi.org/10.1207/s15566935eed1103_3).
- McGrath, L.M., Peterson, R.L., Pennington, B.F., 2020. The multiple deficit model: progress, problems, and prospects. *Sci. Stud. Read.* 24 (1), 7–13. <https://doi.org/10.1080/10888438.2019.1706180>.
- Modak, A., Ronghe, V., Gomase, K.P., 2023. The psychological benefits of breastfeeding: fostering maternal well-being and child development. *Cureus*. <https://doi.org/10.7759/cureus.46730>.
- Motaitianu, I.R., 2015. Parent-child connection — emotional synchronization and playing: a possible model to combat the child's unsafe attachment. *Procedia Soc. Behav. Sci.* 180, 1178–1183. <https://doi.org/10.1016/j.sbspro.2015.02.238>.
- Nelson, C.A., Bhutta, Z.A., Burke Harris, N., Danese, A., Samara, M., 2020. Adversity in childhood is linked to mental and physical health throughout life. *BMJ* m3048. <https://doi.org/10.1136/bmj.m3048>.
- Nomura, Y., Wickramaratne, P.J., Pilowsky, D.J., Newcorn, J.H., Bruder-Costello, B., Davey, C., Fifer, W.P., Brooks-Gunn, J., Weissman, M.M., 2007. Low birth weight and risk of affective disorders and selected medical illness in offspring at high and low risk for depression. *Compr. Psychiatry* 48 (5), 470–478. <https://doi.org/10.1016/j.comppsy.2007.04.005>.
- Oddy, W.H., Kendall, G.E., Li, J., Jacoby, P., Robinson, M., de Klerk, N.H., Silburn, S.R., Zubrick, S.R., Landau, L.I., Stanley, F.J., 2010. The long-term effects of breastfeeding on child and adolescent mental health: a pregnancy cohort study followed for 14 years. *J. Pediatr.* 156 (4), 568–574. <https://doi.org/10.1016/j.jpeds.2009.10.020>.
- O'Driscoll, D.N., McGovern, M., Greene, C.M., Molloy, E.J., 2018. Gender disparities in preterm neonatal outcomes. *Acta Paediatr.* 107 (9), 1494–1499. <https://doi.org/10.1111/apa.14390>.
- Ortuño-Sierra, J., Aritio-Solana, R., Fonseca-Pedrero, E., 2018. Mental health difficulties in children and adolescents: the study of the SDQ in the Spanish National Health Survey 2011–2012. *Psychiatry Res.* 259, 236–242. <https://doi.org/10.1016/j.psychres.2017.10.025>.
- Pant, S.W., Holstein, B.E., Ammitzbøll, J., Skovgaard, A.M., Pedersen, T.P., 2024. Community health nurses' concerns about infant regulatory problems are predictive of mental disorders diagnosed at hospital: a prospective cohort study. *Front. Child Adolesc. Psychiatry* 2. <https://doi.org/10.3389/frcha.2023.1330277>.
- Prior, M., Smart, D., Sanson, A., Oberklaid, F., 1993. Sex differences in psychological adjustment from infancy to 8 years. *J. Am. Acad. Child Adolesc. Psychiatry* 32 (2), 291–305. <https://doi.org/10.1097/00004583-199303000-00009>.
- Ramchandani, P.G., Stein, A., O'Connor, T.G., Heron, J., Murray, L., Evans, J., 2008. Depression in men in the postnatal period and later child psychopathology: a population cohort study. *J. Am. Acad. Child Adolesc. Psychiatry* 47 (4), 390–398. <https://doi.org/10.1097/CHI.0b013e31816429c2>.
- Rezapor, M., Hansen, L., 2022. A machine learning analysis of COVID-19 mental health data. *Sci. Rep.* 12 (1), 14965. <https://doi.org/10.1038/s41598-022-19314-1>.
- Rice, F., Jones, I., Thapar, A., 2007. The impact of gestational stress and prenatal growth on emotional problems in offspring: a review. *Acta Psychiatr. Scand.* 115 (3), 171–183. <https://doi.org/10.1111/j.1600-0447.2006.00895.x>.
- Rothberg, W.A., Bizzego, A., Esposito, G., Lansford, J.E., Al-Hassan, S.M., Bacchini, D., Bornstein, M.H., Chang, L., Deater-Deckard, K., Di Giunta, L., Dodge, K.A., Gurdal, S., Liu, Q., Long, Q., Oburu, P., Pastorelli, C., Skinner, A.T., Sorbring, E., Tapanya, S., Alampay, L.P., 2023. Predicting adolescent mental health outcomes across cultures: a machine learning approach. *J. Youth Adolesc.* 52 (8), 1595–1619. <https://doi.org/10.1007/s10964-023-01767-w>.
- Sellers, R., Maughan, B., Pickles, A., Thapar, A., Collishaw, S., 2015. Trends in parent- and teacher-rated emotional, conduct and ADHD problems and their impact in prepubertal children in Great Britain: 1999–2008. *J. Child Psychol. Psychiatry* 56 (1), 49–57. <https://doi.org/10.1111/jcpp.12273>.
- Singh, S., Gupta, H., Singh, P., Agrawal, A.P., 2022. Comparative analysis of machine learning models to predict depression, anxiety and stress. In: 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), pp. 1199–1203. <https://doi.org/10.1109/SMART55829.2022.10047752>.
- Smith, A.S., Tabbaa, M., Lei, K., Eastham, P., Butler, M.J., Linton, L., Altshuler, R., Liu, Y., Wang, Z., 2016. Local oxytocin tempers anxiety by activating GABA

- receptors in the hypothalamic paraventricular nucleus. *Psychoneuroendocrinology* 63, 50–58. <https://doi.org/10.1016/j.psyneuen.2015.09.017>.
- Sonego, M., Pichiule, M., Gandarillas, A., Polo, C., Ordoñas, M., 2018. Mental health in girls and boys exposed to intimate partner violence. *Public Health* 164, 26–29. <https://doi.org/10.1016/j.puhe.2018.07.003>.
- Stein, A., Pearson, R.M., Goodman, S.H., Rapa, E., Rahman, A., McCallum, M., Howard, L.M., Pariente, C.M., 2014. Effects of perinatal mental disorders on the fetus and child. *Lancet* 384 (9956), 1800–1819. [https://doi.org/10.1016/S0140-6736\(14\)61277-0](https://doi.org/10.1016/S0140-6736(14)61277-0).
- 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. <https://doi.org/10.1371/journal.pone.0230389>.
- Tennant, R., Goens, C., Barlow, J., Day, C., Stewart-Brown, S., 2007. A systematic review of reviews of interventions to promote mental health and prevent mental health problems in children and young people. *J. Public Ment. Health* 6 (1), 25–32. <https://doi.org/10.1108/17465729200700005>.
- Theunissen, G., D'Souza, S., Peterson, E.R., Walker, C., Morton, S.M.B., Waldie, K.E., 2022. Prenatal determinants of depressive symptoms in childhood: evidence from growing up in New Zealand. *J. Affect. Disord.* 302, 41–49. <https://doi.org/10.1016/j.jad.2022.01.076>.
- Thompson, J.M.D., Slykerman, R.F., Wall, C.R., Murphy, R., Mitchell, E.A., Waldie, K.E., 2021. Factor structure of the SDQ and longitudinal associations from pre-school to pre-teen in New Zealand. *PLoS One* 16 (3), e0247932. <https://doi.org/10.1371/journal.pone.0247932>.
- University of Essex, Institute for Social and Economic Research, 2023. Understanding Society: Waves 1–13, 2009–2022 and Harmonised BHPS: Waves 1–18, 1991–2009 [data collection]. UK Data Service. <https://doi.org/10.5255/UKDA-SN-6614-19>. SN: 6614.
- University of Essex, Institute for Social and Economic Research, 2024. Understanding Society: Pregnancy and Early Childhood (PEACH), 2009–2022 [data collection], 3rd edition. UK Data Service. <https://doi.org/10.5255/UKDA-SN-9075-3>. SN: 9075.
- Upadhyay, R.P., Naik, G., Choudhary, T.S., Chowdhury, R., Taneja, S., Bhandari, N., Martines, J.C., Bahl, R., Bhan, M.K., 2019. Cognitive and motor outcomes in children born low birth weight: a systematic review and meta-analysis of studies from South Asia. *BMC Pediatr.* 19 (1), 35. <https://doi.org/10.1186/s12887-019-1408-8>.
- Vaishnavi, K., Nikhitha Kamath, U., Ashwath Rao, B., Subba Reddy, N.V., 2022. Predicting mental health illness using machine learning algorithms. *J. Phys. Conf. Ser.* 2161 (1), 12021. <https://doi.org/10.1088/1742-6596/2161/1/012021>.
- Vater, A., Schröder-Abé, M., 2015. Explaining the link between personality and relationship satisfaction: emotion regulation and interpersonal behaviour in conflict discussions. *Eur. J. Personal.* 29 (2), 201–215. <https://doi.org/10.1002/per.1993>.
- Wallander, J.L., Berry, S., Carr, P.A., Peterson, E.R., Waldie, K.E., Marks, E., D'Souza, S., Morton, S.M.B., 2019. Patterns of exposure to cumulative risk through age 2 and associations with problem behaviors at age 4.5: evidence from growing up in New Zealand. *J. Abnorm. Child Psychol.* 47 (8), 1277–1288. <https://doi.org/10.1007/s10802-019-00521-w>.
- Wheelock, M.D., Hect, J.L., Hernandez-Andrade, E., Hassan, S.S., Romero, R., Eggebrecht, A.T., Thomason, M.E., 2019. Sex differences in functional connectivity during fetal brain development. *Dev. Cogn. Neurosci.* 36, 100632. <https://doi.org/10.1016/j.dcn.2019.100632>.
- Winston, R., Chicot, R., 2016. The importance of early bonding on the long-term mental health and resilience of children. *Land. J. Prim. Care* 8 (1), 12–14. <https://doi.org/10.1080/17571472.2015.1133012>.
- Zahn-Waxler, C., Shirtcliff, E.A., Marceau, K., 2008. Disorders of childhood and adolescence: gender and psychopathology. *Annu. Rev. Clin. Psychol.* 4, 275–303. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091358>.
- Zhang, C., Chen, X., Wang, S., Hu, J., Wang, C., Liu, X., 2021. Using CatBoost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry Res.* 306, 114261. <https://doi.org/10.1016/j.psychres.2021.114261>.