INFANT GUT MICROBIOME: A LONGITUDINAL ANALYSIS OF RELATIONSHIPS WITH FEEDING PRACTICES, AND BEHAVIOURAL DEVELOPMENT IN EARLY LIFE.

EMMA-LOUISE ALVING-JESSEP

Doctor of Philosophy

ASTON UNIVERSITY

DECEMBER 2022

For my family, Luc, Ryan, Scarlett, and James.

Thank you for being my strength and inspiration.

Personal Acknowledgements

First of all, I would like to thank Professor Jackie Blissett. To say this Ph.D. has been a bumpy ride would be a gross understatement. Navigating through data collection and research during a global pandemic seemed daunting and impossible at times, but your continued support guided me through. When all my plans went awry, you were there to guide me to the best possible outcome, pushing to get the best out of me. For that I will be forever grateful. I would not have been able to complete this without you.

I would also like to thank my associate supervisors Professor Amanda Wood, and Professor Anthony Hilton. Amanda, thank you for always playing devil's advocate and challenging me to question further. Anthony, thank you for re-introducing me to microbiology laboratory work, for listening to my in-depth discussion about how to handle faecal samples, and advising me about the best way to create synthetic infant poo.

To the members of the PEACh research group of Aston University, for welcoming me into the team and helping me feel at home. I could not have asked for a more friendly supportive group to belong to, and I am proud to be a part of the group as it grows. Additional thanks go, to Clare Langman, Library Services, Aston University, for advising me during the preparation of my Systematic Review.

To my fellow Ph.D. students in SW506, thank you for being some of the best friends I could have asked for. For listening to my discussions about baby poo, encouraging me to keep going when things were hard, and for being there when we were isolating during lockdown.

Finally, I would like to thank my family. My husband Luc for being there through every stressful event, and all the late-night statistics conversations. To my children, who have cheered me on every step of the way, and who have so often been the inspiration for my research.

Collaborator Acknowledgements

I would like to thank the Barwon Infant Study of the Deakin University and Murdoch Research Children's Institute Australia, for agreeing to collaborate with Jackie, Amanda, Anthony, and myself, on chapters 3-5 of this thesis I would like to thank Professor Peter Vuillermin for the continued support throughout the project. Special thanks also go to Dr Martin O'Hely, and Dr Loughman for fielding my very many questions about bioinformatic analysis.

I would also like to thank Dr. Edith Botchway for collaborating during the synthesis and quality assessment of papers collected for my Chapter two systematic review. Thank you for so generously giving your time, and for your company during our late night/ early morning meetings.

[signed - Author's signature removed from open access version of thesis]

ASTON UNIVERSITY

Infant gut microbiome: A longitudinal analysis of relationships with feeding practices, and behavioural development in early life.

Emma-Louise Alving-Jessep

Doctor of Philosophy Health and Life Sciences

ABSTRACT

Background: Previous research has established that there is an association between the composition and diversity of the gut microbiota (GM) and human behaviour. Current literature also highlights the links between dietary intake and behaviour. However, to date there is a paucity of literature investigating the complex relationship between all three aspects across early development.

Aims: This thesis aimed to i) investigate changes in diversity and composition of the GM during GM maturation, ii) examine how this is related to temperament and behavioural development throughout early childhood, and iii) explore how this relationship can be influenced by diet.

Methods: A systematic review was conducted investigating the influence of GM upon temperament during early childhood. Secondly, 324 participants recruited to the microbiota sub-cohort of the Barwon Infant Study (BIS), which focused on the role of specific environmental factors in early life development, provided faecal samples analysed using 16S rRNA illumine MiSeq to extract Amplicon Sequence Variant (ASV), and dietary information at 1-, 6-, and 12-months, and child behaviour was measured at 4-years.

Results: It was established that temperament during early childhood and behaviour measured at 4-years is significantly associated with GM composition and diversity. Furthermore, diet characterised by exclusivity of breastfeeding at 1-month of age, and by solid foods introduced at 6- and 12-months, was a significant moderator of the relationship between GM and behaviour measured at 4-years.

Discussion: This thesis has established a relationship between GM, temperament, and behaviour during early childhood. Furthermore, diet is established as a significant modifying factor of the relationship between GM and behaviour. Future research would benefit from using whole genome sequencing to investigate functional and taxonomic composition, addressing the issue of sample sizes, and improved study design by using concurrent GM, diet, and behavioural measures.

Keywords.

Child; Infant; Gut microbiota; Development; Temperament; Behaviour; Gut-brain axis; Diet; Solid Food Introduction; Breastfeeding.

CONTENTS

LIST OF ABBREVIATIONS

5-HTP – 5- Hydroxytryptophan

ADHD – Attention Deficit Hyperactivity Disorder

ANCOVA – Analysis of Covariance

ANS – Autonomic Nervous System

APrON - Alberta Pregnancy Outcomes and Nutrition

ASD – Autism Spectrum Disorder

ASQ3 – Ages and Stages Questionnaire 3rd edition

ASV – Amplicon Sequence Variant

BIS – Barwon Infant Study

BEBQ – Baby Eating Behaviour Questionnaire

BF – Breast Fed

CBQ – Child Behaviour Questionnaire

CBCL – Child Behaviour Checklist

CC – Conventionally Colonised

CEBQ – Child Eating Behaviour Questionnaire

CFI – Comparative Fit Index

CNS – Central Nervous System

DAG – Directed Acyclic Graph

ECBQ – Early Childhood Behaviour Questionnaire

EDS - Edinburgh Depression Scale

ENS – Enteric Nervous System

FDR – False Discovery Rate

FF – Formula Fed

FFQ – Food Frequency Questionnaire

GABA – Gamma Amino Butyric Acid

GBA – Gut-Brain Axis

GEMS - Genome-scale Metabolic network models

GF – Germ Free

GI – Gastro-intestinal

GM – Gut Microbiota

GO – Gene Ontology

HMO – Human Milk Oligosaccharides

HPA – Hypothalamic Pituitary Adrenal Axis

HTA – Human Tissue Act

IBQ – Infant Behaviour Questionnaire

IBQ-R – Infant Behaviour Questionnaire – Revised

IBQ-R-SF – Infant Behaviour Questionnaire – Revised Short Form

IBS – Irritable Bowel Syndrome

IFQ – Infant Feeding Questionnaire

KEGG – Kyoto Encyclopedia of Genes and Genomes

LefSE – Linear discriminant analysis of effect size

MAR – Missing At Random

MF – Mixed Feeding

NHLBI – The National Heart, Lung, and Blood Institute

OTU – Operational Taxonomic Units

PCA – Principal Component Analysis

PERMANOVA – Permutational multivariate analysis of variance

PPD – Postpartum Depressive Disorder

PPE – Personal Protective Equipment

PPI – Patient Participant Involvement

PSS - Perceived Stress Scale

RMSEA – Root Mean Square Error of Approximation

SCFA – Short Chain Fatty Acid

SDI – Shannon Diversity Index

SEIFA - Socio-Economic Indexes for Areas

SEM – Structural Equation Modelling

SDQ – Strengths and Difficulties Questionnaire

STSI – Short Temperament Scale for Infants

WHO – World Health Organisation

LIST OF APPENDICES

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 7

CHAPTER 1
INTRODUCTION: DEVELOPMENT OF THE GUT MICROBIOME, BEHAVIOUR, TEMPERAMENT AND THE INFLUENCE OF DIET

Behavioural development is known to be influenced by many factors including parenting, temperament, personality, and environmental factors such as school, culture, geographic location, and diet. Recently the gut microbiota (GM) has become a focal point of interest, specifically the relationship between changes in the composition and diversity of the GM and how this can influence both physical and mental health. However, the complex interplay between host, GM, and environment in the form of diet, has yet to be thoroughly explored. This narrative literature review aims to consider the development of the gut microbiota across the first year of life detailing how early feeding practices, including milk-based diet and solid food introduction, influence the development of the composition and diversity of the GM through its maturation. This literature review will then discuss the maturation of the GM and its role in the development of child temperament and behavioural outcomes during early childhood, including the development of problem behaviours. Additionally, this literature review will discuss the role of early feeding practices in the development of child temperament and behavioural outcomes during childhood, and finally the influence of diet upon the relationship between GM and behaviour will be discussed. This chapter will also present detailed aims and hypotheses of this thesis.

In recent years there has been growing interest in the gut microbiota and microbiome, and how variation in the diversity of microbial colonisation can impact both physical and mental health. A healthy colonisation of the gut microbiome promotes a healthy immune function and reduces risk of adverse outcomes such as asthma (Attar, 2015; Moossavi, Miliku, Sepehri, Khafipour, & Azad, 2018), type 2 diabetes (Upadhyaya & Banerjee, 2015) and obesity (de Vos, Tilg, Van Hul, & Cani, 2022; Ley, Turnbaugh, Klein, & Gordon, 2006; Turnbaugh & Gordon, 2009) Furthermore, increasing evidence is highlighting the role of the GM in influencing both behaviour and temperament. Temperament traits can be observed very early in development. These traits include an individual's emotional, attentional and motor tendencies and the ability to regulate themselves (Rothbart, 2007). Behaviour, for the purpose of this thesis, shall be defined as an individual's actions that they engage in, which can be performed either consciously or not (Morales, Amir, & Lee, 2017). In this thesis behavioural outcomes, referring to the set of behaviours that may develop, in part, or as a result of either dietary, or GM influence shall be explored. Furthermore, an in-depth exploration of the development of behavioural problems shall be conducted. Approximately 13-20% of children show a developmental trajectory that may lead to the development of problematic behaviours (Avenevoli et al., 2013; Ghandour et al., 2019). This can be classified as the development of internalising (anxiety, and depression), or externalising problems (hyperactivity, aggression, conduct) (Operto et al., 2021). The focus of this

thesis will be to describe the relationship between GM, temperament, and behaviour throughout early childhood and how this can be influenced by diet.

There has also been an increased awareness of an interaction between the gut and the brain which has been linked with susceptibility to depression, Attention Deficit Hyperactivity Disorder (ADHD) and many other mental health and psychological outcomes in adults (Adesman, Soled, & Rosen, 2017; Foster & McVey Neufeld, 2013; Rook, Raison, & Lowry, 2018). There is increasing evidence that alterations in the composition of the gut microbiome can be linked to behavioural outcomes (Loughman et al., 2020). This mechanism of communication is now known as the gut-brain axis (GBA), a bidirectional communication route between the central nervous system (CNS) and the enteric nervous system (ENS) (Wang et al., 2018). Recent evidence has even shown that there is the potential for microbes that reside within the intestinal mucosa of the gut communicate with the neural mitochondria of the brain, highlighting another potential mechanism of communication of the GBA (Zhu et al., 2022). The gut microbiome is influenced from birth and throughout the lifespan by many factors such as birth method, geographic location, urbanicity, antibiotic usage, and dietary intake (Rodriguez et al., 2015). Further realisation of the diversity and scale of the gut microbiome has given us greater understanding of the relationship between ourselves as hosts and the microbial colonies. In 2004, it was reported that there are over 3.3 million non-redundant genes present in the human gut microbiome, which makes the approximately 22,000 genes in our own genome seem small by comparison (International Human Genome Sequencing Consortium). The scale of the potential influence that the gut microbiome has upon our health has led many to name it as an organ in its own right (Jiménez et al., 2008; Perez-Munoz, Arrieta, Ramer-Tait, & Walter, 2017) and led to a proliferation of research in this area, posing the questions of to what extent the gut microbiome impacts our physical health, mental health, and behaviour, and, in turn, what influences it is possible to have upon the gut microbiome.

Interest in early life influences that impact the development of the gut microbiome has grown considerably in the last 10-years. It was thought that infants' intestinal tracts, or guts, were born sterile, completely absent of colonisation by microorganisms. First colonisation was thought to occur through contact with bacteria during vaginal delivery and subsequent contact with the mother, including breastfeeding (Mackie, Sghir, & Gaskins, 1999; Palmer, Bik, DiGiulio, Relman, & Brown, 2007). This theory has since been challenged as studies of infant meconium (the first infant stool) provide evidence that the colonisation of the infant gut begins *in utero* (Stewart et al., 2018a). It is, however, known that the mode of birth significantly impacts the colonisation of the gut microbiome (Ma et al., 2023). It was also thought that mode of birth would significantly affect incidence health concerns specifically associated with the gastrointestinal tract and the gut microbiota. In a study of over 1 million individuals carried out in Sweden, it was found that there is a link between caesarean section and

Crohn's disease, diverticulosis, and cholecystitis, which suggests that mode of birth significantly impacts the health of the gut, which has been strongly linked to gut microbiota composition (Hellsing et al., 2022). During a normal vaginal birth, the infant is predominantly born facing towards the dorsal or rear of the mother, which allows the face, in particular the mouth, to come into contact with the maternal vaginal microbiome and also the gut microbiome through contact with faecal matter. A caesarean section will bypass this contact and set the infant on a different trajectory of bacterial colonisation and maturation of the gut microbiota (Korpela, 2021). During maturation the GM will undergo many changes throughout the first years of life until reaching maturity at approximately two to three years old (Carlson et al., 2018; Hoskinson et al., 2023; Knickmeyer et al., 2008). Recent evidence has identified three distinct phases of microbiome progression, the developmental phase between 3-14 postnatal months, transitional phase 15-30 months, and the stable phase at 31-46-months (Knickmeyer et al., 2008; Roswall et al., 2021; Stewart et al., 2018b). Once the GM reaches maturation it remains relatively stable and in healthy individuals will remain in this stable state for the majority of adulthood (Backhed et al., 2012). In a recent study Olsson et al. (2022) investigated the dynamics of normal gut microbiota development in infants. They found that 23% of the total compositional variance was explained by individual differences between people. These individual differences in gut microbiota development may be explained by genetic factors, but also by the previously mentioned environmental factors, such as urbanicity, family environment including number of siblings as well as the diet, which is the focus of this chapter. Further discussion of the maturation of the GM from birth will be addressed in sections 1.1 – 1.5 with reference to the influence of maternal feeding practices, milk-based diet, and solid food introduction, and the impact of GM diversity and composition upon and temperament and behavioural development.

It has been established that breastfeeding facilitates a healthy GM, however, there has been no investigation into the relationships between the developing GM across the first year of life, behavioural development, and dietary intake. The first year of infant life is a period of dramatic change; not only is this a period of rapid physical growth (Weaver, 2006), this is also a critical period of dynamic brain development and neurodevelopment, i.e. structural and functional maturation, respectively (Carlson et al., 2018; Cowan, Dinan, & Cryan, 2020; Knickmeyer et al., 2008). Disturbances to the development of the brain in the first 2-years of life are associated with a number of neurodevelopmental disorders (Knickmeyer et al., 2008). Evidence of the involvement of the gut microbiome and the gut-brain axis in brain development and neurodevelopmental outcomes, is only starting to emerge (Gao et al., 2019). Investigation into maternal feeding practices, the introduction of solid food, and the diet established during the first year of life is essential in order to understand the effects and potential mechanisms by which diet choices influence GM and thus behavioural development. Establishing these relationships will contribute to understanding of the interaction between diet, microbiome and behavioural outcomes and may lead to development of interventions in the future.

1.1 Gut microbiome and early maternal feeding practices

1.1.1 Early feeding practices and milk-based diet

Parents or primary care givers are responsible for nurturing infants, ensuring that they receive appropriate care to enable the infant to progress along a healthy developmental pathway. The nutritional requirements of infants, like many aspects of development in the first postnatal year, undergo large changes, from feeding at all times of the day and night, to establishing routine meal times (Grummer-Strawn, Scanlon, & Fein, 2008). There are also several decisions to be made with regard to infant feeding, including the decision to breast or formula feed, whether to supplement, when and how to begin the process of solid food introduction and the process of breastfeeding cessation.

Early maternal feeding practices have been associated with a number of health outcomes as a child matures into adulthood, these include appetite and susceptibility to obesity (Haszard, Russell, Byrne, Taylor, & Campbell, 2019). Although it is often the case that both parents are significantly involved in parenting of a child, it is commonly accepted that mothers are more often the caregiver most involved in the care of infants and food provision. In a study carried out by Patrick, Nicklas, Hughes, and Morales (2005), investigating caregiver feeding styles, they found that the mother was identified as the primary caregiver in 92% of cases. Therefore, for the purpose of this study, the term maternal feeding practices has been adopted.

There is an abundance of research that highlights the benefits of breastfeeding infants. Since 2011, the World Health Organisation (WHO) has advised new mothers to exclusively breastfeed for the first 6-months of infant life (Kramer & Kakuma, 2012). The benefits of breastfeeding can be seen in psychological, behavioural and neurological development of the infant (Golding, Emmett, & Rogers, 1997). There is also extensive research into the role of breastfeeding in reduction of the risk of obesity. A meta-analysis performed by Arenz, Rückerl, Koletzko, and von Kries (2004) investigated 9 studies with over 69,000 individuals with evidence that breastfeeding protects infants against obesity in later life, although effects were small, they were consistent. Additionally, Harder, Bergmann, Kallischnigg, and Plagemann (2005), analysed evidence of the effect of duration of breast feeding upon risk of obesity and concluded that they are inversely related, the longer before breastfeeding cessation, the lower the risk of obesity later in life. Although it is understood that breastfeeding has numerous benefits for infants, there are many occasions when parents choose to formula feed. These can include maternal stress, illness, maternal physical difficulties breastfeeding (including breastmilk production and let-down), distress, tiredness, and problematic infant feeding behaviour such as poor latching (Cloherty, Alexander, & Holloway, 2004).

1.1.2 Milk-based diet and the influence upon GM maturation

Human breastmilk is a complex fluid, which undergoes compositional change to provide the optimal immunological support and nutritional requirements for infants as they mature (Bardanzellu, Fanos, & Reali, 2018). Not only does breastmilk develop along with infant gestational age, but the composition will vary with lactation phase (Anatolitou, 2012; Sundekilde et al., 2016). The major constituents of breastmilk are water, fats, proteins, immune factors, and the carbohydrates lactose and oligosaccharides (Miller et al., 2013). The composition of breastmilk not only plays an important role in the physiological growth and health of the infant (Plaza-Diaz, Fontana, & Gil, 2018), but also serves as a substrate, which specifically caters for colonisation of the infant gut by bacteria that promote good health. Breastmilk also alters in composition reflecting developmental stages of the infant. Immediately following birth, the milk produced, colostrum, is low in fat but high in proteins and specific immune factors, serving as a high energy source able to bolster the infant's immune system (Ogra, Rassin, & Garofalo, 2006). Following the production of colostrum, mothers will start to produce transitional milk, which increases in fat content as the infant is able to digest more complex nutrients. This change coincides with the initial stage of colonisation of the GM (Martin & Sela, 2013). Mature breastmilk is produced from 3-4 weeks onwards and will increase in volume, meeting the needs of the infant until introduction of solid food (Ogra et al., 2006). It is important to capture the very early variation in composition of the GM and how this is shaped by dietary intake, as human breastmilk is specifically designed for optimal development of the infant.

Human milk oligosaccharides (HMO), which are sugars, a very simple form of carbohydrate, are the third most abundant component of breastmilk (Fanos, Reali, Marcialis, & Bardanzellu, 2018), which varies greatly between individuals and across the course of lactation from birth to cessation (Chaturvedi et al., 2001). Additional factors such as ethnicity, geographic location and breastfeeding exclusivity also influence the HMO element of breastmilk, influencing the concentration of HMOs present in the breastmilk (McGuire et al., 2017). Specifically, there were no differences in concentrations between urban and rural settings, however there were significantly lower concentrations of HMOs in women in Ghana. Interestingly, in breastfed infants, the reach of HMOs is not limited to the infant enteric system, approximately 1% are absorbed into the peripheral bloodstream and are potentially distributed to all organs in the body (Bode, 2015). HMOs promote the growth of *Bifidobacterium* and *Bacteroides* (Cong et al., 2016), with a diverse colonisation of *Bifidobacteria* species typically seen, as they are specifically able to break down and metabolise HMOs. Other species, such as pathogenic Enterobacteriaceae*,* are unable to metabolise HMOs (Cong et al., 2016) and therefore are less able to colonise. Competitive colonisation and the ability, or inability, to metabolise HMOs results in the predominant *Bifidobacteria* profile often seen in breastfed infants (Turroni et al., 2011). Competitive colonisation of the immature infant GM can be essentially

thought of as a race. Multiple bacteria are introduced through the birth process, and exposure to the world, through ingestion. These bacteria then compete with each other to occupy their niche environment, competing for resources, and vying for domination in order to be the most reproductively successful. *Bifidobacteria* also significantly influences the GM environment to ensure dominance in colonisation of the GM. Through its metabolic processes *Bifidobacteria* produces acetate and lactate, which lowers the pH of the environment, in doing so this makes the environment less hospitable for other genera of bacteria including potentially pathogenic species (Frese et al., 2017; Henrick et al., 2018). Further to the HMOs contained within human breastmilk, other constituents play an important role in infant health, milk fat globules have recently been explored for their interaction with probiotic bacteria, specifically influencing their ability to colonise the infant gut in a more effective and efficient way (Yadav, Kapoor, Verma, & Ambatipudi, 2022). As Yadav et al. (2022) further note there are many more constituents, such as the milk extracellular vesicles and immunoglobulins, that are being further explored for their impact upon the composition and health of the infant gut microbiota. However, this work is in the beginning stages and full comprehension of the complex interaction between breastmilk and infant gut microbiota is yet to be fully discerned.

Breastfed (BF) infants undergo large changes in GM from birth to breast feeding cessation, at which point the alpha diversity of the GM changes from being lower to higher than those who were exclusively Formula Fed (FF) (Davis, Wang, & Donovan, 2017). A study by Carlson et al. (2018) carried out an investigation into clustering of GM in both FF and BF infants. They found that for FF infants, colonisation of the gut can be clustered into groups that resemble adult enterotypes, with dominance of *Faecalibacterium* and Ruminococcaceae*.* Furthermore, infants who were still breastfed at 1-year of age clustered into a separate group, which is dominated by Bacteroides, providing initial evidence of GM maturation driven by the duration of breastfeeding and subsequent cessation. In breastfed infants, the shift away from breastfeeding results in the overall lowering of the relative abundance of the *Bifidobacterium*, *Lactobacilli* and *Enterococcae* groups, which serve to protect the health of the child. Interestingly in the longitudinal study investigating compositional dynamics of the infant gut microbiota carried out by Olsson et al. (2022) *Fecalibacterium prausnitzii,* and two species of *Bifidobacterium* were the species predominantly linked to intra-individual variability in microbiota composition. In contrast, the *Bacteroides* and *Firmicutes* increase in relative abundance (Davis, Wang, & Donovan, 2017; O'Callaghan & van Sinderen, 2016; Voreades, Kozil, & Weir, 2014). Current evidence indicates that it is the cessation of breastfeeding, rather than the introduction of solid food, that triggers the transition to an adult like GM in breastfed infants (Aatsinki et al., 2023; Koenig et al., 2011; John Penders & van Best, 2022; Stewart et al., 2018a; Valles et al., 2014). Stewart et al. (2018a), investigated the maturation of the GM of children between the ages of 3-, and 46-months. Using metagenomics sequencing techniques, they were able to cluster microbiome by relative abundance of bacterial taxa and relate this to current infant diet. Infants who were introduced to solid

food but were also still receiving breastmilk had overall lower alpha diversity and a dominance of *Bifidobacteria* in their faecal samples, closely resembling the profile of exclusively breastfed infants. As the proportion of breast milk feeding reduced then the profile of the GM changed, moving away from dominance of *Bifidobacteria,* to an increase in diversity and an increase in the phylum Firmicutes*.* Formula-fed infant profiles began to cluster towards the adult enterotypes sooner than those of infants receiving any level of breastmilk. In a study carried out by Carlson et al. (2018), the GM was investigated at 1-year of age. It was found that the GM already clusters into groups defined by their microbial composition, in terms of predominance of *Bacteroides, Fecalibacterium,* or Ruminococcaceae*.* Two of these groups, *Bacteroides* and Ruminococcaceae, closely resemble the adult enterotypes supported by previous studies (Arumugam et al., 2011; Falony et al., 2016). For the group *Bacteroides,* the predominant factor that distinguished this group was infants who were receiving breastfeeding regardless of whether they were already introduced to complementary feeding, which further supports the idea that breastfeeding delays maturation of the infant gut microbiota. Stewart et al. (2018a) further confirms that the maturation of the GM is understood to be initiated by the cessation of breastfeeding, as they investigated maturation in several time points through infancy, and those who were still receiving breastfeeding presented with a GM that was yet to start the maturation process, regardless of whether they had been introduced to solid food. This contradicts the previous view that the introduction of solid food was the driving force of transition to the stable mature microbiome. Since this study several others have since also asserted that it is the cessation of breastfeeding that is the driving force behind the maturation of the GM, and that earlier cessation speeds up this process (Aatsinki et al., 2023; Penders & van Best, 2022)

Formula feeding and supplementary feeding impacts the composition of GM from as early as birth and may lead to changes in developmental trajectories. Although infant formulas strive to achieve the same nutritional value and physiological functions as breastmilk, this is a highly complex task to achieve. Studies of faecal samples taken from neonates have shown that the GM is initially colonised by facultative anaerobes (Jost, Lacroix, Braegger, & Chassard, 2012), organisms that will grow in either the presence or absence of oxygen, unlike obligate anaerobes that require the absence of oxygen to thrive. These facultative anaerobes serve to prepare the gastro-intestinal (GI) tract for colonisation by obligate anaerobes such as *Bifidobacterium* (Fan et al., 2013). A number of studies have identified variation in both alpha-diversity (the average species diversity within any environment) and bacterial-composition of the GM between breastfed and formula-fed infants (Azad, Becker, et al., 2013; Penders et al., 2006; Wang et al., 2015). Differences in GM colonisation occur as early as the first week of life, with breastfed (BF) infants showing lower levels of alpha diversity than formula-fed (FF) infants (Fan et al., 2013). A study by Azad et al., (2013) found that formula fed infants presented with an increased richness of species with increased relative abundance of *C.dificile.* Breastfed infants present with increased relative abundance of *Lactobacillus* and Bacteroides (Stewart et al., 2018a).

Recent research into GM development across the first year of life has shown that type of milk feeding (breast/formula feeding) explains a significant amount of variance in the infant gut colonisation (Adlerberth & Wold, 2009; Davis et al., 2017). However, progress is continuing and recent developments in the production of infant formula have led to a formula balanced in HMOs, which has been found to shift the colonisation of the infant gut microbiome towards that of a breastfed infant (Bosheva et al., 2022). This also highlights the important influence that HMOs have upon the composition of infant gut microbiome.

Many studies have focused on the genus *Bifidobacterium*, which shows significantly higher levels of abundance in BF vs FF infants (Adlerberth & Wold, 2009; Davis et al., 2017). BF infants, however, also show significantly lower levels of *Bacteroides and Faecalibacterium* in the phyla Bacteroidetes and Firmicutes than FF infants (Gomez-Llorente et al., 2013; Thompson, Monteagudo-Mera, Cadenas, Lampl, & Azcarate-Peril, 2015). FF infants show a greater prevalence of *C.difficile* and *E. coli* than BF infants (Azad, Konya, et al., 2013), although evidence for *E.coli* presence in both FF and BF infants is inconsistent (Davis et al., 2017). FF infants have been shown to initially have an GM with higher levels of alpha diversity, dominated by facultative and obligate anaerobes, although composition remains relatively stable in comparison to BF infants (Cresci & Bawden, 2015). Research into the GM of infants who receive a combination of both breast and formula feeding, known as supplemented breast feeding or Mixed Feeding (MF), is complicated due to the variability of diet between infants. Both the quantity and timing of introduction of formula to the breastmilk diet can vary widely and therefore accurate recording of nutritional intake can be unreliable (Davis et al., 2017). MF infants have high levels of within group variability in GM composition (Madan et al., 2016). In one study, infants who received more than 50% supplementation during the first three months of life showed significantly different GM composition than those who received less than 50% (Rinne, Kalliomaki, Arvilommi, Salminen, & Isolauri, 2005). Several studies have demonstrated that MF infants predominantly show GM composition resembling that of FF infants rather than BF infants (Penders et al., 2006). In a recent study it was found that there is a distinct difference however between the composition of the gut microbiota of those receiving mixed feeding compared to exclusive breastfeeding within the first 30 days of life. With breastfeeding influencing the composition of the gut microbiota, not necessarily through direct strain transmission, but rather through the SCFA, where one taxa of bacteria is influencing the metabolism of another through the process of cross-feeding (Li et al., 2022). There is a distinct gap in knowledge that reflects the need to investigate how supplementation of breastfeeding with formula feeding influences the development and maturation of the GM. Additionally, although many studies focus on the differences between BF and FF infants, there is less evidence of the maturation process of the GM in formula fed infants. This leads to the questions of whether formula fed infants reach a mature GM at an earlier age, whether maturation in

these infants is driven by solid food introduction and whether formula composition influences GM maturation.

1.2 The infant gut microbiome and introduction of solid food.

1.2.1 Factors influencing the introduction of solid food.

The introduction of solid food into the infant diet of breast or formula fed infants, known as complementary feeding, is a process that influences great change in the GM (Kashtanova et al., 2016). In addition to the introduction of solid food, the process of breastmilk or formula feeding reduction will usually begin. The WHO recommends that complementary feeding is introduced to infants no younger than 6-months of age, and that breastfeeding children have the right to safe and nutritious foods as a human right (Organization, 2017). In Europe, it is recommended that parents can choose to introduce solid foods to infants from 4-6-months, which reflects how recommendations can vary from the global WHO recommendations (Agostoni et al., 2008). In reality, it is often the case that adherence to recommendations is poor and children younger than 4 months are already introduced to some form of solid food (Arden, 2010; Burgess et al., 2019). A study of infant feeding practices in the UK carried out in 2000 found that 85% of mothers had introduced solid food to their infant before the age of 4 months (Hamlyn, 2002). This led to the adoption of the WHO recommendation by the UK Department of Health (UK Department of Health, 2003). By 2005, the number of infants being introduced to solid food before 4 months dropped significantly to 51%, however the number receiving solid food before the WHO recommended age of 6-months was still at 98% (Bolling, Grant, Hamlyn, & Thornton, 2007). Results of the UK-based Infant Feeding Survey 2010 (McAndrew et al., 2012) show that nationwide the number of infants being fed solid food by 4-months dropped to 31%, with 75% introduced by 5 months. However, still only 6% of mothers waited until the after recommended 6-months to begin introduction of solid foods. The maturation process of the GM in formula fed children is more complicated as this is dependent upon the composition of the formula that they were given (Marques et al., 2010; Oozeer et al., 2013). Modern formulas often contain probiotics as well as sugars that attempt to mimic the oligosaccharides contained within the human breast milk. HMO replacements are derived from plant-based sugars and are in the form of either galactooligosaccharides or fructooligosaccharides (Boehm & Stahl, 2007). Whilst this is not an exact like-for-like substitution of HMOs, it does influence the GM in a positive way and may influence the way in which the GM matures in formula fed infants (Oozeer et al., 2013). Furthermore, it is thought that introduction to solid food has larger influence upon the GM of formula fed infants compared to infants who are still receiving any breastmilk, this includes both exclusively breastfed infants and mixed feeding.

There are a variety reasons that prompt mothers to introduce solid foods at a younger age, despite awareness of the WHO recommendation. These include the belief that the child needs more substantial food, peer pressure amongst new mothers to achieve solid food introduction earlier, confusion with food labelling suggesting suitability from 4-6-months and a number of cues from the baby such as curiosity and interest in adult food (Arden, 2010). Overall, Arden (2010) found that there is a conflict between the importance given to the recommendation and signs that parents perceive from the baby that they are ready for solid food at an earlier age. Although the infants may indicate that they are ready for solid food behaviourally, it is not necessarily true that they are ready for solid food physiologically. Prior to 6- months the gastrointestinal tract is immature, and early introduction of solid food may have a negative health impact, (discussed in greater detail at the end of this section). Infant gender has also been found to influence the decision to introduce solid food earlier, with male infants being perceived as hungrier and requiring more energy than females (Brown & Rowan, 2016). Infants who are higher in birthweight were also more likely to be introduced earlier, which concurred with the work of Huh, Rifas-Shiman, Taveras, Oken, and Gillman (2011) who found that larger infants are perceived to need solid food earlier by their mothers. Additionally, infant temperament may play an important role in the decision to introduce solid food. Rogers and Blissett (2019), found that infants whose mothers perceive them to smile and laugh more readily are more likely to introduce solid food at an earlier age. Given that temperament may influence, and be influenced by, early feeding practices, it is essential to further understand the relationship between GM and infant temperament across the first year of life. This need arises due to the bi-directional nature of the GBA, the specific nature of the relationship is not yet clear. It may be that the gut microbiome influences an individual's temperament through the GBA, or that an individual's temperament may lead to the individual's altered response to the environment around them, which can subsequently lead to influences upon the GM.

Early introduction of solid food affects subsequent development of the infant GM, (Bergstrom et al., 2014), specifically, the younger the infant at the time of solid food introduction, the quicker the GM resembles a stable adult-like state. One potential interpretation of this trajectory is that a more rapid maturation toward the adult like GM is beneficial to infants, however studies investigating introduction of solid food earlier than the recommended age have found that there are negative health implications for the infant. For example, Wang et al. (2016), found that there is an increased risk of obesity for children aged between 2-, and 12-years if they were introduced to complementary feeding before the age of 4-months. Interestingly, Papoutsou et al. (2018) found that timing of solid food introduction was associated with obesity, but this was also dependent upon breast/formula feeding practices. The relationship is highly dependent upon timing of breastfeeding cessation and complex. Infants that were exclusively breastfed for longer than 6-months, with a later introduction to solid foods, had increased risk of obesity compared to those who received solid foods at 6-months of age.

Interestingly, early solid food introduction also protected against obesity in infants who experienced cessation of breastfeeding prior to 4 months. Finally, those children who received solid food at 6-months of age, with continued breastfeeding for up to 12-months of age were significantly less likely to be at risk of obesity. This indicates that solid food should be introduced no later than 6-months of age in children still receiving breastmilk, and for those who cease breastfeeding early, an earlier time of solid food introduction has a protective influence against obesity. However, the underlying biological mechanisms by which breastfeeding, and timing of solid food introduction impacts the risk of obesity has yet to be understood, and may further be explained by the changes and maturation of the GM.

1.2.2 The influence of diet types upon GM maturation

In addition to the timing of solid food introduction and cessation of breastfeeding/formula feeding, the type of food introduced to the infant diet also plays a significant role in the diversity and characteristics of the GM. Several studies have investigated the impact of diets introduced to infants and the enterotypes that emerge during maturation in order to establish which enterotype can be associated with a "healthy" microbiome. The phrase healthy in terms of microbiome is site specific, in this case the environment of the gut, showing stability or resistance to change, low levels of pathogenic species, and a community structure that functions to promote both the physiological and mental wellbeing of the host (Backhed et al., 2012). There are three adult enterotypes that have been identified with a dominance of *Bacteroides, Prevotella* or *Ruminococcus* (Kashtanova et al., 2016). These enterotypes are heavily dependent upon the diet that is introduced during childhood, which in turn is influenced by country of origin and culture. De Filippo et al. (2010), found that the proportion of *Firmicutes and Bacteroidetes,* are strongly influenced by variation in the diet between Western developed countries and rural areas in West African countries. They compared children, between 1- and 6- years of age, raised on the typical Western diet, which is associated with higher levels of animal protein, sugar, starch and fat intake and lower levels of dietary fibre, to that of children from a rural village in Burkina Faso, whose diets contained low fat and animal protein and high levels of fibre and plant polysaccharides. Analysis of the gut microbiome showed that the bacterial phyla, *Actinobacteria* and *Bacteroidetes* had lower abundance in the children from Africa compared to the children from Europe. These bacteria have been associated with a high-fat diet when investigated in animal models (Wu et al., 2011). In particular, they found that *Prevotella* and *Xylanibacter* colonisation was present in the gut microbiome of children brought up in Africa, and completely absent in those brought up in Italy. Although this highlights the potential influence that diet has upon gut microbiome composition, it is not possible to ascertain at which point in the development of the child the divergence of microbiome composition occurs. Furthermore, without more in-depth investigation it is not possible to ascertain the other potential influences at play, such as ethnicity, which may influence GM composition through genetic variation. It is for these reasons that it is necessary to explore further the introduction of

complementary feeding and its impact upon GM development before stabilisation to the adult enterotype. Additionally, environmental factors may play a role in shaping the compositional differences seen between these two groups. It is known that sanitised environments and levels of urbanisation reduces the alpha diversity of the gut microbiome (Rampelli et al., 2015). Finally, dietary changes that occur as a result of seasonal changes may have a greater impact upon countries with greater variation in temperature (Davenport et al., 2014).

Early introduction of fibrous plant material into the infant diet allows for a group of bacteria that produce butyrate to increase in number (Voreades et al., 2014). These bacteria metabolise complex polysaccharides, such as starches and complex carbohydrates found in plant material, that are resistant to digestion by human enzymes. Here a symbiotic relationship is formed with the host to allow for the digestion of complex carbohydrates, as a biproduct of this metabolic process they produce a Short Chain Fatty Acid (SCFA) that interacts with the GBA, known as butyrate, giving this group of bacteria the general name of butyrate producing bacteria. There are several beneficial intestinal effects of the production of butyrate including decreased pH of the colon (which decreases risk of pathogenic colonisation), improvement of gut barrier function, reduction of inflammation and promotion of satiety (Riviere, Selak, Lantin, Leroy, & De Vuyst, 2016). When complex plant-based material is introduced to the infant at a young age, the digestive system is unable to fully digest the material. A mature pancreas exocrine system is responsible for excreting amylase into the duodenum of the small intestine to breakdown the complex carbohydrates before they reach the colon. Due to the immaturity of the infant pancreas exocrine state the digestive system is unable to begin the process of breaking down the plant material, allowing many complex carbohydrates to enter directly into the colon (Fallani et al., 2011b). This allows the butyrate producing bacteria to establish and colonise the GM. From this information it would appear that early introduction of these butyrate producing bacteria would be beneficial, however, this may lead to an over colonisation by this group and disrupt the balance of the GM. Later introduction allows for further maturation of the infant digestive system so that colonisation by commensal species occurs at the optimal time. Breastfeeding protects the immature infant digestive system from over colonisation by this group. The longer that infants are either breast or formula fed the lower the relative abundance of the butyrate producing bacteria (Voreades et al., 2014). As complementary feeding progresses, the digestive system matures, with increased pancreatic function and small intestine absorption (Parrett & Edwards, 1997). This reduces the amount of undigested carbohydrates reaching the site of the large intestine and contributes towards stability of the GM.

In addition to plant material, investigation of the introduction of protein into to the complementary feeding diet of infants has increased in recent years. Prior to solid food introduction protein intake in formula fed infants has been associated with greater weight gain trajectories when compared to

breastfed infants (Luque, Closa-Monasterolo, Escribano, & Ferre, 2015). The high protein content associated with the Western diet, is associated with a gut microbiome dominated by *Bacteroides* (Arumugam et al., 2011). A number of dietary factors have been found to influence the microbiome in adults. This includes protein type; diets rich in protein from chicken and fish are associated with increased abundance of *Bifidobacterium* and *Bacteroides,* whereas a beef rich diet is associated with increased relative abundance of *C.perfingens* (Shen, Chen, & Tuohy, 2010). However, investigations into the western style diet have only been substantiated in populations after the expected point of maturation, and therefore further investigation is necessary to understand the way in which protein content during solid food introduction can influence the maturation process of the GM.

Another component of the diet that is often investigated is dietary fat. Dietary fat is predominantly absorbed in the small intestine, with only approximately 7% reaching the large intestine to be excreted with faeces (Gabert et al., 2011). High fat diets in adults are inversely related to the relative abundance of both SCFA, a biproduct of bacterial metabolic processes, and *Bifidobacteria* (Brinkworth, Noakes, Clifton, & Bird, 2009*).* Of particular interest is a diet high in palm oil, which has been shown to shift the microbiome profile to one high in relative abundance of *Firmicutes* and *Clostridium* (Riviere et al., 2016). It has been found in animal models that high fat diets can have a number of effects upon the gut due to the effect upon the microbiome, including increased inflammation (Benoit et al., 2015) and increased bile-acid production, which can be metabolised to produce carcinogenic secondary bile acids (Ou, DeLany, Zhang, Sharma, & O'Keefe, 2012). A study of the effect of High Fat Diet consumption upon health in mice has shown that there is a significant increase in instance of colorectal cancer when fat intake is increased (Yang et al., 2022). Specifically, it was found that impaired gut barrier function was related to dysbiosis of the GM and further cancer incidence. Although researchers are becoming increasingly aware of the effect of diet upon the gut microbiome, the studies mentioned above focus on adult gut microbiomes. There is very little research into the effect of the timing of first introduction to solid food, differences in food choices, and the effect it has upon the microbiome as it develops in the first year of life. It is an important step therefore to investigate differences in diet at this formative stage of life.

It can be seen from the information above that the introduction of complementary feeding and cessation of breast/formula feeding is a time of great change for the GM. The enterotype that is shaped during this time may have an impact upon the health and developmental trajectories of a child. Further impact of dietary intake, upon behavioural development in relation to the development of the GM will be discussed in the following sections.

1.3 The influence of the gut microbiome upon temperament and behaviour development.

In order to explore the role of the gut in behavioural development, it is first essential to understand the means by which the gut can influence the host and the way in which the gut and the brain communicate. The gut-brain axis is a system of bidirectional communication, which is known to use mechanisms involving the central nervous system , enteric nervous system and the hypothalamic pituitary adrenal axis (HPA) (Skonieczna-Żydecka, Marlicz, Misera, Koulaouzidis, & Łoniewski, 2018). The HPA axis is a system in the body that is involved in the stress response of humans and animals and allows for the processing of stressors in the environment, allowing the body to adapt and return to homeostasis. This is achieved through a series of endocrine pathways, involving the hypothalamus, the adrenal gland and the pituitary gland (Sheng et al., 2021). Currently, evidence is emerging to support the notion of a gut microbiome GBA (Carabotti, Scirocco, Maselli, & Severi, 2015). The mechanism by which the CNS and ENS communicate involves many processes. The autonomic nervous system (ANS) (in particular, the brain and spinal cord) communicates with the ENS using both neural and endocrine pathways. This is achieved through activation of the sympathetic ANS and HPA, which increases permeability of the intestine (Marlicz et al., 2018). The ENS communicates with the CNS through activation of the vagus nerve and biochemical transmission of cytokines, chemokines, and other neurotransmitters (Paolicelli, Bergamini, & Rajendran, 2019). Evidence has shown that the by-products of microbiological metabolic activity enter into circulation and influence both the GBA and HPA. This is inclusive of molecules such as gamma aminobutyric acid (GABA), 5-Hydroxytryptophan (5-HTP), and SCFAs (Dinan, Stilling, Stanton, & Cryan, 2015). In adults, these metabolic by-products are essential for the normal functioning of both of these systems, and therefore these bacteria work in a positive symbiotic relationship with the host.

Animal models have shown that dysregulation of the gut microbiome is associated with lasting impact upon brain chemistry, affecting stress response, cognition and behaviour relating to anxiety and depression (Cryan & Dinan, 2012; Kelly et al., 2016; Zheng et al., 2016). Additionally, animal studies investigating the administration of probiotics, specifically *Lactobacillus plantarum, L. rhamnosus,* and *B. longum,* have shown that use of these bacteria can have a positive influence upon the composition of the GM and can have beneficial anti-anxiety and anti-depressive effects (Abildgaard, Elfving, Hokland, Wegener, & Lund, 2017; Barros-Santos et al., 2020), although further investigation into the use of probiotics is still necessary in both animals and humans before the mechanism of influence can be truly established. In humans, the HPA axis (which plays an important role in emotional regulation and stress response) has been shown to be affected by gut microbiome dysregulation (de Weerth, 2017). Alterations in the functioning of the HPA in response to early life stress from birth in children aged between 1-, and 6-years of age has been associated with problems in emotional regulation, temperament, later behaviour, and psychopathology (Loman & Gunnar, 2010; Simons, Beijers,

Cillessen, & de Weerth, 2015). Alterations of the composition of the GM has been found by several studies to be associated with dysregulation of the HPA axis (Frankiensztajn, Elliott, & Koren, 2020; Misiak et al., 2020; Rosin et al., 2021). Interestingly, increased HPA activity, resulting in abnormal levels of cortisol, may also result in altered composition of the GM (Zijlmans, Korpela, Riksen-Walraven, de Vos, & de Weerth, 2015). As this relationship is bidirectional in nature it is difficult to establish a causal pattern and clearly state whether it is the GM that initiates alterations in the HPA axis functioning, or whether it is the altered functioning of the HPA axis that changes the GM composition. Due to the immaturity of both HPA axis and GM, alterations in function are likely to occur during early infancy, as they are both more susceptible to stress during this time (Frankiensztajn et al., 2020). The HPA axis is activated by the gut microbiota in several ways, including via cytokines and prostaglandins (Frankiensztajn et al., 2020). Current evidence has found that disruption of the GM alters the HPA functioning by influencing the hormonal levels and cytokines produced in the bacterial metabolic processes (Frankiensztajn et al., 2020). This can subsequently lead to anxiety, depression, and stress disorders (Kinlein, Phillips, Keller, & Karatsoreos, 2019; O'Mahony, Clarke, Dinan, & Cryan, 2017; Rosin et al., 2021). In adult humans with mental health disorders, there is often a dysregulation of the gut microbiome present (Skonieczna-Żydecka et al., 2018). For instance, decreased *Faecalibacterium* and *Bifidobacterium,* and increased *Bacteroidetes* and *Proteobacteria* have been associated with depression in adults (Aizawa et al., 2016; Jiang et al., 2015). Accurate interpretation of the mechanisms leading to dysregulation of the gut microbiome in disordered populations is difficult as there are a number of factors to which it can be attributed, e.g., poor self-care, including diet, after the onset of symptoms. It is therefore important to understand variation of GM composition in typically developing children to help us interpret variation in those with diagnoses (mental health or otherwise).

When investigating predictors of behavioural problems and risk of mental health or psychopathology, one of the earliest assessment tools in children is measures of temperament. Temperament characterises an individual's behaviour and encompasses sets of traits that can be used to predict psychological health outcomes. Temperament traits can be thought of as a subset of personality traits, which has historically been described as a system that determines an individual's ability to interact with their environment (Allport, 1937). As described earlier in this chapter, temperament traits include emotional, motor, and attentional reactive tendencies (Rothbart & Bates, 2006). Although these traits were once thought to be stable across the lifespan, they can be influenced through several mechanisms including genetics, exposure to prenatal stress, neurochemistry, culture, gender and a number of other environmental factors (Davis et al., 2007; Gartstein & Skinner, 2018; Saudino, 2005; Schmitz, Saudino, Plomin, Fulker, & DeFries, 1996). These temperamental traits influence development over the whole lifespan , and are indicators of potential risk of psychopathology that develop as the child matures, including internalising, and externalising disorders, and are even related

to physical health (Muris & Ollendick, 2005; Rigato, Charalambous, Stets, & Holmboe, 2020) . Through investigations into the factors that influence temperament in early childhood it is possible to understand ways in which these internalising, and externalising disorders may be prevented or the effects lessened. There are several measures of temperament that are designed to be used in infancy and early childhood, these include the Infant Behaviour Questionnaire (IBQ) (Putnam, Helbig, Gartstein, Rothbart, & Leerkes, 2014), the Early Childhood Behaviour Questionnaire (ECBQ) (Samuel P Putnam, Gartstein, & Rothbart, 2006), and the Child Behaviour Questionnaire (CBQ) (Rothbart, Ahadi, Hershey, & Fisher, 2001). Investigations into the underlying mechanisms of individual differences in temperament highlight dysregulation of the GM as a potential contributing factor, which can be explained through altered functioning of the HPA axis. Luczynski et al. (2016), investigated the effect of GM dysregulation upon the HPA axis by comparing the volume of the amygdala and hippocampus between germ free (GF) mice and those whose GM were conventionally colonised (CC). They found that that there were significant structural differences between GF and CC mice that result in altered functioning of the HPA. Again, although this is a first step in understanding the possible mechanisms that result from altered GM status influencing temperament it is not yet clear how this develops in humans and children. Furthermore, it is yet to be established how this influences later behavioural problems.

As mentioned previously behavioural problems are predominantly characterised as either internalising or externalising problems. Externalising problems are characterised by a tendency towards aggression, and disobedience on the conduct scale, and overactivity, impulsivity, distraction, and a lack of concentration on the hyperactivity/inattention scale. These problems can create a large amount of stress for both child and parents, and if they persist can be predictive of adolescent and adult mental health and behavioural problems (Reef, Diamantopoulou, van Meurs, Verhulst, & van der Ende, 2010; Tremblay, 2010). Internalising behaviour problems, characterised by heightened anxiety or worry, depressive features, and increased fear on the emotional problem subscale, and inability to interact with peer group, and solitary tendencies on the peer problems subscale, is also a predictor for mental health outcomes such as depression (Costantini, Paul, Caldwell, López-López, & Pearson, 2020). For these reasons it is essential to understand the underlying mechanisms that lead to these types of problems in early childhood. One of the most commonly used tools for assessing behavioural development and the risk of behavioural problems is the Strengths and Difficulties Questionnaire (SDQ). A study conducted by Ou, Belzer, Smidt, and de Weerth (2022), used the SDQ to investigate the development of the gut microbiota and the influence upon internalising and externalising behaviour, following children from birth to 10-years of age. They found that increased relative abundance of *Prevotella_9* and *Phascolarctobacterium,* measured at 6- and 10-years of life, was positively associated with maternal measures of externalising behaviour at the age of 10. In a second study, also investigating children from birth, found that *Prevotella* measured during the first year of life,

was negatively associated with increased problem behaviours measured using the child behaviour checklist (CBCL) at 2-years. From these two studies it can be seen that it is possible to see patterns of associations between composition of GM and behavioural problems through childhood. However, as the most sensitive time of GM development is during very early childhood it is necessary investigate these patterns in order to be able to understand the influence the development of the GM has upon behaviour before the more stable adult-like enterotypes are established.

1.4 The gut microbiome, diet, and behaviour.

The relationship between GM composition and diversity, diet and behaviour has yet to be explored thoroughly. There is a significant research gap that exists because it is currently only established that diet significantly impacts the composition and diversity of the GM, that GM compositions that vary significantly in either diversity or composition have been shown to be associated with poorer mental health and behavioural outcomes, and that diet has been associated with behavioural and cognitive outcomes in direct relationships. Figure 1.1 below presents the dynamic relationship between host, environment, and microbiota, including the bidirectional nature of these relationships. This illustrates fundamentally some of the difficulties in establishing causal pathways as well as the hypothetical models that can be explored in this type of research. In this instance the environmental factor of interest is dietary intake, which of course is heavily influence by the host in terms of quantity, types of food taken in, taste preferences. It is known that dietary intake also significantly influences the host, however the pathway that is of particular interest in this thesis is the influence of interaction between microbiota and environment that may influence the host behaviour. This section explores further the relationship between diet, behaviour and cognition, and any current initial findings that have been established with the influence of diet upon the relationship between GM and behaviour.

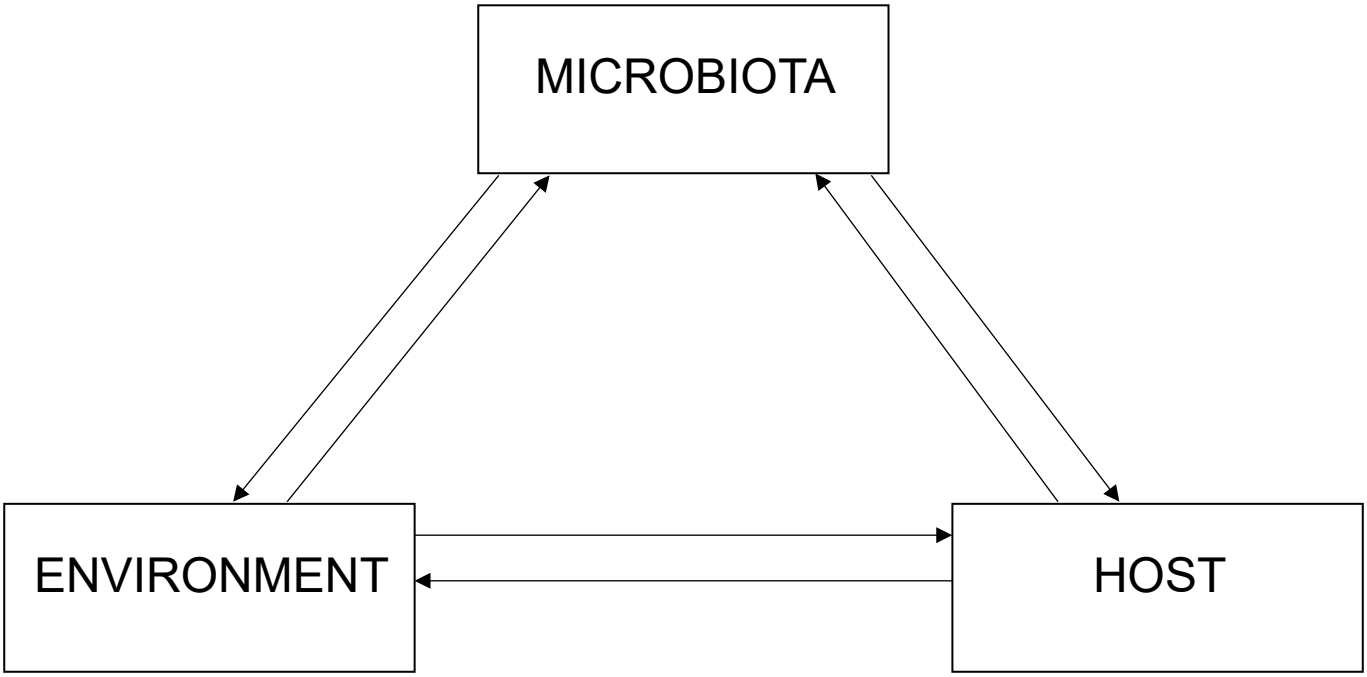


Figure 1.1 the dynamic relationship between microbiota, environment, and host.

Several studies have investigated the influence of diet upon behavioural outcomes, in particular there is a growing wealth of literature that associates poor dietary quality with diagnoses of attention deficit hyperactivity disorder. In addition to influencing the GM, diet in early childhood has been shown to be a good candidate as an influencing factor upon behaviour. The ALSPAC study (n=4000), a longitudinal study following UK children from birth, first investigated the effect of a 'junk food diet' on behavioural outcomes. In this study 'junk food' was defined as a diet high in processed foods and soft drinks. Dietary intake was measured using a food frequency questionnaire collected by maternal report when the child was between 37- and 54-months of age. The measure of dietary intake was then used to identify a 'junk food' factor, which in turn was used to further investigate the effects upon behavioural problems in children. The Strengths and Difficulties Questionnaire (Goodman, 1997) was used to assess mental health of children on five subscales: emotional symptoms, conduct problems, hyperactivity, peer problems and prosocial behaviour. They found that there was a strong relationship between 'junk food' intake at mean age of 4.5-years and greater hyperactivity measured using the SDQ at 7-years of age (Wiles, Northstone, Emmett, & Lewis, 2009). Thus, there appears to be a domain-specific association between diet and behavioural outcomes in children. As the study included measurements at 38-and 57-months for the dietary data and 47-months for the SDQ, it may be possible to determine some elements of causality. However, the dietary data was quantified as frequency of consumption of different types of food. This data could have been more accurate if the quantity of each type of food was also recorded, and therefore would give clearer understanding of the results. This overall result of this paper is replicated in several studies and is summarised in a

systematic review and meta-analysis by Del-Ponte, Quinte, Cruz, Grellert, and Santos (2019), where a total of 14 studies were found to have investigated the associations between diet and ADHD. Additionally, this review considered a pre-specified set of confounders including sex, socioeconomic variables, and maternal schooling. The results of the meta-analysis found that an overall healthy diet was protective against the development of ADHD, whereas an unhealthy diet was indicative of risk of developing ADHD. Healthy diets were most often characterised as consumption of fruits, vegetables, and whole grains, whereas 'unhealthy' diets were characterised as consumption of saturated fats and refined sugars. This is consistent with the result of the ALSPAC study and the notion of 'unhealthy' diet aligning with the consumption of 'junk food.' Furthermore, these results occurred after adjustment for confounders showing that higher levels of processed food were associated with higher levels of ADHD (Yan et al., 2018), and likewise similar patterns were seen with increased levels of sweet or fat consumption (Azadbakht & Esmaillzadeh, 2012), which further indicates that it may be dietary intake that is driving the associations.

A number of studies have investigated development in infants and the impact that breast/formula feeding has upon cognitive outcomes (Bauer, Ewald, Hoffman, & Dubanoski, 1991; Horwood & Fergusson, 1998). Infants who are breastfed for less than 3 months have shown poorer cognitive development at 1 and 5 years when compared to those who receive breastfeeding until 6-months (Angelsen, Vik, Jacobsen, & Bakketeig, 2001). The mechanism by which diet may affect cognitive performance is not yet fully understood. Animal models have investigated variation in dietary components, including fat and sugars, and the impact that these dietary components have upon cognitive function. Magnusson et al. (2015), investigated the impact of diet upon anxiety measures, memory, and cognitive flexibility in mice. Male, 2-month-old mice were randomly assigned diets that varied in fat, carbohydrate, and sugar content. They found that the microbiome of mice in the high-fat and high-sugar groups were significantly increased in the bacterial taxa *Clostridiales*. The high-sugar groups also presented with significant decreases in *Bacteroidales* and significant increases in *Lactobacillales*. Short-and long-term memory and cognitive flexibility was measured during performance in the Morris water maze. The Morris water maze is a navigation task designed to measure spatial learning and memory in rodents. Cognitive flexibility is assessed by moving the location of an invisible platform. Mice in the high-sucrose group showed significant problems with cognitive flexibility, working memory, and spatial bias in long-term memory compared to those in the high-fat group. Furthermore, increases in the bacteria *Clostridiale* and decrease in *Bacteroidales* are associated with poor cognitive flexibility. Carlson et al. (2018), explored the influence of the gut microbiome in humans at 1-year of age upon cognitive outcomes at 2-years of age. They found that higher alpha diversity at 1-year of age was associated with poorer overall scores on the Mullen scales of early learning, and particularly influenced expressive language, at 2-years. Whilst this gives a first insight into the association between the gut microbiome and cognitive outcomes in children, the

dietary measures included were limited to whether the infant was currently receiving any breastmilk or not. Further investigation into the relationship between GM, cognitive development and the role of infant diet is therefore necessary to understand how diet may shape developmental trajectories.

In addition to influence upon cognitive development there have been several investigations into the potential influence that the gut microbiome may have upon temperament. Specifically, there have been numerous studies that investigated the influence that temperament has upon infant feeding behaviours, food acceptance and the dyadic feeding relationship between mother and child. The relationship between mother and child is also of a bi-directional nature. It is necessary for the infant to provide hunger and satiety cues that are interpreted by the mother correctly, and for the mother to respond in an appropriate and timely manner (McMeekin et al., 2013). The interpretation and response to hunger and satiety cues by the mother is known as responsive feeding and can vary in response to infant temperament. Infants who are seen as having difficult temperaments by their mothers often have difficulties during feeding. Farrow and Blissett (2006), found that maternal reports of feeding difficulties and infant temperament were associated, with unadaptable and fussy difficult behaviour relating to greater feeding problems. Temperament may therefore predict more narrow dietary range, particularly the consumption of fruit and vegetables, and thus in turn may contribute to a less diverse GM. McMeekin et al. (2013) further investigated the associations between infant temperament and early feeding practices as part of the NOURISH study of Australia. First time mothers and their infants (n=698) were recruited into the study when the infants were between 2-7 months of age. Using the Short Temperament Scale for Infants (STSI) and the Infant Feeding Questionnaire (IFQ), they found that infant temperament was associated with maternal feeding practices. More specifically, mothers of infants who were reported to have more difficult temperaments scored lower on the IFQ, which is a self-report of maternal feeding practices including items such as “I know when my baby is hungry.” Additionally, infants rated as higher temperamental “difficulty” had mothers who used food more frequently to calm their infant. This in turn may lead to behaviour associated with eating for reasons other than hunger and may lead to consumption of sweet/fatty foods often seen with emotional eating (Zellner et al., 2006).

As mentioned previously there is evidence that the composition of the GM may influence behaviour through dysregulation of the HPA axis and therefore altered response to environmental stressors. Additionally, appetite, eating behaviour, frequency of food and choice of food consumed are also associated with increased susceptibility to stress. Sominsky and Spencer (2014), reviewed the current literature relating eating behaviour to stress response, and identified that after a period of HPA activation in response to acute stress, there is a period of glucocorticoid stimulation that inhibits hunger. However, during periods of chronic stress, the glucocorticoids in the blood stream are elevated, which activate the hypothalamus and increase appetite. Chronic stress also enhances the

likelihood that an individual will consume high calorie food. The resultant eating behaviour in turn works as negative feedback, which is shown to suppress the HPA axis and reduce the feeling of stress. This pattern of behaviour in adults often leads to obesity and further exacerbation of the gut microbiome dysregulation. Infants, however, are not solely responsible for the food that they consume. What is not clear in the current literature is how both dietary intake and GM dysregulations are linked in altering an individual's HPA and stress response. It is therefore necessary to explore the factors that contribute to changes the dietary intake of infants, such as temperament and the mother-infant feeding relationship and how these further contribute to alterations in development of the gut microbiome.

Despite the well-established links between diet and behaviour, and diet and GM, there is little true evaluation of the influence of diet upon the relationship between GM and behaviour. The majority of research focuses on the influence of GM composition, upon behavioural outcomes, through investigation of the relative abundance of individual genera of bacteria or through cluster analysis of the dominant species within the GM. Diet is often included only as a covariate in these analyses, with no true investigations of the influence of dietary patterns. For this reason, this thesis shall take a novel approach to investigate very early dietary patterns including milk-based diet and first solid foods introduced to the diet, and the role that this plays in shaping the established relationship between GM and behavioural development.

1.5 Summary

From the research gathered in this literature review it can be understood that the period of maturation of the GM is highly influenced by dietary intake, and that both GM diversity and composition, and diet have strong associations with development of temperament and behaviour during early childhood. Research is yet to thoroughly investigate the role of diet upon the relationship between GM and behaviour. Whilst the relationship between diet and GM, and diet, temperament and behaviour has been researched previously and established, there is little evidence linking all three relationships together. Given that the first years of life present with sensitive periods of development that align for both GM and the brain (Cowan, Dinan, & Cryan, 2020), which subsequently may impact upon temperament and behavioural development, it may be possible to develop interventions to prevent later poor mental health and behavioural problems. Diet is a particularly good candidate for possible interventions. However, before it is possible to develop these interventions it is essential to understand the underlying mechanisms, and contributors to behavioural development, and to further understand the GM, and dietary factors that lead to healthy behavioural development.

1.6 Aims, Objectives and Hypotheses

The overall aim of this thesis is to investigate changes in diversity and composition of the gut microbiome during the period of GM maturation from birth, how this is related to temperament and behavioural development throughout early childhood, and how this relationship can be influenced by diet. This aim will be achieved through the following three objectives:

1. To synthesise previous literature to investigate the role of GM and the influence upon temperament during early childhood. A systematic review will be conducted to thoroughly investigate this relationship in previous literature, also including evaluation of microbiota sampling and analysis techniques, and evidence of associations with key confounding variables. The systematic review will be presented in chapter 2 of this thesis.

2. To investigate the maturation of the GM, and the influence of dietary intake upon behaviour outcomes measured in 4-year-olds through a longitudinal approach. The first objective is to investigate whether early infant and childhood gut microbiota composition characteristics are related to behaviour at 4-years of age. Furthermore, a multi-mediation approach will be used to investigate whether gut microbiota diversity and composition mediates the relationship between early childhood diet and behaviour at 4-years. Through investigation of previous literature, *a priori* hypotheses were developed with *Bifidobacteria* and butyrate producing bacteria as candidate bacteria of choice in these investigations. These results will be presented in chapter 3. It is hypothesised that;

    a. diversity and composition of the GM will be associated with SDQ scores: specifically, there will be a) a negative association between diversity and SDQ scores, b) a negative association between the relative abundance of *Bifidobacteria* and total butyrate producing bacteria, and SDQ scores.
    b. Diet, both milk-based and first solid foods introduced, would be associated with SDQ scores: specifically, the relationship between diet and SDQ scores will be mediated by microbiota composition.

3. To investigate the entire microbiota in order to establish bacteria of interest and investigate the relationship between these bacteria, dietary intake, and behaviour, through use of moderation analysis. The results of this investigation will be presented in chapter 4. It is hypothesised that
    a. The composition of the GM will be associated with SDQ scores for bacteria of interest identified at 1-month, 6-months, and 12-month of age.

b. Diet will significantly moderate the relationship between GM composition and behaviour measured using the SDQ at 4-years.

4. To investigate the prediction of risk of behaviour problems that can be established as clinically abnormal, when measured at 4-years of age, from relative abundance of bacteria of interest established through investigation of the whole microbiota. This will be investigated through use of a logistical regression approach, the results of which will be presented in chapter 5. It is hypothesised that;
    a. The composition of the GM will significantly predict clinically significant scores on the scales of internalising problems, externalising problems and total difficulties measured at 4-years of age on the SDQ.

5. To design a future study and determine whether the proposed project will be seen as feasible and acceptable by target participants. The objective is to synthesis evidence from empirical work carried out in this study, to design a future study that will extend current literature and understanding of the relationship between GM, diet, and behaviour. In order to achieve this the proposed research design and methodology will be presented to a target audience through patient participant involvement (PPI) and feedback will be gathered regarding the project suitability. The proposed study and results of the PPI will be presented in chapter 6.

CHAPTER 2

THE DEVELOPMENT OF THE GUT MICROBIOME AND TEMPERAMENT DURING INFANCY AND EARLY CHILDHOOD: A SYSTEMATIC REVIEW

This chapter was accepted for publication in May 2022 and made available online using the link below. Any variations from publication have been made to accommodate thesis guidelines.

Alving-Jessep, E., Botchway, E., Wood, A. G., Hilton, A. C., & Blissett, J. M. (2022). The development of the gut microbiome and temperament during infancy and early childhood: A systematic review. *Developmental Psychobiology*, *64*(7), e22306. **https://doi.org/10.1002/dev.22306**

CHAPTER 2
THE DEVELOPMENT OF THE GUT MICROBIOME AND TEMPERAMENT DURING INFANCYAND EARLY CHILDHOOD: A SYSTEMATIC REVIEW

ABSTRACT

Background: Temperament in early childhood is a good predictor of later personality, behaviour, and risk of psychopathology. Variation in temperament can be explained by environmental and biological factors. One biological mechanism of interest is the gut microbiome (GM), which has been associated with mental and physical health.

Method: This review synthesised existing literature evaluating the relationship between GM composition and diversity, and temperament in early life. Web of Science, PsycInfo, PubMed, and Scopus were searched, data was extracted according to PRISMA guidelines.

Results: In total 1562 studies were identified, of which 6 remained following application of exclusion/inclusion criteria. The findings suggest that there is an association between higher alpha diversity and temperament, greater Surgency/Extraversion and High-Intensity Pleasure in males, and lower Effortful Control in females. Unique community structures (beta diversity) were found for Surgency/Extraversion in males and Fear in females.

Discussion: An emerging pattern of positive temperament traits being associated with GM communities biased toward short-chain fatty acid production from a metabolism based on dietary fibre and complex carbohydrates was observed and is worthy of further investigation. To gain deeper understanding of the relationship, future research should investigate further the functional aspects of the microbiome and the influence of diet.

2.1 Introduction

The global prevalence of mental health disorders in children and adolescents, aged between 6 and 18 years, is close to 15% (Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). For example, it is estimated that currently in the U.K., 1 in 8 children aged between 5 and 19 years old are diagnosed with at least one mental health problem, including emotional, behavioural and/or hyperactivity disorders (Baker, 2020). These disorders typically have a pervasive effect upon individuals throughout their childhood, adolescence and into adulthood (Kretschmer et al., 2014). Identification of early childhood markers of later behavioural and mental health problems provides the possibility of developing interventions and prevention programmes targeted at early life stages. Measures of temperament in early childhood, through parent, teacher, or observational measures, have been shown to be good predictors of personality, behaviour, and risk of psychopathology in later childhood and adolescence (Muris & Ollendick, 2005), and thus are good early targets for investigation of preceding markers.

Temperament, which refers to an individual's patterns of behaviour, including emotional responsiveness, mood, and the speed and intensity of reactions, is often considered to be a fundamental component of personality, present early in life (Sanson, Hemphill, & Smart, 2004). During early childhood, an individual's reactions to their environment are predominantly influenced by temperament (Rothbart, 2012) and temperamental traits in children are closely linked with the broad factors used to describe personality traits in adulthood (McCrae et al., 2000). Temperament in childhood can give insight into later behaviour through its close relationship with personality (Rothbart, Sheese, Rueda, & Posner, 2011), suggesting that it is a strong marker of later behavioural phenotypes. Although temperamental traits have previously been considered to be stable over time, it is possible for them to undergo change during an individual's development (Rothbart & Bates, 2006). Evidence for individual differences in temperament has shown that between 20-60% of phenotypical variance in personality can be accounted for by genetics (Saudino, 2005). Nevertheless, data from twin and adoption studies have also shown that environmental factors play an important role in individual differences in child temperament (Saudino, 2005).

Composition of the gut microbiome is likely to influence children's temperament, given that several studies in children have implicated the GM in a range of other physical and mental and developmental outcomes. The GM includes both the composition of the communities of bacteria, viruses, archaea, and fungi which colonise the gut, as well as the collective genome. In contrast, the term 'microbiota' refers to the composition of a community including bacteria, viruses, archaea, and fungi but not its collective genome. In this review, the use of GM refers exclusively to gut microbiome and where microbiota is the topic of discussion, this term is written in full.

Studies in children have implicated the GM in several health outcomes, including physical health conditions, such as obesity (Murugesan et al., 2018) and asthma (Attar, 2015; Moossavi et al., 2018), as well as mental health conditions such as attention deficit hyperactivity disorder (ADHD; (Adesman et al., 2017). Furthermore, bacterial colonisation of the gut has been shown to be directly related to the maturation of both the central nervous system (CNS) and enteric nervous system (ENS) in children (Barbara et al., 2005; Stilling, Dinan, & Cryan, 2014). From a developmental perspective, the most rapid phase of colonisation of the gut starts at birth and continues until maturation of the GM at approximately 31-46-months (Stewart et al., 2018b). The GM and brain are thought to share similar sensitive periods of development during infancy that are known to extend up until the second year of life (Borre et al., 2014; Heijtz et al., 2011). Sensitive periods in the development of the microbiota include birth and the early postnatal period, as well as during complementary feeding (the period of introduction to solid food, typically around 6-months of age as recommended by the World Health Organisation; WHO). These periods align with neurodevelopmental periods of plasticity including sensory function, language, learning, and memory (for a review see (Cowan et al., 2020)).

The GM and brain share a bidirectional relationship, and this communication route between them is known as the gut-brain axis (GBA; Wang et al., 2018). The GBA comprises of several pathways including the CNS, ENS and the hypothalamic pituitary adrenal axis (HPA; (Skonieczna-Żydecka et al., 2018). Through their metabolism of several substrates including dietary fibre and carbohydrates, bacteria produce short-chain fatty acids (SCFA) consisting primarily of acetate, propionate, and butyrate (Silva, Bernardi, & Frozza, 2020). Each bacterium can be categorised by the SCFA that they produce, and each may produce one or more, through different metabolic processes (Louis & Flint, 2017). Production of SCFAs in the gut plays an important role in maintaining gut health, including prevention of inflammation and maintenance of intestinal barrier function. SCFAs additionally play a central role in the communication in the GBA (Silva et al., 2020). Bacteria within the gut also play an important role in the metabolism of tryptophan, an amino acid precursor of serotonin production, with the serotonergic system being key to the regulation of mood (Jenkins, Nguyen, Polglaze, & Bertrand, 2016). For these reasons, altered composition of the GM, through colonisation of aberrant species or changes in overall diversity or composition, may disrupt the communication of the GBA and further impact both physical and mental health of an individual. It is plausible that these effects may be evident in early development of temperament.

Animal models provide further evidence for the mechanism by which GM may influence the development of temperament. Studies of germ-free (GF) animals (specially raised animals that are free from all microorganisms), have shown that dysregulation of the GM is associated with lasting impact upon brain chemistry affecting stress response, cognition, and behaviour relating to anxiety and depression (Cryan & Dinan, 2012; Zheng et al., 2016). In humans, the HPA axis (which plays an

important role in emotional regulation and stress response) has been shown to be affected by GM dysregulation (de Weerth, 2017). It is plausible that these changes in the composition of the GM may alter both the functioning of the HPA axis and the relationship between the ENS and the CNS respectively, which has been suggested as a mechanism that drives individual differences in temperament (Luczynski et al., 2016).

To understand how variation in temperament is related to later adverse development, it is important to understand the relationship between temperament and its underpinning biological mechanisms, specifically the development of the GM. Despite emerging and developing interest in the relationship between the GM and temperament, there is no current consensus regarding specific bacterial composition or diversity of the GM and its relationship with different aspects of temperament. Thus, the aim of this systematic review was to gather and synthesise existing evidence relating to the relationship between GM composition and diversity and temperament in early childhood.

2.2 Methods

A systematic review was conducted using methods set out in the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2011). A protocol for this systematic review was registered on PROSPERO, registration number CRD42020196919.

2.2.1 Data sources

A search of the academic databases Web of Science, PsycInfo, PubMed, and Scopus was conducted from September to October 2020. Search terms were established relating to gut (gut, intestin*, enterotype), microbiome (microbiome, microbio* or bacteri*, composition, diversity), and temperament (temperament, personality, anxiety, sociability, “negative affect”, fear, shyness, mood, stress). Following establishment of search terms, Boolean operators were applied, and appropriate adjustments were made for each database. This process was repeated in June 2021 to ensure the most up to date papers were included.

2.2.2 Eligibility criteria

All studies reporting on the relationship between variation in composition of GM and temperament in children up to the age of 6-years 11-months. From the age of 7 years the temperament measures used move toward the middle and late childhood, 8-14-years of age. Therefore, for the purpose of this review the age range was limited to the early childhood years and related measurements. The study types of interest were cross-sectional and longitudinal observational studies. Inclusion criteria for this review were:

- Studies that measured the composition and/or diversity of GM using either whole genome sequencing, 16S rRNA, or shotgun sequencing methods.
- Studies that used established scales to measure behavioural outcome measures of temperament.
- Studies involving healthy child participants (and their parents), from birth to 6-years 11-months.
- Studies involving children born at full term, after 36-weeks but before 42-weeks.
- Studies with children born either via vaginal delivery or caesarean section.
- Studies involving children with either single or multiple births, with or without siblings.

Exclusion criteria for this review were:

- Studies not written in English.
- Reviews, meta-analyses, book sections, book chapters, and studies not published in peer-reviewed journals.
- Studies that did not measure the outcomes of interest and/or did not use next generation sequencing to measures GM composition.
- Studies with children born prematurely <36-weeks gestation, or late >42-weeks gestation.
- Studies of children with diagnosed gastrointestinal health conditions, such as studies focused on Crohn's disease.
- Studies focused on children with severe/multiple allergies.
- Studies of children with diagnosed genetic disorder or syndrome, learning or developmental disorder, or acquired injuries that have known links with altered GM composition or behaviour, such as studies focused on autism.

2.2.3 Study selection and data extraction

Endnote version x9 software (Hupe, 2019) was used to collate articles from academic database searches and to remove duplicate articles. Two reviewers, EAJ and EB, independently screened each article by title and abstract to remove articles that either did not meet the inclusion criteria or met the exclusion criteria. Full articles were then independently screened for inclusion by two reviewers (EAJ and EB), and final selection was made for data extraction. A third reviewer (JB) was consulted to settle discrepancies in review decisions at each stage of the review process. Researcher EAJ extracted all data for the final articles included in this review. Data were first extracted into an excel sheet for synthesis, including details on author and location of study, publication year, study design, participant demographics including age at each time point, GM technique including, collection method, sequencing method, hypervariable regions, temperament measurement, GM diversity and/or

composition measures. Statistical methods and study results were also extracted including significance levels and effect sizes with confidence intervals, where possible.

2.2.4 Quality assessment

All articles that met inclusion criteria for this review were assessed for risk of bias independently by two reviewers (EAJ and EB). The National Heart, Lung, and Blood Institute (NHLBI) quality assessment tool for observational cohort and cross-sectional studies (Health, 2018) was used to rate the articles. This tool has 14 questions evaluating the inclusion and quality of the research question, study population, sample size justification, exposure measures, outcome measures, statistical analyses, timeframe, blinding, and repeated exposures. Each question was scored as either 'yes', where the criteria is satisfied; 'no', where the criteria is not met; or 'not applicable'. A score of 'yes' corresponded with one and 'no' or 'not applicable' corresponded with zero. Scores for each article were totalled, and a grading system developed by Uloko et al. (2018) was employed to rate the selected articles into: 'Good' (≥ 70%), 'Fair' (≥ 50%), and 'Poor' (< 50%).

2.3 Results

A total of 2176 articles were identified (1698 in the first search (S1), 478 in the second search (S2)). Duplicate articles (Total n =614, S1= 544, S2=70) were removed, and 1562 (S1=1154, S2=408) articles were screened by abstract and title. Following the first screening 1128 did not meet the eligibility criteria and were removed. Full text screening was carried out on a total of 30 articles (S1 = 26, S2 = 4) 24 articles were unanimously excluded, including one for which the third reviewer (JB) was consulted to resolve conflict in the review decision. A total of 6 articles were included in this systematic review, each reporting on a unique study, that met the review criteria as shown in Fig. 1, PRISMA flowchart (Moher et al., 2011).

2.3.1 Study characteristics

The six studies included a total of 733 participants, with individual studies including 40 to 301 participants (Aatsinki et al., 2019; Christian et al., 2015; Flannery et al., 2020; Kelsey et al., 2021; Loughman et al., 2020; Wang et al., 2020). Child participants ranged in age from birth to 6-years 11-months old. Publication dates of the articles ranged from 2015 to 2021 (See Table 2.1). Two studies were longitudinal studies (Aatsinki et al., 2019; Loughman et al., 2020) and four were cross-sectional observational studies. Both longitudinal studies recruited families during the prenatal period and have continuing follow ups, one focused on the age range from birth to 6-months of age, and the other from birth to 2-years of age. Location of study samples occurred across several continents; 3 were conducted in the USA, one in Australia, one in Europe (Finland), and one in China.

2.3.2 Quality assessment

Overall, the quality of articles was varied, two papers were assessed as ‘Good’, three “Fair’, and one “Poor’ (Flannery et al., 2020; see Table 2.2). All articles included a clear research question, well defined exposures, including levels of measures, and outcome measures. Only one of the articles included in this review presented effect sizes, which satisfied question 5 ‘was a sample size justification, power descriptions, or variance and effect estimates provided?’ (Kelsey et al., 2021). None of the articles included power descriptions or sample size justifications. For all studies, blinding of the assessor to the exposures of participants was marked as not applicable; no studies were interventions. Cross-sectional studies were scored as “no” to the following questions: question 6, ‘For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?’; question 7, ‘Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?’; question 10, ‘Was the exposure(s) assessed more than once over time?’ and question 13, ‘Was loss to follow-up after baseline 20% or less?’. This resulted in lower overall quality scores for all cross-sectional designs. The scoring was performed in line with the instructions of the NHLBI quality assessment tool for observational cohort and cross-sectional studies (NHLBI, 2017), whilst acknowledging the systematic impact of the scoring system on ratings of cross-sectional designs.

2.3.3 Microbiome analyses

All the studies included in this review investigated the composition of the microbiota, studies 5 and 6 additionally investigated the functional composition of the GM. GM diversity was assessed in all studies, except Study 4. Alpha diversity was measured in studies 1, 2, 3 and 6, Beta diversity was measured in studies 1 and 3, and functional beta diversity was measured in Studies 5 and 6. Four of the studies included in this review used 16S rRNA sequencing to investigate the gut microbiota, and two studies (study 5 & 6), used shotgun metagenomics to investigate the microbiome. Three of the 16S rRNA studies used the Illumina MiSeq platform to sequence the data, except study 1, which used Roche 454 FLX Titanium system. Studies investigating the 16S rRNA varied in hypervariable region selection. Studies 2 and 3 used V4 region only, Study 4 used both V3 and V4 and Study 1 used V1-3 (See Table 2.3). Both the SILVA taxonomic data base and the GreenGenes reference database were

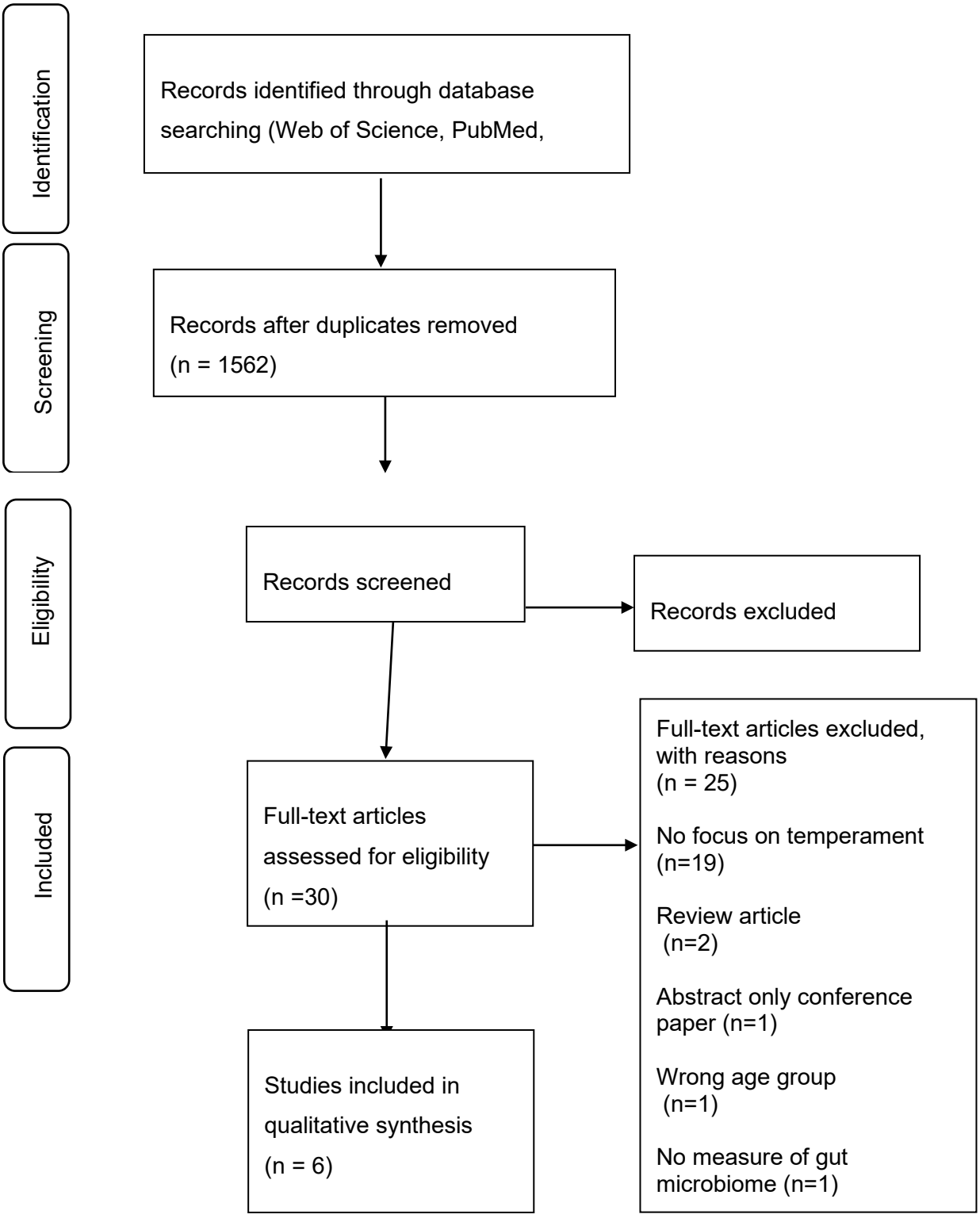


Figure 2.1. PRISMA flowchart illustrating systematic review screening process.

used to identify genera of bacteria from the Operational Taxonomic Units (OTUs) assigned. Flannery et al. (2020), prepared the raw sequences for analysis using the Shot Cleaner Workflow, following the guidelines laid out by the Human Microbiome Project Consortium (2012). Furthermore, Flannery et al. (2020) used Shotmap to quantify group relative abundance by function using the Kyoto Encyclopedia of Genes and Genomes (KEGG) group relative abundance by function (Sharpton 2017) and Metaphlan2 (Truong et al., 2015) was used to quantify taxon relative abundance. Kelsey et al. (2021), used a series of pipelines developed in-house to analyse their microbiome samples. The JAMSalpha pipeline was used to obtain taxonomic and functional relative abundances. Metagenomic contigs were taxonomically classified using the k-mer analysis in Kraken 2.

Sample collection methods also varied, with two of the studies (1 and 2) requesting that participants stored the samples chilled at +4ºC until collection/drop off at the laboratory. Studies 1,2, and 6 requested that samples were brought to the laboratory within a 24-hour window. Studies 5, and 6 collected at ambient room temperature, and study 3, collected samples either fresh or chilled in a home freezer at typically -18 ºC. Study 3 collected samples at time of the home visit for completion of the temperament scale, samples were transported in a cooler at +4ºC for an average of 1.5 hours until they reached the lab. Following collection, all but study 2 froze their samples at -80 ºC until DNA extraction was performed ready for analysis. Study 2 began DNA extraction as soon as samples reached the lab.

2.3.4 Temperament measures

Temperament was typically measured using several well-established scales (See Table 2.3). Study 1 used the Early Childhood Behaviour Questionnaire (ECBQ; (Putnam et al., 2006), studies 2 and 6 used the Infant Behaviour Questionnaire – Revised Short Form (IBQ- R SF), study 4 used the Chinese version of the Infant Behaviour Questionnaire – Revised (IBQ-R; (Putnam et al., 2014), and study 5 used the Child Behaviour Questionnaire (CBQ; Rothbart et al. (2001). In addition, study 3 used a single 5-point Likert scale to measure ‘temperament’ developed by Ponsonby, Dwyer, and Couper (1997). This scale has not been validated as an in-depth measure of temperament in infants.

Table 2.1 Summary of article characteristics (N=6)

Authors	Country	Number of participants	Age range child	Average age parents (if given)	Ethnicity	SES proxy markers
1. Christian et al. (2015)	USA	77 mother-toddler dyads	18-27 months of age, mean = 23.14 (SD = 2.00)	Maternal mean = 31.1 (SD = 5.43)	87% (n=67) White, 9.1% (n=7) Black, 3.9% (n=3) Asian	None Given.
2. Aatsinki et al. (2019)	Finland	Sub-cohort of 301 (159 boys, 142 girls) [a.]	From birth followed until 6-months of age	Mother mean = 30.8 (SD = 4.3)	Not Specified	Maternal level of education reported. Upper Secondary n = 67 (22.3%) Vocational School n = 97 (32.2%) Tertiary education n = 128 (42.5%) NA n = 9 (3%)
3. Loughman et al. (2020)	Australia	Sub-cohort of 201[b]	From birth followed until 2-years of age	Maternal mean = 32.4 (SD = 4.2) Paternal mean = 34.2 (SD 5.1) In Sub-cohort	Not Specified	Socio-Economic Indexes for Areas SEIFA - Australia Low - n = 59 (29.4%) Middle - n = 73 (36.3%) High - n = 69 (34.3%) Unknown - n = 0

Authors	Country	Number of participants	Age range child	Average age parents (if given)	Ethnicity	SES proxy markers
4. Wang et al. (2020)	China	51 mother infant dyads	mean age 12.3 months (SD = 0.25).	Maternal mean = 32.3 (SD = 4.51)	Chinese	Maternal level of education reported.
5. Flannery et al. (2020)	United States	40 mothers and infants recruited	5 to 7 years of age Mean age = 6.12 (SD = 0.69)	Not given	55.26% (n=21) Caucasian; 18.42% (n=7) Mixed Race; 18.42% (n=7) Hispanic/Latinx; 7.89% (n=2) Native American/American Indian	Socioeconomic risk was indexed using the socioeconomic status and the Life Events Checklist.
6. Kelsey et al. (2021)	United States	63 participants took part. 23 participants were excluded due data quality.	Mean Average 24 days. Range 9 days to 56 days	Not given	Measured as white (73%) and non-white.	Measured as household income Less than $15,000 = 5 (8%) $15,001 - $30,000 = 5 (8%) $30,001 - $45,000 = 3 (5%) $45,001 - $60,000 = 1 (2%) $60,001 - $75,000 = 2 (3%) $75,001 - $90,000 = 9 (15%) $90,001 - $110,000 = 7 (11%) $110 ,001 - $125,000 = 7 (11%) $125,001 - $175,000 = 2 (3%) $175,001 - $225,000 = 8 (13%) $225,001 - $275,000 = 8 (13%) $275,001+ = 3 (5%)

Notes. a). Main cohort is part of the FinnBrain Birth Cohort Study (Karlsson et al., 2018) with a total of 5790 participating families. b). Main cohort is part of the Barwon Infant Study (Vuillermin et al., 2015) with a total of 1074 participants.

Table 2.2. Quality Assessment Scores According to The NHLBI Quality Assessment Tool for Observational Cohort and Cross-sectional Studies[1].

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Quality Rating
1. Christian et al. (2015)	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	NA	No	No	7/14 (Fair)
2. Aatsinki et al. (2019)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	NA	Yes	Yes	11/14 (Good)
3. Loughman et al. (2020)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	NA	Yes	Yes	12/14 (Good)
4. Wang et al. (2020)	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	NA	Yes	Yes	9/14 (Fair)
5. Flannery, et al. (2020)	Yes	No	CD	CD	No	No	No	Yes	Yes	No	Yes	NA	No	Yes	5/14 (Poor)
6. Kelsey et al. (2021)	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	NA	No	Yes	8/14 (Fair)

Note. Each question is answered either Yes, No, Cannot Determine (CD), Not Applicable, (NA), or Not Reported, (NR) as per guidance provided with this quality assessment tool.

[1] Criteria Questions. 1. Was the research question or objective in this paper clearly stater? 2. Was the study population clearly specified and defined? 3. Was the participation rate of eligible persons at least 50%. 4. Were all the subjects selected or recruited from the same or similar populations (including the same period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants? 5. Was a sample size justification, power description, or variance and effect estimates provided? 6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured? 7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed? 8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to outcome (e.g., categories of exposure, or exposure measured as continuous variable)? 9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? 10. Was the exposure(s) assessed more than once over time? 11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? 12. Were the outcome assessors blinded to the exposure status of the participants? 13. Was loss to follow-up after baseline 20% or less? 14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposures(s) and outcome(s)?

2.3.5 Microbiome and temperament outcomes

2.3.5.1 Microbiome diversity

Alpha diversity, a measure of the diversity of species within a given ecosystem or environment, was measured in four studies (studies 1, 2, 3, and 6). Study 1 used a phylogenetic diversity measurement, PD_whole_tree4, and the Shannon Diversity Index (SDI). In this study, microbiota measures were investigated separately for boys and girls. There were significant associations between alpha diversity, measured using phylogenetic diversity, and higher scores on the Surgency/Extraversion subscale, in both boys and girls aged between 18-27 months. There was no significant relationship between SDI and Surgency/Extraversion for either boys or girls. In boys, High-Intensity Pleasure was positively associated with both phylogenetic diversity and SDI, whereas in girls Effortful Control was negatively associated with SDI but not phylogenetic diversity. Study 2 found that alpha diversity scores of SDI and Chao1, measured at 2.5 months of life, had no statistically significant associations with temperament measured at 6-months of life. Study 2 also included several covariates in addition to gender, including gestational age, infant age, mode of delivery, breastfeeding status, and antibiotic intake age. In the adjusted models, alpha diversity was associated with negative emotionality and fear reactivity. The Chao1 measure of richness was not associated with temperament in the adjusted models. Although study 3 included alpha diversity measures, taken at 1-month, 6-months, and 12-months, the relationship between alpha diversity and temperament was not assessed. Study 6 found no significant associations between the behavioural temperament measures of negative emotionality, regulation/orienting, and surgency/positive emotionality, and alpha diversity in children aged 9 – 56 days (M = 24 days). This was consistent across all measures of alpha diversity both taxa diversity measures, Shannon-taxa, and Chao1-taxa, and Chao1 functional terms diversity, virulence factors, resistome and Gene Ontology (GO) terms. Virulence factors include the cellular structures that allow bacteria to invade and colonise a host, suppress immunity, and divert nutrition from the host. Resistome diversity is the genes within bacteria that code for products that increase resistance to antibiotics. GO terms characterise the contribution of individual genes to the biological make up of an organism. They did however find significant indirect effects, which suggested that the relationship between taxa diversity and negative emotionality may be mediated by homologous-interhemispheric connectivity. Similarly, they also found significant indirect effects for virulence factors diversity, and both negative emotionality and regulation/orienting, when mediated by homologous-interhemispheric connectivity. Studies 4 and 5 did not measure alpha diversity.

Beta diversity, measured using weighted and unweighted UniFrac distances, is a measure of amount of compositional difference between communities or environments. Study 1 investigated beta diversity in 18-27-month-old children. Using the Adonis statistic, they found that Surgency/Extraversion was associated with a unique microbiota structure measured on unweighted UniFrac, but not weighted

distances, in boys. Subscale analysis highlighted three subscales, Sociability, High-Intensity Pleasure, and Activity Levels, which drive the effect seen with unweighted differences. In girls, only one subscale, Fear, was associated with a unique community structure, measured using unweighted UniFrac distances. Study 1 was the only study to investigate beta diversity related to temperament. Additionally, due to beta diversity being an index of the unique community structures of microbiota within a study population, these results are difficult to generalize further beyond that specific study population, and therefore further work is needed to establish the importance of beta diversity in the development of temperament. Figure 2 illustrates the relationship between alpha and beta diversity and temperament of all included studies.

Table 2.3. Summary of key measures included in all articles (N = 6).

Authors	Microbiome sequencing technique	Hyper Variable Region	Fresh/chilled/frozen Samples	Age(s) of microbiome measurement	Temperament measure	Time points and age of Temperament measurement
1. Christian et al. (2015)	16S rRNA, Sequencing conducted on the Roche 454 FLX Titanium System, PyNAST was used for sequence alignment with the GreenGenes core reference database	Primers 27F/519R were used to extract V1-3 hypervariable regions.	Samples refrigerated at home +4C. Samples were transported on ice to lab, (temperature not specified), stored at -80C until pyrosequencing conducted.	Collected once at 18-27 months of age	Early Childhood Behaviour Questionnaire. Negative affectivity, Surgency/Extraversion and Effortful Control are measured	Collected once at 18-27 months of age
2, Aatsinki et al. (2019)	16S rRNA sequencing Illumina MiSeq, Qiime (v1.9) was used to check sequences against GreenGenes database and annotate the OTUs	V4 region selected.	Samples were chilled by participants at +4C	Collected once at 2.5 months of age	Infant Behaviour Questionnaire Revised Short Form (IBQ-R SF) Cronbach's alpha across subscales ranged from 0.65 to 0.84	Collected once at 6-months of age
3. Loughman et al. (2020)	16S rRNA sequencing, Illumina MiSeq Platform, Mothur software was used to assign Taxa using SILVA v123 Nr99 taxonomic database. Samples with fewer than 2500 reads were excluded	V4 region selected.	Stool samples stored at -80C, analysis adjustment of samples stored in home freezer (typically -18C) prior to delivery to the laboratory.	3 times total at 1, 6 & 12-months of age	Child behaviour checklist (Achenbach, 1999), and Temperament measured at 12-months1-, 6-, and 12-months, using a 5-point Likert scale	Collected once at 2-years of age

Table 2.3 continued

Authors	Microbiome sequencing technique	Hyper Variable Region	Fresh/chilled/frozen Samples	Time points and age of microbiome measurement	Temperament measure	Time points and age of Temperament measurement
4. Wang et al. (2020)	16S rRNA sequencing Illumina MiSeq, Taxa of each sequence was analysed using RDP Classifier algorithm, against SILVA database using confidence interval at 70%.	V3 & V4 hypervariable regions were selected	Chilled in a cooler during transport +4C, frozen until analysis at -80C.	Collected once at 12-months of age	Infant Behaviour Questionnaire - Revised (IBQ-R) Chinese Version.	Collected once at 12-months of age
5. Flannery et al. (2020)	Shotgun Metagenomic Sequencing was conducted. Raw metagenomic sequences were prepared for analysis using the Shotcleaner workflow, following the Human Microbiome Project Consortium data processing guidelines. Group relative abundance, by function, was assigned using the Kyoto Encyclopedia of Genes and Genomes. Metaphlan2 was used to quantify taxon relative abundance.	NA - Shotgun metagenome measured.	Collected at ambient temperatures and then stored at -80 until analysis.	Collected once at 5 to 7 years of age	The Child Behaviour Questionnaire and the Child Behaviour Checklist (Achenbach, 1999).	Collected once at 5 to 7 years of age

Table 2.3 continued

Authors	Microbiome sequencing technique	Hyper Variable Region	Fresh/chilled/frozen Samples	Time points and age of microbiome measurement	Temperament measure	Time points and age of Temperament measurement
6. Kelsey et al. (2021)	Shotgun metagenomic sequencing was conducted. A series of pipelines was used developed in house using the R language. The JAMSalpha pipeline was used to obtain taxonomic and functional relative abundances. Kraken2 and k-mer analysis were used to classify taxonomy.	NA - Shotgun metagenome measured.	Samples reached the lab within 24 hours of the study visit - average 7.96 hours. Samples were aliquoted into cryovials containing 20% Glycerol and 80% Phosphate =- Buffered Saline Solution. Samples were then frozen at -80oC.	Once, ranging from 9 – 56 days of life.	Infant Behaviour Questionnaire Revised Short Form (IBQ-R SF)	Once at the same time as the microbiome sample ranging from 9 – 56 days of life.

Table 2.4. Summary of Key findings and limitations across all included articles (N = 6).

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
1. Christian et al. (2015)	Alpha Diversity measured with phylogenetic diversity measurement, PD_Whole_tree, and the Shannon Diversity Index (SDI). Beta Diversity was measured using both unweighted and weighted UniFrac distances.	Genus Abundances were used.	1. Showed significant sex differences in temperament ratings including, higher scores for boys in motor ratings, High-Intensity Pleasure, while girls had higher rating of inhibitory control and soothability. 2. Among boys higher scores of Surgency/Extraversion were associated with greater phylogenetic diversity, but not associated with Shannon Diversity Index (SDI). High-Intensity Pleasure was associated with both greater phylogenetic diversity and SDI. 3. Surgency/Extraversion was associated with greater phylogenetic diversity, in girls, but not associated with SDI. There were also significant negative associations between the composite scores of Effortful Control and SDI, but not phylogenetic diversity. 4. In boys, sociability was positively associated with the abundances of an undefined genus in the family Ruminococcaceae, and Parabacteroides. High-Intensity Pleasure was positively associated with the genus Dialister, and an undefined family in Rikenellaceae. In girls, fear was positively associated with an undefined genus in the family Rikenellaceae.	1. This study was cross-sectional and observational in approach and does not allow for the determination of causal direction of effects. 2. In this study it was not possible to look at microbial function, as it would require a metagenomic or metatranscriptomic approach.

Table 2.4 Continued

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
2. Aatsinki et al. (2019)	Alpha - Shannon Index, Chao1.	OTU counts were used for composition measures.	1. Three distinct clusters were identified. 2. Clusters and infant temperament - Bifidobacterium/Enterobacteriaceae presented the highest scores and Bacteroides cluster with the lowest scores in trait of regulation (Kruskall-Wallis H test x2 = 5.8, FDR = 0.23), and the subscales of High-Intensity Pleasure, cuddliness, and duration of orienting. 3. Alpha diversity and infant temperament - Neither Shannon Index nor Chao1 were associated with any of the temperament traits. When adjusted for gestational age, infant age, sex, mode of delivery, breastfeeding and antibiotic intake, diversity was associated with negative emotionality (B= -0.17, FDR, =0.17, adjusted R2 = 0.016), and fear reactivity (B= 0.27, FDR = 0.17, adjusted R2 = 0 032) Chao, richness, was not associated after adjustment. 4. Genus Level investigation revealed that the temperament trait surgency is positively associated with streptococcus and regulation is positively associated with Erwinia, after controlling for sex, mode of delivery, infant age and gestational age, antibiotic treatment, and breastfeeding status.	1. Temperament assessment was based on report by mothers, which may be influenced by her own temperament and other characteristics. 2. Both GMC and temperament were assessed only at single time points. Serial and concurrent measurements should be undertaken in the future. 3. 16S rRNA offers comprehensive taxonomic profiling, but other methods such as shotgun sequencing, could offer better resolution.

Table 2.4 Continued

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
3. Loughman et al. (2020)	Alpha - Shannon, Simpson, Chao1 and Observed Species indices. Beta diversity both weighted and unweighted UniFrac distance.	Voom method from the limma package was used for differential normalised abundance testing.	1. No evidence of associations between one or 6-month alpha or beta diversity and behavioural outcome measured at 2-years. 2. No differential normalised abundance in microbiota of one month old associated with either behavioural case vs non-case. 3. In 6-month faecal samples Sutterella appeared lower in the case group but was attenuated for following adjustment for storage. 4. 12-month-olds Alpha diversity showed weak evidence of increased risk of elevated behavioural problems (OR: 2.42[0.92-6.97], p=0.087). 5. PERMANOVA analysis of unweighted UniFrac distances suggested differences in microbiota community structure between groups (R2 = 0.0092, p=0.018) 6. Normalised abundance of two bacterial groups were substantially different in the 12-month-old behavioural case infants. Prevotella was detected in only 4% of the cases vs 44% of non-case infants (logFC = -1.46, p < 0.0001, q < 0.0001). Lachnospiraceae was detected in 91% of case infants vs 69% of non-case infants (logFC = 2.09, p=0.0009, q=0.054)	1. Insufficient information to estimated dietary intake of fermentable fibre. 2. Relatively small number of cases with elevated behavioural problems. 3. Absence of potential metagenomic or transcriptomic data from which to infer potential mechanisms underlying the observed data.

Table 2.4 Continued

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
3. Loughman et al. (2020)			7. No association between Temperament and presence or relative abundance of Prevotella or Lachnospiraceae, when assessed for reverse causation.	
4. Wang et al. (2020)	None given	Relative abundance at OTU level.	1. Six genera were associated with infant temperament. Bifidobacteria was positively associated with soothability. Cuddliness was negatively correlated with Hungatella 2. Boys and girls showed no significant differences in temperament at 12-months old.	1. This study was a small-scale test of the associations between infant's gut microbiota and temperament at the age of 12-months. 2. Most infants in the study were taking probiotics. 3. The study did not address the role of diet quantity or quality at the time of stool sample collection, which may account for some of the differences in the associations.
5.Flannery et al. (2020)	Functional beta diversity measured.	Taxonomic and functional composition measures of the microbiome.	1. The quality of caregiver-child relationship moderates the associations between socioeconomic risk and both the structure and functional capacity of the gut microbiome. 2. The quality of caregiver-child relationship moderates the association between measures of behavioural dysregulation and the gut microbiome's functional capacity.	1. The study was cross-sectional and therefore it was not possible to determine which child later developed a mental health disorder. 2. It was not possible to discern a causal role of the microbiome upon behavioural dysregulation.

Table 2.4 Continued

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
5.Flannery et al. (2020)			3. Specific gut microbial taxa associate with socioeconomic risk and behavioural dysregulation. 4.Bacteroides fragilis was associated with both higher socioeconomic risk and behavioural dysregulation, and reduced levels of aggressive behaviour, emotional reactivity, externalizing behaviour, sadness, and impulsivity. 5. Butyrate producing taxa, specifically Coprococcus comes, Eubacterium rectale, are positively associated with elevated anxious depression and reduced inhibitory control. 6. Roseburia inulinivorans is associated with a decrease in depressive problems.	
6. Kelsey et al. (2021)	Alpha diversity - Shannon Diversity Index and Chao1, was calculated for both taxa and Chao1 functional terms for virulence factors, resistome, and GO terms) using the vegan R package.	Both taxonomic and functional relative abundance was used.	1. There were no significant associations between alpha diversity (Shannon-taxa/Chao1-taxa) and behavioural temperament measured using the IBQ-R. 2. A mediation analysis suggests that the relationship between increased alpha diversity and negative emotionality may be mediated by homologous-interhemispheric connectivity. 3. There were no significant indirect effects found for the relationship between taxa diversity and regulation/orienting.	1. The study is limited to one time point in early development. 2. A single faecal sample was collected in the home environment and not in the laboratory environment.

Table 2.4 Continued

Authors	Microbiome Diversity Measures	Microbiome Composition Measures	Main Results	Limitations
6. Kelsey et al. (2021)	Beta diversity was measured to calculated as the relative abundance in parts per million of each feature used.		4. There was a significant indirect effect found to suggest that the relationship between virulence factor diversity and negative emotionality may be mediated by homologous-interhemispheric connectivity. 5. A similar significant result was found to suggest that the relationship between virulence factor diversity and regulation/orienting may also be mediated by homologous- interhemispheric connectivity through a negative association. 6. Both negative emotionality and regulation/orienting were marked by an enrichment in Bifidobacterium. Particularly B. pseudocatenulatum was enriched in both the high negative emotionality and high regulation/orienting groups. 7. In a linear model including regulation/orienting, negative emotionality, and surgency as fixed effects, Thermovibrio guaymasensis was identified as a significant biomarker for negative emotionality in the unadjusted model.	3. Although rs-fNIRS was used to measure brain connectivity, as it allows the infant to remain with their mother, it is limited to monitoring activity in superficial structures, and does not allow for measurement of deeper cortical and subcortical structures. 4. By adjusting the model for some confounding factors some of the association effects are no longer statistically significant. 5. Even though participants were instructed to bring the samples into the laboratory within a 24h window, it was not possible to freeze stool samples immediately after collection.

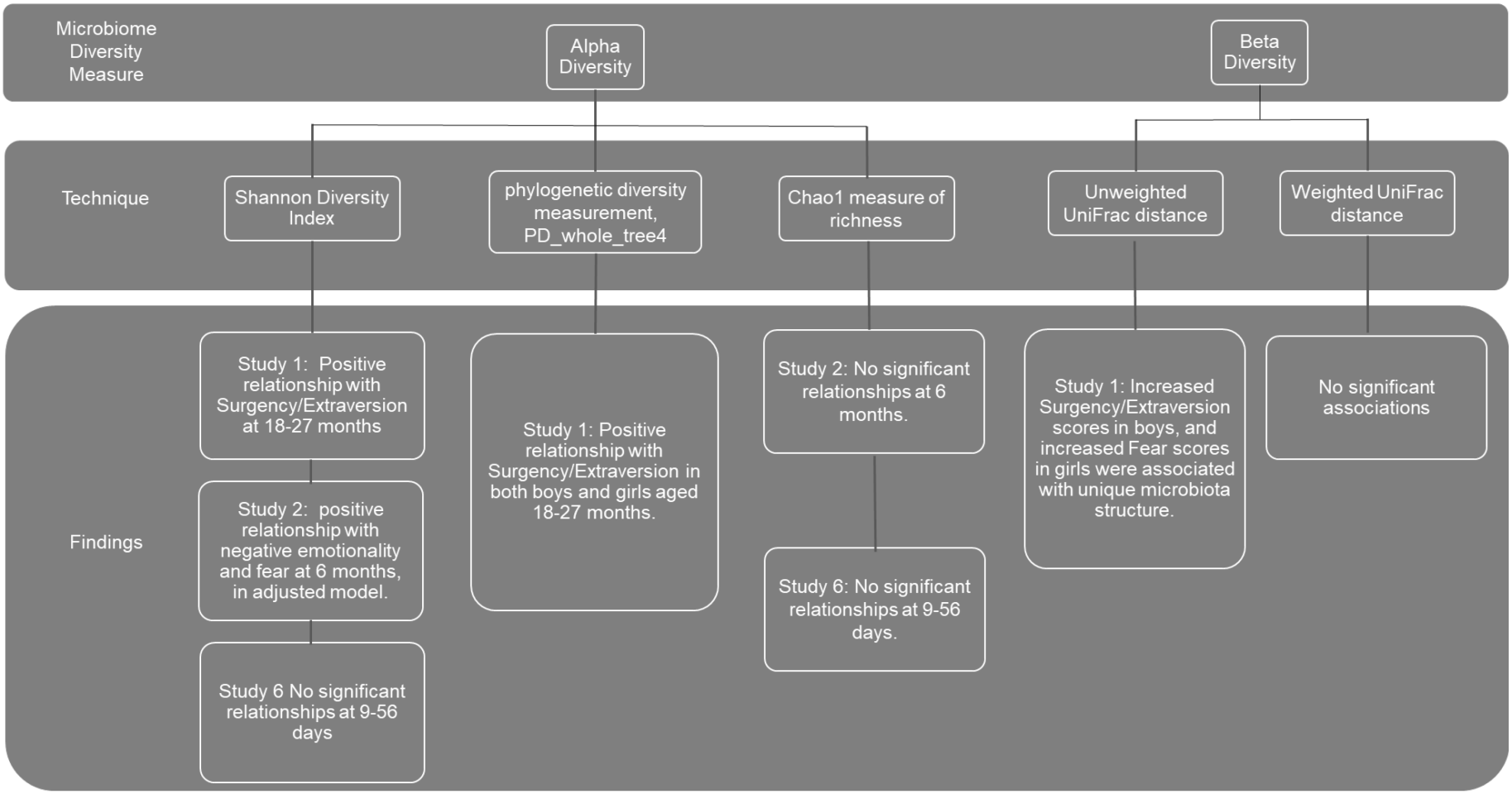


Figure 2.2. Synthesis figure illustrating the relationship between alpha and beta diversity and temperament for each study.

2.3.5.2 Microbiome composition

Several approaches were employed to look at the association of microbiota composition with temperament (See Table 2.4). Study 1 examined genera that made up at least 1% of the total sample by relative abundance, in children aged between 18 – 27 months, which was done in order to focus on the dominant, highly abundant genera (Christian et al., 2015). This included the top 20 and 18 genera in boys and girls, respectively. Notably, temperament sub-scales loading onto the Surgency/Extraversion composite scale were found to correlate with specific genus abundance for boys. Sociability was positively associated with an undefined genus in the *Ruminococcaceae* family, and the genus *Parabacteroides* (see Figure 3 for synthesis of composition results). The genus *Dialster* and an unidentified genus in the family *Rikenellaceae* was positively associated with both High-Intensity Pleasure and activity level. Interestingly, in girls, fear was positively associated with an unidentified genus in the family *Rikenellaceae.* Study 4 established that the abundance of *Bifidobacterium* was positively related to soothability, and *Hungatella* was negatively correlated to cuddliness. This relationship was controlled for delivery mode, feeding type (breast or formula) and probiotic consumption.

Study 2 used a cluster analysis approach, identifying three distinct community types in infants aged 2.5 months that were related to temperament traits at 6-months of age. Within these clusters, five OTUs presented as the most discriminating and represented the taxa *Veillonella dispar, Clostridium neonatale, Bacteroides including Bacteroides fragilis* (named as Bacteroides cluster)*, Enterobacteriaceae,* and *Bifidobacterium.* The cluster dominated by *Bifidobacterium* and *Enterobacteriaceae*, presented the highest scores in the temperament trait of regulation and subscales of High-Intensity Pleasure, Cuddliness, and Duration of Orienting, which is a measure of the time an infant spends paying attention to or interacting with a single object (Putnam et al., 2006). The lowest scores for each of the same trait and subscales was found in *Bacteroides* dominant cluster.

Taxonomic composition analysis carried out in study 5 found (using pairwise comparisons) that *Bacteroides fragilis* was associated with reduced levels of sadness and impulsivity, and increased levels of inhibitory control in children, measured using the CBQ, aged 5 to 7 years. They identified three known butyrate producing taxa, specifically *Coprococcus comes* and *Eubacterium,* that were positively associated with elevated anxious depression and reduced inhibitory control. Interestingly the third butyrate producing bacterium *Roseburia inulinivorans,* was associated with a decrease in depressive problems. Using the shotgun metagenomic technique, Study 5 also investigated the functional capacity of the GM. They found that fear was positively associated with both heme/iron biosynthesis and biosynthesis of melatonin metabolised from tryptophan. Tryptophan metabolism was additionally positively associated with impulsivity.

Study 6 used linear discriminant analysis of effect size (LefSE) to identify 5 microorganisms as potential biomarkers for temperament in infants aged 9-56 days old. Both negative emotionality and regulation/orienting were associated with increased levels of *Bifidobacterium,* specifically increased negative emotionality and regulation/orienting were found in those individuals whose gut microbiome was enriched by *B. pseudocatenulatum*. Further analysis using microbiome multivariate associations with linear models (Maaslin2), found an additional biomarker of *Thermovibrio guaymasensis,* which was associated with negative emotionality.

Finally, in a longitudinal approach, study 3 assessed reverse causation considering associations between early temperament, measured using a 5-point Likert scale, and the candidate bacteria, *Prevotella* and *Lachnospiraceae.* These bacteria were established as candidates for further investigation via earlier examination of the link between microbiota composition and risk of elevated behaviour problems in 2-year-olds. There were no associations found between temperament measured at one, six, and 12-months and presence or abundance of either bacterium. Furthermore, the relationship between normalised abundance of *Prevotella* and behaviour measured at 2-years was not attenuated by adjusting for temperament.

2.3.6 Associations between covariates, the gut microbiome, and temperament.

In addition to the relationship between GM and temperament, several covariates were discussed within 4 of the five studies. Study 3 did not adjust for covariates in their reverse causation investigation of temperament.

Study 1 focused primarily on differences between genders in GM composition and temperament scores and found that there were significant differences between males and females (See Table 2.4). Study 2 also found positive associations between Surgency subscales in boys and relative abundance of *Bifidobacterium* OTUs. In addition to gender, study 2 also considered the potential effects of several covariates, including gestational age, infant age, mode of delivery, breastfeeding status, and antibiotic intake age. Results of adjusted models are presented above. Study 4 investigated gender differences in temperament only and found no significant difference in scores measured on the IBQ-R. Maternal education level was positively related to the temperament measures of soothability. Furthermore, this study controlled for several covariates in their model including delivery mode, feeding type (breast or formula) and probiotic consumption.

Study 5 investigated covariates of gut-related history and diet categories using a daily diary of basic food categories that the child ate at breakfast, lunch, and dinner in the week prior to the laboratory visit. Diet was categorised as the average number of days a child's diet contained food in any one of the following categories: grains, vegetables, fruit, meat, other type of protein, dairy, yoghurt (separate than dairy), beans/nuts/seeds, sugars/fats/oils. In addition, the average number of food categories

(diversity in diet) that a child had per day was also measured. They found that 12.5% of the variation in functional composition, and 25.3% of the taxonomic composition was explained by these variables.

Study 6 included several covariates and found that Shannon-taxa diversity was significantly associated with birthweight, income, breastfeeding, gestational age, and head circumference. There were no associations between Chao1-taxa and any covariates. For functional term diversity, there were significant associations with resistome diversity and income, gestational age, and maternal depression scores. Virulence factor diversity was also significantly associated with income and antibiotics administered at the hospital after birth. Significant associations were also found between the temperament measurement of negative emotionality, and the covariates infant age, and income. The results above present the adjusted models.

2.4 Discussion

The GM composition, diversity, and function, and its relationship with the gut-brain axis, is emerging as an important area of research in understanding the causal pathways of behavioural and mental health problems in later childhood, adolescence, and into adulthood. This systematic review aimed to determine whether there was empirical evidence supporting the relationship between GM diversity and composition and temperament outcomes in children from birth to the age of 6-years 11-months. A total of 6 articles were identified, each from a unique study sample that examined both the GM and temperament in early childhood.

2.4.1 Findings regarding microbiome diversity

The findings from the studies examining alpha diversity fall into two patterns that are distinguishable by age from birth to 12-months, and 12-months and over. Twelve months of age is a significant time of maturation of the gut microbiota: as the diet moves away from milk-based to solid food intake, the microbiota moves towards a more diverse composition in healthy individuals. Aatsinki et al. (2019) and Kelsey et al. (2021) both presented results consistent with a tentative pattern of no significant associations between diversity of the microbiota and temperament outcomes before 12-months of age. Kelsey et al. (2021) did however find significant indirect

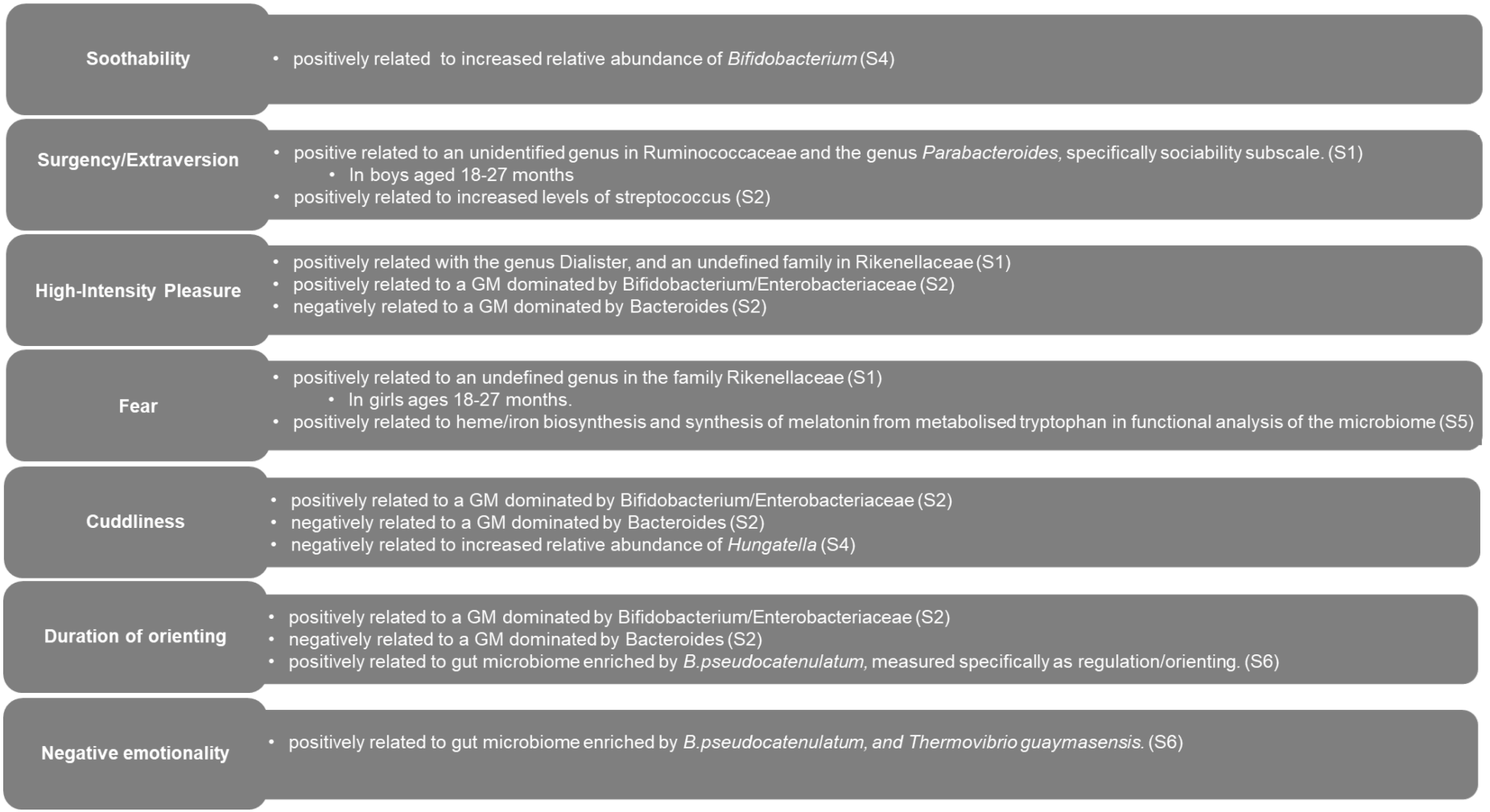


Figure 2.3. Synthesis figure illustrating the relationship between temperament and microbiota composition for each study.

Note: The numbers in brackets refer to the study number to which the results belong.

associations between alpha diversity of taxa (Shannon and Chao1) and negative emotionality, and alpha diversity indices for functional terms (virulence factors) and both negative emotionality and regulation/orienting when mediated by homologous-interhemispheric neural connectivity. The pattern of this relationship suggests that increased connectivity at this stage of development is an aberrant response, which would not be expected later in childhood when increased alpha diversity would be beneficial. The mechanism underlying this warrants further exploration. For example, do increases in the strength of connectivity at that stage of development reflect delayed maturation of usual brain networks or a response to altered microbiome? Kelsey et al. (2021) compare their findings to previous literature that examined the link between alpha diversity and cognitive performance in infants aged between 1 and 2-years of age (Carlson et al., 2018). However, it is important to note that cognitive performance, behaviour, and temperament/personality measure very different aspects of child development, and therefore this limits the conclusions that can be made from comparison of the role of alpha diversity in these differing developmental outcomes.

Findings in children aged over 12-months of age, show that higher alpha diversity was associated with Surgency/Extraversion in both males and females, and High-Intensity Pleasure in males. Higher alpha diversity was negatively associated with Effortful Control in females (Christian et al., 2015). Variation in gut microbiota community structure, measured as beta diversity unweighted UniFrac distances, showed that there is a unique community structure associated with the temperament trait Surgency/Extraversion in males, and Fear in females (Christian et al., 2015). Interestingly, Surgency/Extraversion in males was associated with both higher alpha and unique beta diversity. In summary, there is a very small amount of evidence to support the idea of a link between GM diversity and temperament. The tentative pattern showing no association between temperament and alpha diversity before 12-months of age should be viewed cautiously due to methodological shortfalls in the papers reviewed. These include differences in microbiota analysis selected (including use of both 16S rRNA (Christian et al., 2015; Aatsinki et al., 2019) and shotgun metagenomics (Kelsey et al., 2021) methods), and limited control of important confounding factors (e.g., environmental factors). Of the three papers evaluating these relationships, two studies were conducted in the U.S.A and one was conducted in Finland, but no mention was given to whether participants lived in rural or urban locations, which has previously been shown to be associated with the diversity and richness of the gut microbiota (Salim, Kaplan, & Madsen, 2014; Zuo, Kamm, Colombel, & Ng, 2018). Given that only three papers have measured these relationships and there is a lack of overlap in study design and measures, the patterns of findings are not wholly consistent; more research is needed to increase confidence in the absence or existence of any causal relationships and meaning of these tentative associations. It would be premature to draw the conclusion that the pursuit of further investigations of the relationship between microbiome and temperament prior to 12-months of age are not necessary on the basis of the small amount of work in the field to date. More work is required to investigate the

temporal and causal relationships between microbiome and temperament in these early months and our review highlights a range of factors that are important to consider for optimal study design in future studies.

2.4.2 Findings regarding microbiome composition

In contrast to measures of diversity, which tell us about the number of different taxa found within the microbiota, and the number of functional differences between them, the composition of the GM allows us to identify specific taxa of interest and how they shape the relationship between the gut microbiota and temperament. When investigating the taxonomic composition of the GM and temperament, tentative patterns from the results of the 6 studies identified were found. Significant associations between abundance of *Bacteroides* and temperament were found in two studies (Aatsinki et al., 2019; Flannery et al., 2020). Microbiota dominated by *Bacteroides* in 2.5-month-olds were associated with lower scores of High-Intensity Pleasure, cuddliness and duration of orienting measured at 6-months of age (Aatsinki et al., 2019). Specific associations with increased relative abundance of *B.fragilis*, measured in 5- to 7-year-olds, were associated with reduced levels of sadness and impulsivity and increased levels of inhibitory control (Flannery et al., 2020). Whilst the results of Aatsinki et al. (2019) and Flannery et al. (2020) appear to contradict each other, it should be noted that the composition of the microbiota undergoes large changes between 2.5 months and 1-year, as solid food is introduced into the diet and the microbiota matures. Thus, cross sectional patterns observed in early infancy may not be predictive of later relationships.

Interestingly two studies (Aatsinki et al., 2019; Kelsey et al., 2021) found significant relationships between *Bifidobacterium* and temperament. Kelsey et al. (2021) found that in children aged between 9 and 56 days of life, higher abundance of *Bifidobacterium* were significantly associated with both high negative emotionality and high regulation/orienting. Aatsinki et al. (2019) found similar results, with higher durations of orienting at 6-months of age in infants whose microbiota was dominated by *Bifidobacterium* at 2.5 months of age. Additionally, they found higher scores of High-Intensity Pleasure in those children whose microbiota was dominated by *Bifidobacterium* at 2.5 months of age. When combined, the findings of both Aatsinki et al. (2019) and Kelsey et al. (2021) allude to a potential link between relative abundance of *Bifidobacteria* and emotional regulation. As this was measured in very early infancy, future research should investigate whether the link between gut microbiota and emotional regulation persists through to later childhood given the large amount of variation and change that occurs during the maturation of the infant gut microbiota. Overall, there is a need for more longitudinal research in this area, which would allow for the mapping of *changes* to the microbiota, and the impact this can have upon the development of infant temperament.

Flannery and colleagues (2020) found that two butyrate producing bacteria, *C.comes* and *Eubacterium rectale* were associated with elevated anxious depression and reduced inhibitory control.

Conversely, they also found that another butyrate producing bacterium, *R.inulinivorans,* was associated with a decrease in depressive problems. *Ruminococcaceae,* found to be associated with sociability in boys (Christian et al., 2015), is a family of bacteria also known to produce butyrate. Although these results are somewhat contradictory, these data provide support for the notion that the influence of butyrate producing bacteria upon temperament should be an important focus for further investigation. Future research could focus on butyrate producing bacteria known to colonise the GM, and their overall role in the relationship between GM and temperament. Furthermore, butyrate producing bacteria metabolise complex carbohydrates and dietary fibre and have previously been shown to be beneficial to cognitive function, social behaviour, and mental health in animal models (Stilling et al., 2016). Interestingly, the family *Rikenellaceae,* found to be positively associated with fear in girls (Christian et al., 2015), has been associated with diets high in fat and low in dietary fibre in animal models (Nagano and Yano 2020). Assessment of the functional composition of the GM indicates that metabolism of tryptophan found in the diet is associated with fear and impulsivity measured on the CBQ (Flannery et al., 2020). Tryptophan is consumed in dairy products, proteins, such as turkey and chicken, and nuts and seeds. It is also found in breast milk and is used in the production of melatonin, which is further associated with mood and depressive state (De Crescenzo et al., 2017; Lanfumey, Mongeau, & Hamon, 2013; Srinivasan et al., 2006). However, due to the lack of defined study population and sample size, the quality of the Flannery et al. (2020) paper was judged as poor, and therefore caution exercised in its interpretation. It will be particularly important to replicate such findings before firm conclusions are made. Further investigation should also attempt to probe these relationships by examining the role of diet and the influence this has upon the relationship between GM and temperament.

The main question this review sought to address was whether there was evidence of associations between GM composition and diversity, and temperament in children during early childhood. Although there are some interesting patterns emerging, the evidence is still clearly preliminary and only tentative patterns can be discerned. The findings of this review show that replication and extension of existing research is needed in the field of GM in order unlock more of the potential links with temperament during early childhood. This would then pave the way towards targeted interventions in early childhood that could alter future wellbeing.

2.4.3 Limitations of captured studies and the current review

There were several limitations of the studies that may explain some of the variability in findings, including GM factors (e.g., microbiome analysis technique and hypervariable region chosen), study design, and time-points analysed. Of the 6 studies in this review, two identified that sample size was small. When looking at the quality assessment carried out for all 6 studies, none of the studies identified power calculations or presented sample size justification, although one study (Kelsey et al.,

2021) did provide effect sizes. This is not currently unusual in this field; it is not common practice in GM studies because there is no standard approach for a-priori sample size calculation (La Rosa et al., 2012), which can be a major limitation of this type of study.

Additionally, the selection of hypervariable region for analysis is an important part of the GM analysis pipeline. Of the 4 studies analysing 16S rRNA, three separate combinations of hypervariable region were selected. Selection of the V4, or V4-V5 regions have been shown to alter or even miss the relative abundance of important taxa in samples taken from the young, such as *bifidobacteria* species, and substantially increase the abundance of Firmicutes (Alcon-Giner et al., 2017; Biol-Aquino, Perdiz, Borlagdan, Alcantara, & Mallillin, 2019). This variation in selection of hypervariable region may contribute to the lack of a distinct pattern emerging between GM composition and temperament. Furthermore, the variety of collection, processing, and analysis pipelines used in the studies contained within this review further impede the ability to generalise the results between gut microbiota and temperament. The field of GM analysis is also increasingly moving towards a whole genome or shotgun metagenomic approach, which provides both higher resolution and additional functional information (Jovel et al., 2016). Two studies (Flannery et al., 2020; Kelsey et al., 2021) used a shotgun metagenomic approach to investigate the relationship between GM and temperament, however, neither of these two studies primarily focused on the relationship between GM and temperament. Flannery et al. (2020) included several early childhood environmental exposures, such as quality of caregiving and life experiences, and Kelsey et al. (2021) focused on functional neural connectivity, and the mediating effect this has upon the relationship between gut microbiome and behavioural temperament. Thus, despite the promise of this technique, there are insufficient data to date that reliably explore the association with temperament.

A further limitation of the studies selected in this review was the study design, which in many cases did not allow for discernment of the causal role of the GM upon temperament. The first year of life is a window of critical development of both the GM and neurodevelopment (Knickmeyer, et al., 2008; Carlson et al., 2018; Stewart et al., 2018). Selection of a single measure of both GM and temperament gives only a snapshot of the interaction that is occurring. To discern the causal role of the GM and to measure developmental trajectories, a longitudinal approach with measures taken concurrently for both GM and temperament would be beneficial. Additionally, future studies should carefully consider the role of confounding variables such as diet, gender, and environmental factors known to influence the microbiome.

Finally, regarding the measures of temperament for each study, all studies used a measure that was completed by the mother. Only Aatsinki et al. (2019) identified this as a limitation to their study, stating that choosing maternal reports of temperament may show different results to laboratory-based assessments as maternal measures of child temperament are known to be influenced by the mother's

own temperament and other characteristics (Bayly & Gartstein, 2013). To improve upon this limitation, future studies should consider collecting temperament measures from more than one source, such as additional questionnaires completed by another primary caregivers, or inclusion of laboratory-based observations in addition to parental/caregiver ratings.

This review had some limitations. First is the limited number of studies included, influenced by the low number of studies examining both the GM and temperament. Another limitation is the heterogeneity in the methodologies used across studies, including the data collection and GM analysis pipeline. Most studies used 16S rRNA techniques, however all studies varied in hypervariable region selection, library selection and statistical approach which resulted in synthesis of the results being more challenging. Overall, there was a lack of overlap between measures and study design, which, in combination with the small number of studies, impedes the generalisability of results.

2.4.4 Future research recommendations

The findings of this review highlight key areas for improvement in future research that investigates the association between GM and temperament in infancy and early childhood.

- Development of a standard method to determine sample size and calculate power would vastly improve the field and allow for more consistent and robust GM analysis.
- Increased use of shotgun or whole genome sequencing approaches would allow assessment of the functional role that species play in the development of the gut-brain axis as well as identifying the presence of species within the community.
- Future studies should also employ longitudinal approaches that takes measurements of GM and temperament both concurrently and in series, to establish causal pathways between GM and temperament. This would require careful prospective control for known or theoretically likely confounding variables.
- Inclusion of dietary measures in studies of GM and temperament. Temperament in infancy is linked to diet quality (Lipsanen et al., 2020), in particular, consumption of fewer vegetables and increased consumption of sugar sweetened drinks and desserts, a dietary pattern associated with lower GM diversity and higher colonisation of aberrant species (Martinez, Leone, & Chang, 2017). In contrast, animal models have shown that dietary fibre increases abundance of butyrate producing bacteria. This may highlight the potential for subsequent development of dietary intervention that has relevance to the GM/child temperament association.
- Finally, the tentative association between the butyrate producing bacteria and temperament appears to be an important one which warrants further investigation.

2.5 Conclusion

This systematic review synthesises current evidence for the relationship between temperament and GM diversity and composition in infancy and early childhood. Several tentative patterns have emerged from this review. Firstly, the direct relationship between alpha diversity and differences in community structure, beta diversity, and temperament, is only evident in children over 12-months of age. Secondly, there is some indication that bacteria that metabolise dietary fibre and complex carbohydrates are important taxa of interest when investigating the relationship between GM and temperament. Finally, from the perspective of temperament, the results indicate that there is a link between variation in the diversity and composition of the GM, and both emotional regulation and fear.

Previous research has generally adopted a cross-sectional approach, or included only a single measure of GM, which limits the ability to identify causal pathways in the relationship between GM and temperament. To improve this, longitudinal approaches should be adopted using both serial and concurrent measures. Additionally, most research in this area has used a 16S rRNA approach to investigate the composition of the GM. To gain a deeper understanding of the relationship, future research should consider using whole genome methods to understand functional aspects of the GM, and further investigate the potential metabolomic relationship between the GM and temperament.

CHAPTER 3

FEEDING THE MIND: EARLY LIFE INTERACTIONS BETWEEN DIET, GUT MICROBIOTA, AND PRESCHOOL CHILDREN'S STRENGTHS AND DIFFICULTIES.

ABSTRACT

Background: Emerging evidence of the relationship between the gut microbiota and behaviour has to date only investigated dietary influence as a confounder rather than as a predictor of that relationship. The current study aimed to investigate, through a longitudinal approach, the relationship between the development of gut microbiota across the first year of life and behavioural outcomes in early childhood, whilst examining the influence of early dietary choices in this pathway.

Method: A total of 1074 infants were recruited into the Barwon Infant Study of Geelong Australia. Of these, 324 participants were recruited into a sub-cohort in whom microbiota data were acquired. Faecal samples were collected at 1-, 6-, and 12-months postpartum, and analysed using 16S rRNA Illumina MiSeq, to generate ASVs. Breastfeeding status was recorded at each time point and introduction to solid food was recorded at 6- and 12-months. At 4- years of age the Strengths and Difficulties Questionnaire (SDQ) was reported by the primary caregiver.

Results: The results of this study showed that there was an effect of diet upon behavioural outcomes measured by the SDQ, however this was not via a pathway that was mediated by the GM. It was, however established that the relationship between microbiota and behavioural outcomes is moderated by the first foods introduced into the infant diet.

Discussion: This study furthers our understanding of the complex relationship between development of the gut microbiota, behaviour, and dietary intake in early life, through establishing the moderation effect of diet upon the relationship between GM and behavioural outcomes. For example, higher-than-average consumption of meat at 6-months influences the relative abundance of *Bifidobacteria,* which in turn predicts conduct problems measured at 4-years. Future research should expand to investigate the whole gut microbiota and examine moderation of the relationship between GM and behaviour by diet measured in greater detail.

3.1 Introduction

Globally, the prevalence of emotional and mental health disorders in children is estimated at 1 in 8 which is associated with further impairment and increased risk of mental health disorders as the child matures to adulthood. Identification of causal antecedents of such disorders offers the best opportunity for intervention. One causal pathway of interest is that of the relationship between the gut microbiota (GM), its development and relationship with the gut-brain axis, through which it later influences behaviour and mental health.

The human gut microbiota has been associated with many behavioural, developmental, and mental health outcomes such as Autism Spectrum Disorder (ASD), ADHD, and depression (Adesman et al., 2017; Foster & McVey Neufeld, 2013). Studies of adults with mental health disorders has shown that there is often an associated dysregulation of the GM. For instance, decreased *Fecalibacterium* and *Bifidobacterium,* and increased *Bacteroidetes* and *Proteobacteria* are associated with depression in adults (Aizawa et al., 2016; Jiang et al., 2015). The mechanism for influence is a bidirectional communication route between the gut and the brain known as the gut-brain axis (GBA). The GBA is known to use mechanisms involving the central nervous system (CNS), enteric nervous system (ENS) and the hypothalamic pituitary adrenal axis (HPA) (Skonieczna-Żydecka, Marlicz, Misera, Koulaouzidis, & Łoniewski, 2018). Further explorations of the development of, and interactions between, the gut microbiota and behaviour in human children have shown that there are associations between the composition of the gut microbiota and later behavioural outcomes. The Barwon Infant Study (BIS), found that decreased abundance of *Prevotella* at 12-months of age was associated with increased behavioural problems, measured on Child Behaviour Checklist (CBCL) subscales for Externalizing, Internalizing, and Total Problems, at 2-years of age. Further investigation found that the best predictor of the decreased abundance of *Prevotella* was antibiotic exposure, which highlights potential influences of environmental factors upon the gut-brain axis.

Alterations in the functioning of the HPA in response to early life stress from birth, in children aged between 1 and 6-years of age, have been associated with problems in emotional regulation, temperament, later behaviour and psychopathology (Loman & Gunnar, 2010; Simons, Beijers, Cillessen, & de Weerth, 2015). Dysregulation of the HPA axis is associated with alterations of the composition of the gut microbiota (Frankiensztajn et al., 2020; Misiak et al., 2020; Rosin et al., 2020). Due to the bidirectional nature of the relationship between HPA axis and GM, increased HPA activity, resulting in abnormal levels of cortisol, may also result in altered composition of the GM (Zijlmans et al., 2015). In particular, this may be more likely to occur during early infancy, when both the HPA axis and GM are immature and more susceptible to stress. The HPA axis is activated by the gut microbiota in several ways, including cytokines and prostaglandins. There is evidence alluding to the disruption of the GM, altering the HPA functioning by influencing the hormonal levels and cytokines produced . This

can subsequently lead to anxiety, depression, and stress disorders (Kinlein et al., 2019; O'Mahony et al., 2017; Rosin et al., 2021). Additionally, this relationship between HPA axis and alterations in GM composition have been shown to commence from birth (de Weerth, 2017). This in turn may contribute to the influence of GM upon behavioural and neurodevelopment.

There are many environmental factors that influence the composition of the GM, of which diet has been shown to have the largest effect. Emerging evidence has shown that there is a possible interaction between a typically Western diet and the composition of GM. The typical Western diet is associated with higher levels of animal protein, sugar, starch and fat intake and lower levels of dietary fibre, compared to a non-Western diet, which is associated with high levels of fibre and plant polysaccharides. The crucial time of dietary influence upon microbiota development occurs in the first months from birth, reaching maturity between 36 and 41-months of life (Stewart et al., 2018b). This starts at birth and is heavily influenced by the first milk-based diet, i.e., the introduction of breast or formula feeding. Breastfed infants have been shown to have higher relative abundance of *Bifidobacteria,* and lower abundance of *Bacteroidetes* compared to formula fed infants (Davis et al., 2017). Formula fed infants additionally show higher levels of concurrent alpha diversity (Cresci & Bawden, 2015). However, with the introduction of solid food and the cessation of breast feeding, the GM undergoes a dramatic change. The alpha diversity of breast-fed infants increases dramatically and is then ultimately higher than that of formula fed infants. A longitudinal study of gut maturation conducted in children between 3- to 46-months, found that as breastmilk ceases, the bacterial dominance moves from *Bifidobacteria* to *Firmicutes*. Formula fed infants begin to cluster closer to adult enterotypes sooner than those of infants receiving any breastmilk (Stewart et al., 2018b).

A second dietary phase that has influence upon the GM is the complementary feeding stage, when the first solid foods are introduced, usually from around 6-months of life. In fact, the introduction of the first solid foods during the complementary feeding stage is considered one of the most important events relating to the development of the infant gut microbiota (Fallani et al., 2011a; Zimmer et al., 2012). Despite this there is a paucity of research in this area. One study investigated the impact of the first solid foods introduced during complementary feeding in two international cohorts, one in the Netherlands and the other in Canada. They found that the diversity of foods introduced was positively associated with stability of the microbiota during the complementary feeding stage, which suggests that a varied diet stabilises the GM. However, the size of the cohorts in this study was relatively small n = 15 and 9 respectively, and more evidence is needed to confirm this finding and understand the influence of the complementary feeding stage upon GM maturation. It is during the period of breast/formula feeding cessation and entering the complementary feeding stage, that the microbiota undergoes the largest compositional change, and it is for this reason that understanding the influence of diet during this time is so important to the potential development of interventions to maximise GM health and functioning within the GBA. These data also highlight the potential role of diet influencing

behavioural outcomes through altering composition and diversity of the GM and subsequently the interaction with the GBA, and timely interventions could modify the course of the child's development.

Further to identifying specific timepoints of interest regarding early diet and GM development, there has also been a wealth of literature investigating potential bacteria of interest in this development. It is well documented that during the breastfeeding stage, children who receive exclusive breastfeeding predominantly have a GM composition that is dominated by *Bifidobacterium* (Davis, Dinsmoor, Wang, & Donovan, 2020; Davis et al., 2017). This is a result of the metabolic functioning of *Bifidobacterium,* which lowers the pH of the environment, resulting in a less hospitable environment for other bacteria species to colonise (Donovan & Comstock, 2016). During the milk-based diet, *Bifidobacterium* metabolises Human milk oligosaccharides (HMOs) (Donovan & Comstock, 2016), which occur naturally in breastmilk. Due to the importance of *Bifidobacteria,* formula milks have been developed that replace these HMOs by supplementing with either fructo- or galactooligosaccharides, which allows for colonisation of the GM by *Bifidobacteria.* For this reason, *Bifidobacterium* is a bacterium of interest during investigations of early GM maturation and the influence of diet. Furthermore, De Filiippo et al 2010, investigated the role of Western (Italy) and Non-Western diet (rural Burkina Faso), on the GM of children aged between 1 and 6-years of age. They found that *Prevotella* and *Xylanibacter* were present in the GM of children brought up in Burkina Faso, and completely absent in those brought up in Italy. *Prevotella* and *Xylanibacter* are responsible for the metabolism of cellulose and xylan hydrolysis respectively, and have been associated with plant-rich, high-fibre diets (De Filippo et al., 2010). Diets high in fibre are also associated with a bacteria group that are known as the butyrate producing bacteria. These bacteria metabolise dietary fibre and are responsible for influencing gut barrier function and lowering gut inflammation (Bach Knudsen et al., 2018). Furthermore, they are associated with lowering risk of depression through anti-inflammatory action (Liu et al., 2020), reducing anxiety (Simpson et al., 2021), and in early childhood have been associated with changes in temperament (See chapter 2). Therefore, dietary fibre metabolism and the associated bacteria are of specific interest when looking at the complementary feeding stage and GM maturation.

In addition to influencing the GM, diet in early childhood has been shown to influence behaviour. The ALSPAC study (n=4000), a longitudinal study following UK children from birth, first investigated the effect of a 'junk food diet' on behavioural outcomes. In this study 'junk food' was defined as a diet high in processed foods and soft drinks. Dietary intake was measured using a food frequency questionnaire collected by maternal report when the child was between 37 and 54 months of age. The measure of dietary intake was then used to identify a 'junk food' factor, which in turn was used to further investigate the effects upon behavioural problems in children. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) was used to assess mental health of children on five subscales: emotional symptoms, conduct problems, hyperactivity, peer problems and prosocial

behaviour. They found that there was a strong relationship between ‘junk food’ intake at mean age of 4.5 years and increased hyperactivity measured using the SDQ at 7 years of age. Thus, there appears to be a domain-specific association between diet and behavioural outcomes in children. This result is replicated in several studies and is summarised in a systematic review and meta-analysis by, where a total of 14 studies were found to have investigated the influence of diet upon ADHD. Additionally, this review considered a pre-specified set of confounders including sex, socioeconomic variables, and maternal level of education. The results of the meta-analysis found that an overall healthy diet was protective against the development of ADHD symptoms, whereas an unhealthy diet was indicative of risk of developing symptoms associated with ADHD. Healthy diets were most often characterised as consumption of fruits, vegetables, and whole grains, whereas ‘unhealthy’ diets were characterised as consumption of saturated fats and refined sugars. This is consistent with the result of the ALSPAC study and the notion of ‘unhealthy’ diet aligning with the consumption of ‘junk food.’ Furthermore, these results occurred after adjustment for confounders showing that higher levels of processed food were associated with higher scores indicating symptoms of ADHD (Yan et al., 2018), and likewise similar patterns were seen with increased levels of sweet or fast consumption, which further highlights that it is dietary intake that is driving the associations. However, this is established in only a very specific area, and so it raises further questions about mental health and behavioural development more generally.

To date there have been no longitudinal studies that have investigated in depth the relationship between diet-gut and behavioural outcomes in infancy and early childhood. Although there is some early work on influence of milk-based diet upon the gut microbiota, there is a paucity of literature investigating the complementary feeding stage. Furthermore, there is little work investigating the complementary feeding stage, development of the gut microbiota and behaviour. Examining the dietary components that influence the microbiota might help to identify a causal antecedent of emotional and mental health problems in children and further understand the extent to which the development of the gut microbiota mediates the effect of the diet upon on emotional and behavioural development.

3.2 Aims and Hypotheses

The aims of the study and corresponding hypotheses are as follows:

Aims:

1. To investigate whether early infant and childhood gut microbiota composition characteristics are related to behaviour at 4-years of age, measured using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997).

2. To investigate whether gut microbiota diversity and composition mediate relationships between early childhood diet and behaviour at 4-years.

Hypotheses:

1. Diversity and composition of gut microbiota will be associated with SDQ scores.
    a. Higher diversity at 4 weeks, 6-months, and 12-months of age, and lower relative abundance of genus *Bifidobacterium* and other butyrate producing bacteria, will be associated with a) higher total problem scores, b) higher subscale scores of hyperactivity/inattention, emotional symptoms, conduct problems and peer problems, and c) lower prosocial behaviour subscale scores.
2. Gut microbiota diversity and composition will mediate the relationships between early childhood diet and SDQ scores.
    a. It is predicted that those who are receiving breastmilk, either exclusively or via mixed feeding will have lower alpha diversity scores at 4 weeks, 6-months, and 12-months. It is further predicted that beta diversity will differ between those receiving breastmilk and those not receiving breastmilk. Furthermore, those receiving breastmilk will have higher relative abundance of genus *Bifidobacterium* (at the three timepoints), compared to those not receiving breastmilk.
    b. It is predicted that those children who consume fruits and vegetables, and other fibrous foods, in greater frequency, at 12-months, will have higher relative abundance of butyrate producing bacteria, compared to children who consume those foods in lower frequency.
    c. It is predicted that early diet will be associated with SDQ scores and that the relationship between diet and SDQ scores will be mediated by microbiota composition. Specifically, it is hypothesised that dietary input, characterised by breastfeeding and higher frequency of fibrous food consumption, will predict greater abundance of taxa associated with positive developmental outcomes (as identified in previous literature, e.g., *Bifidobacteria*), which in turn will predict better SDQ scores at 4-years.

NB: It is expected that the number and/or presence of bacteria including butyrate producing bacteria will vary by age from 1-month, to 6- and 12-months. As noted in previous literature, those who are still in receipt of breastmilk are expected to have a microbiota dominated by *Bifidobacteria* and lower alpha diversity until breastfeeding cessation.

3.3 Methods

3.3.1 Study design and participants

Participants were recruited as part of a larger cohort taking part in the Barwon Infant Study (BIS), a population derived birth cohort study, focused on the role of specific environmental factors and early life development (Vuillermin et al., 2015). Participants in the BIS were residents of the Barwon Statistical Division, women pregnant at no more than 28 weeks of gestation at the time of enrolment, planning to give birth at either Geelong Hospital or St John of God Hospital, intending to be available for the duration of the study. Data was collected for both mothers and infants, with 30% of the total cohort randomly recruited as a sub-sample investigating the gut microbiota. The total cohort recruited was 1074, with 324 children being randomly recruited to the microbiota subset. Multiples were included in this sub-cohort, therefore 321 families including 3 pairs of twins were included.

3.3.2 Ethics

Consent was sought from the mothers of all infants and was given in written form prior to participation in the study. The ethics committee at Barwon Health approved the study prior to commencement (reference 10/24).

3.3.3 Microbiota

Stool samples were collected from 324 infants in the microbiota subset at 1-, 6-, and 12-months of age. The participants were asked to collect the samples from infants as close to the appropriate timepoint prior to their in-clinic review. Participants were advised to store the infant samples in their home freezer and then transport to the laboratory using an icebox provided by the study. Some of the samples were collected during the review appointment and were denoted as fresh. There was a 1.6-day median average delay between collection of stool samples and receipt at the laboratory, of these 95% were received within 20 days.

Once in the laboratory, frozen samples were thawed. All samples were aliquoted to approximately 200mg, and then stored at −80ºC for 1-3 years. DNA was extracted using the Qiagen PowerSoil® DNA Isolation Kit, Cat#12888-100. The extracted DNA was then transported to the J. Craig Venter Institute, Rockville, MD, USA. Sequencing was performed on the V4 hypervariable region of the 16S rRNA gene. Amplification of the V4 was performed using adaptor and barcode-ligated V4 primers (forward: 5'-GTGCCAGCMGCCGCGGTAA-3', reverse 5'-GGACTACHVGGGTWTCTAAT-3'). Purification of the amplicons was conducted using Qiaquick PCR purification kit (QIAGEN Inc.), and quantification of the amplicons was conducted using SybrGold, followed by normalisation and pooling in preparation for Illumina Miseq sequencing. The DADA2 v1.12.1 library based upon the standard "big data2" and paired-end big data tutorials was used to determine Amplicon Sequence Variant (ASVs) (Callahan,

McMurdie, & Holmes, 2017). Taxonomy assignment was conducted using the SILVA Nr v128 database as reference. The phylogenetic tree for the ASVs was constructed using PHANGORN v2.5.5 (Schliep, 2011) as described by Callahan, Sankaran, Fukuyama, McMurdie, and Holmes (2016). Samples with fewer than 2500 read pairs and were excluded from further analysis. Microbiota data was managed using the phyloseq package in R , (Team, 2018)

3.3.4 Behavioural measurement

The SDQ is a one of the most commonly used and highly validated instruments for screening psychopathology in children and adolescents (Goodman, 1997). The questionnaire consists of 25 items measuring the dimensions of hyperactivity/inattention, conduct problems, peer relationship problems, emotional symptoms, and prosocial behaviour subscales (See Appendix 1). A total difficulties score is generated based on 20 items measured in the subscales 1-4 not including those items pertaining to prosocial behaviour. A systematic review of the psychometric properties of the SDQ has shown that the questionnaire has good discriminative validity and convergence validity when compared to other scales measuring similar constructs (Kersten et al., 2016). The same review found that the weighted average Cronbach's alpha for the parental version of the SDQ, as used in this paper, was 0.79. The questionnaire has also been validated as an English-Australian version, which is used in this paper.

3.3.5 Dietary measurement

Diet was measured at the time of each microbiota sample, 1-, 6-, and 12-months. Milk based diet was recorded at each time point and was measured during the child health questionnaire and denoted as 1, exclusively breastfed, 2, mixed breast and formula fed infants (any combination of breast and formula feeding in any quantity), 3, exclusively formula fed, 4, other. The measure other was an open question and was to include any other liquids given to the infant as part of their diet, examples of which include, cow's milk, rice milk, and water. Introduction to solid food was measured at 6- and 12-months sample time points. Several food types were recorded in this questionnaire including, fruit, vegetables, nuts, cereals, proteins, fish amongst others. For each food type the participants were asked whether the child had been introduced to that food, whether they still consume the food, and the frequency of consumption over a week period, where 0 = not at all, 1 = less than once a month, 2 = at least monthly but not weekly, 3 = at least weekly but not most days, and 4 =daily/most days.

3.3.6 Data analysis

Data preparation and analysis was performed using the software R. Preparation of the data including quality assessment, removal of duplicates and sub setting was carried out in the *phyloseq* package (McMurdie & Holmes, 2013).

Hypothesis 1: In order to examine the hypothesis that both diversity and composition of gut microbiota were associated with variation in SDQ scores, multiple linear regressions were performed to analyse the relationships between alpha diversity and relative abundance of *Bifidobacterium* and butyrate producing genera in the gut microbiota, with total problem scores and each of the subscales of the SDQ. Alpha diversity was measured as Shannon diversity index (SDI). In order to correct for the number of analyses the Benjamini-Hochberg correction was applied.

Hypothesis 2a: To examine the microbiota alpha diversity and relative abundance of genus *Bifidobacterium* and butyrate producing bacteria in children receiving breast milk, either exclusively, via mixed feeding or not at all, one-way analysis of covariance (ANCOVA) at each timepoint was performed. Furthermore, to examine variation in beta diversity permutational multivariate analysis of variance (PERMANOVA) was conducted.

Hypothesis 2b: To examine the relationship between frequency of fruit, vegetable and other fibrous food intake, and relative abundance of butyrate producing bacteria*,* multiple linear regressions were performed.

Hypothesis 2c: To examine the relationship between early childhood diet, gut microbiota diversity and relative abundance of specific genera, and SDQ scores, a longitudinal multi-mediation structural equation modelling (SEM) approach was used to investigate whether variation in early childhood diet predicts variation in gut microbiota composition and further predicts variation in SDQ scores.

Follow up analyses were conducted to investigate further the relationship between microbiota, SDQ behavioural outcomes and diet. A moderation analysis approach was used, with categorical diet clusters as the moderator. For months 6 and 12, diet was investigated by using a Principal Component Analysis (PCA) and k-means clustering approach. Month one was investigated using breastfeeding status.

3.3.7 Covariates

Candidate covariates were entered into a directed acyclic graph (DAG) , using an online programme DAGitty (Textor, van der Zander, Gilthorpe, Liśkiewicz, & Ellison, 2016), to determine suitable adjustment sets for each analysis: birth weight, sex, number of siblings, mode of delivery, Socio-Economic Indexes for Areas (SEIFA), duration of antibiotic exposure, number of household pets/livestock, known childhood maternal mental health and stress measures when the child is one month old using the Edinburgh Depression Scale (EDS) and Perceived Stress Scale (PSS). Additionally, a dummy coded variable was created to identify samples that arrived at the lab as either fresh or frozen, and models were adjusted accordingly. Figure 1 shows an example DAG generated for hypothesis 1 below (See Appendix 2 for full set of DAGs for all hypotheses).

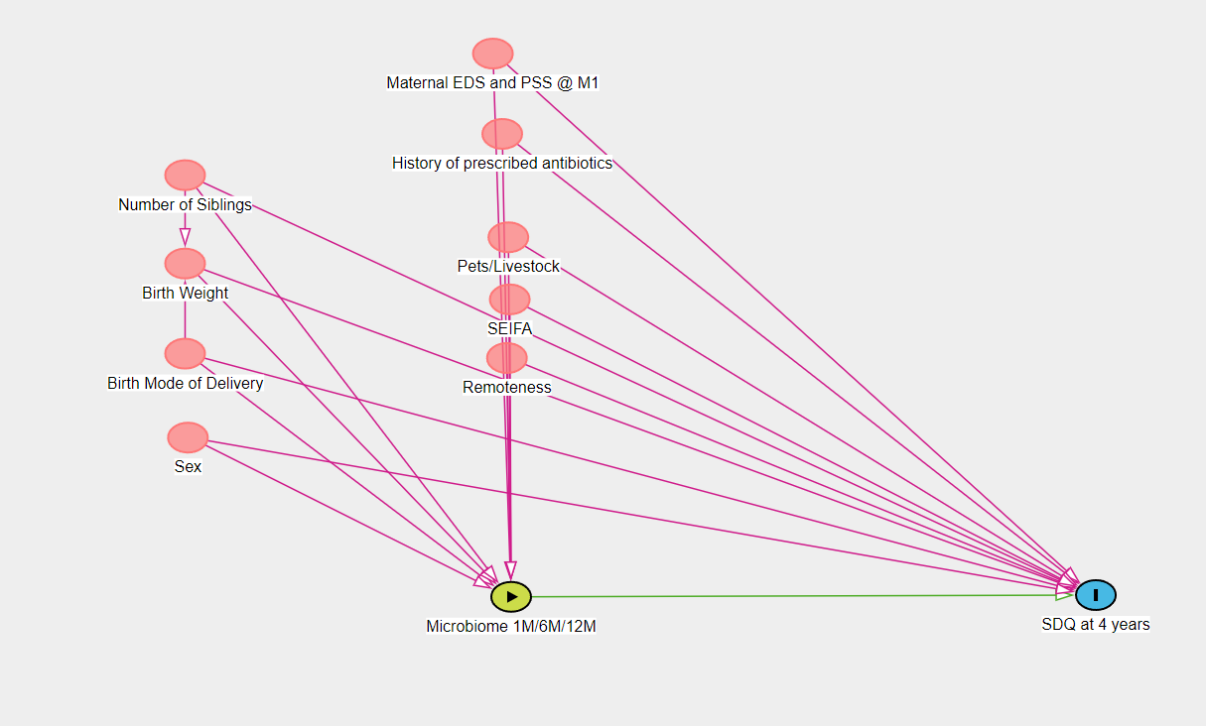


Figure 3.1. DAG for hypothesis 1 identifying covariates to include when investigating the relationship between infant microbiome measured at 1-, 6-, and 12-months, and SDQ measured at 4-years.

3.3.8 Transformations

Microbiota data was prepared for differential abundance testing, taking into account zero-inflation, over dispersion and its compositionality. Zero counts were addressed by adding a pseudo count of one as primary analysis. A centred-log-ratio was generated after normalisation by sequencing, this allows for processing of 16S rRNA data in linear models.

Butyrate producing bacteria were identified using previous literature to generate a list of 12 potential genera of interest. This included *Fecalibacterium, Alistipes, Clostridiodes, Eubacterium, Ruminococcus, Hungatella, Butyricicoccus, Clostridium, Coprococcus, Dorea, Roseburia,* and *Blautia.* These genera were then identified in the dataset and the relative abundance of each was used to create a score for total butyrate producing bacteria, which was used in the analysis as mentioned above.

3.4 Results

3.4.1 Descriptive statistics of study sample

Of the 324 infants randomly selected to the microbiota sub-cohort, 289 microbiota samples were available for one-month-olds following quality control, 298 for six-month-olds, and 251 for 12-month-olds. Table 3.1 presents both characteristics for the full cohort (n=1074), and for each time point. Additionally Figure 3.2 presents heatmaps of each of the bacteria of interest including, *Bifidobacteria,*

Fecalibacterium, Allstipes, Clostridiodes, Eubacterium, Ruminococcus, Hungatella, Butyricicoccus, Clostridium, Coprococcus, Dorea, Roseburia, and *Blautia*. It can be seen that the number of butyrate producing bacteria increases across the first year of life. Additionally, in these samples there is a strong presence of *Bifidobacteria.*

3.4.2 Gut microbiota and SDQ outcomes, hypothesis 1

Increased relative abundance of butyrate producing bacteria measured at 6-months of age was associated with higher scores of both conduct problems (F (12, 185) = 2.606, p < .01, FDR = 0.11, Adj R^2 = 8.9%) and total difficulties (F (12, 185) = 1.984, p <.05, FDR = 0.12, Adj R^2 = 5.7%), however this association was no longer significant following FDR correction for multiple analyses. Furthermore, there were no further significant direct associations between SDQ scores and any microbiota measures at 1-, 6- or 12-months of age. For full results see Table 3.2. This result would indicate that GM measures of alpha diversity, and relative abundance of *Bifidobacteria*, and butyrate producing bacteria are not *directly* associated with SDQ outcomes measured at 4-years.

3.4.3 Gut microbiota and breastfeeding status, hypothesis 2a

To determine if breastfeeding status influenced the microbiota outcomes, a one-way ANCOVA was conducted. Results indicated that breastfeeding status had a significant influence upon the relative abundance of butyrate producing bacteria at one month of age (F (2, 222) = 4.102, p < .05). Post hoc Tukey tests indicate that exclusively formula fed infants had significantly higher levels of butyrate producing bacteria than breastfed, (p < .05). (See Figure 2). There were no significant covariates that contributed to the difference in relative abundance of butyrate producing bacteria between groups of breastfeeding status, which indicates that the differences seen are entirely related to breastfeeding status.

At six months of age, there was a significant effect of breastfeeding status upon alpha diversity (F (3, 214) = 7.471, p < .001), when adjusted for birthweight, mode of birth and sample status upon arrival to the laboratory (fresh vs. frozen). Thus, although several covariates influence the full model, the significant relationship between breastfeeding status and alpha diversity persists when those effects are taken into account. Where models are significant following adjustment for covariates it indicates

Table 3.1. Participant Characteristics for entire birth cohort and at each microbiota sample time point.

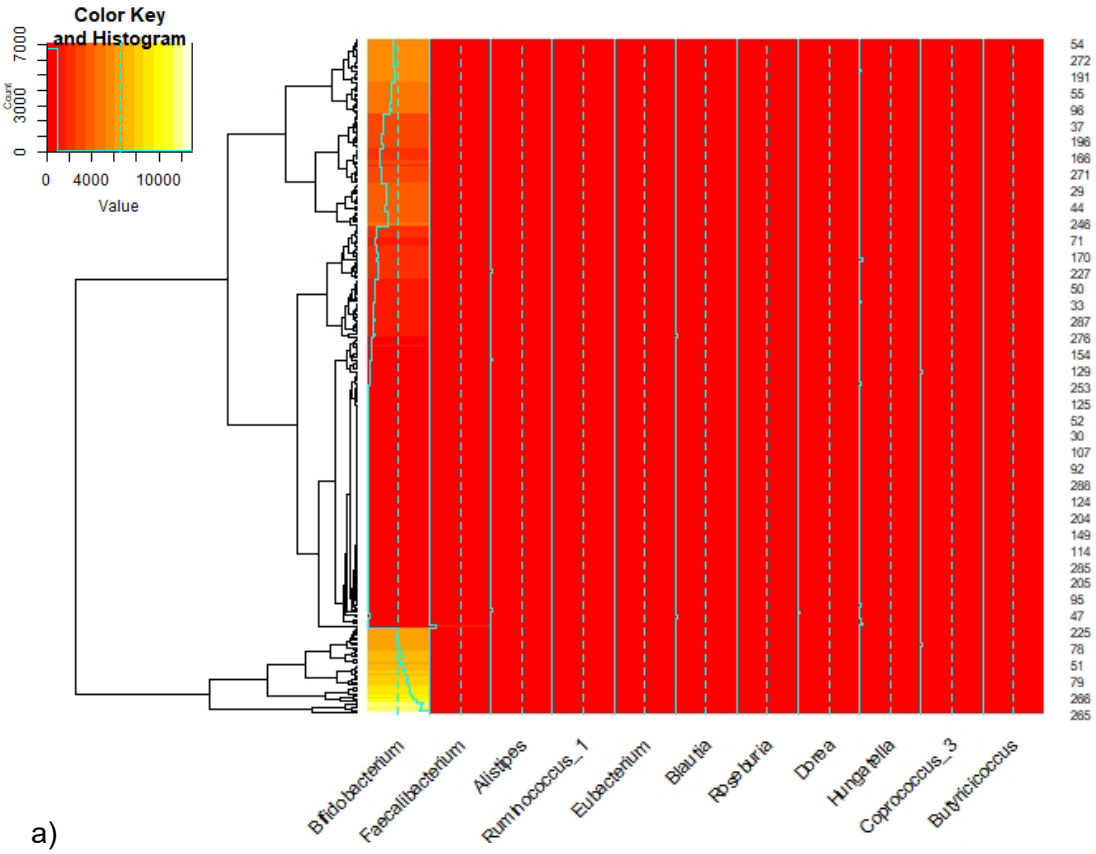
Characteristic	Inception birth cohort (n = 1074)	Study Sample: 1- month (n =289)	Study Sample: 6-month (n= 298)	Study sample: 12-month (n= 251)
Sex of child:				
1. Male	555 (51.7%)	135 (46.7%)	137 (46.0%)	118 (47.0%)
2. Female	519 (48.3%)	154 (53.3%)	161 (54.0%)	133 (53.0%)
Age at time of interview in weeks (mean & SD)		6.1 weeks (1.51)	6.5 months (0.77)	13.0 months (0.76)
Mode of Birth				
1. Unassisted Vaginal Birth	525 (48.9%)	134 (46.4%)	134 (44.9%)	105 (41.8%)
2. Forceps	91 (8.5%)	27 (9.3%)	27 (9.1%)	22 (8.8%)
3. Vacuum	123 (11.5%)	36 (12.5%)	36 (12.1%)	33 (13.1%)
4. Caesarean Planned	178 (16.6%)	57 (19.7%)	63 (21.1%)	60 (23.9%)
5. Caesarean Unplanned	155 (14.4%)	35 (12.1%)	38 (12.8%)	31 (12.4%)
6. No answer	2 (0.1%)	0	0	0
Gestation age in days (mean & SD)	276.1 (10.62)	275.9 (10.33)	275.7 (10.68)	275.5 (9.97)
Birth weight in grams (mean & SD)	3527.3 (519.33)	3550.8 (526.86)	3537.7 (552.79)	3536.8 (542.35)
Number of Siblings				
0	450 (41.9%)	102 (35.4%)	107 (35.9%)	93 (37.1%)
1	382 (35.6%)	114 (39.4%)	118 (39.5%)	100 (39.8%)
2	184 (17.1%)	60 (20.8%)	59 (19.7%)	49 (19.5%)
3	39 (3.6%)	11 (3.8%)	12 (4.0%)	7 (2.8%)
4	12 (1.1%)	1 (0.4%)	1 (0.3%)	1 (0.4%)
5+	4 (0.4%)	1 (0.4%)	1 (0.3%)	1 (0.4%)
No answer	3 (0.3%)	0	0	0
SEIFA (quintiles).				
1	190 (17.7%)	42 (14.5%)	46 (15.4%)	40 (15.9%)
2	230 (21.4%)	72 (24.9%)	68 (22.8%)	61 (24.3%)
3	105 (9.8%)	30 (10.4%)	28 (9.4%)	26 (10.4%)
4	302 (28.1%)	89 (30.8%)	95 (31.9%)	77 (30.7%)
5	244 (22.7%)	56 (19.4%)	61 (20.5%)	47 (18.7%)
No answer	3 (0.3%)	0	0	0
Remoteness				
1. Major Cities of Australia	778 (72.4%)	214 (74.0%)	214 (71.8%)	186 (74.1%)
2. Inner Regional Australia	284 (26.4%)	73 (25.3%)	82 (27.5%)	65 (25.9%)
3. No answer	12 (1.1%)	2 (0.75)	2 (0.7%)	0

Table 3.1. Continued.

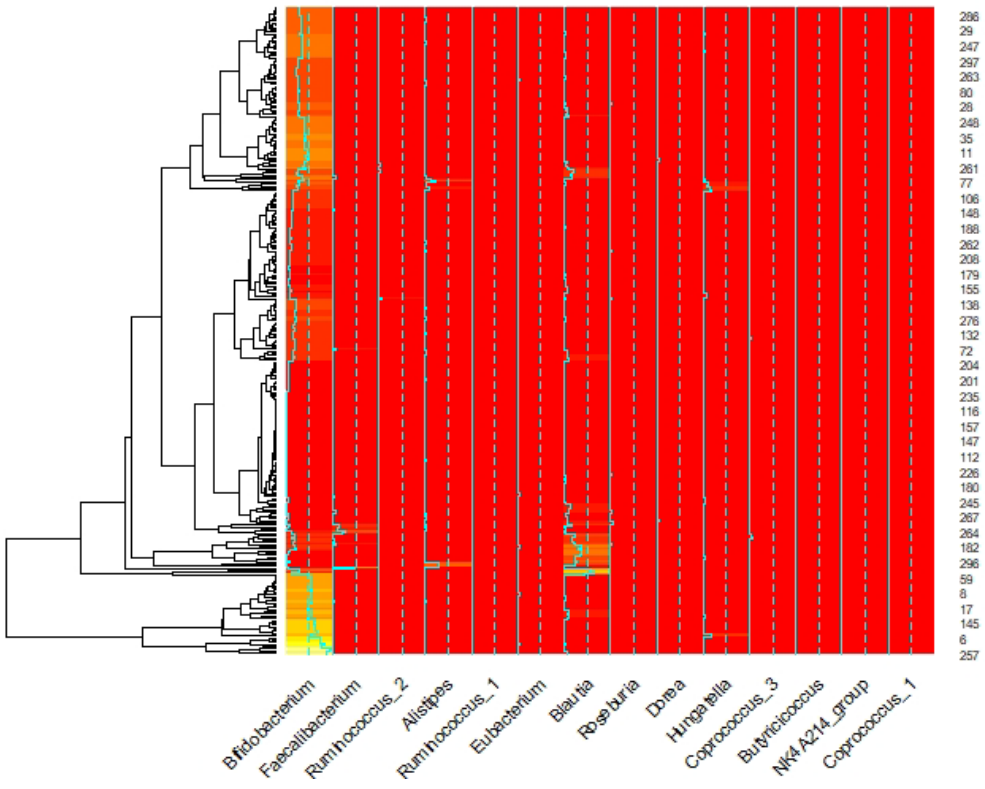
Characteristic	Inception birth cohort (n = 1074)	Study Sample: 1- month (n =289)	Study Sample: 6-month (n= 298)	Study sample: 12-month (n= 251)
Number of animals. (mean & SD)	-	0.74 (0.51)	-	1.68 (1.25)
Antibiotics prescribed				
1. Yes	-	30 (10.4%)	33 (11.1%)	73 (29.1%)
2. No	-	257 (88.9%)	262 (87.9%)	177 (70.5%)
3. No answer	-	2 (0.6%)	2 (0.7%)	1 (0.3%)
Duration of Antibiotics in days (mean & SD).	-	5.8 (2.28)	8.7 (5.90)	8.1 (4.63)
Maternal Perceived Stress Scores (mean & SD)	18.87 (7.43)	18.34 (7.71)	18.39 (7.83)	18.30 (7.70)
Maternal Edinburgh Depression Score (mean & SD)	5.63 (3.99)	5.31 (4.18)	5.29 (4.06)	5.26 (4.03)
SDQ Scores (mean & SD)				
1. Emotion Symptoms	1.52 (1.58)	1.51 (1.52)	1.49 (1.54)	1.62 (1.58)
2. Hyperactivity /Inattention	3.49 (2.28)	3.34 (2.32)	3.32 (2.30)	3.35 (2.22)
3. Conduct problems	1.74 (1.49)	1.72 (1.48)	1.75 (1.50)	1.70 (1.43)
4. Peer relationship problems	1.35 (1.46)	1.44 (1.46)	1.44 (1.46)	1.45 (1.46)
5. Prosocial behaviour	7.66 (1.84)	7.73 (1.79)	7.66 (1.86)	7.73 (1.79)
6. Total Difficulties	8.10 (4.58)	8.01 (4.45)	7.99 (4.45)	8.13 (4.20)
Breastfeeding status				
Exclusively Breastfed	-	189 (65.4%)	129 (43.3%)	77
Mixed Feeding	-	54 (18.7%)	53 (17.8%)	31
Exclusively Formula Fed	-	44 (15.2%)	111 (37.2%)	109
Other	-	0	3 (1.0%)	32
NA	-	2 (0.7%)	2 (0.7%)	2

Table 3.1. Continued.

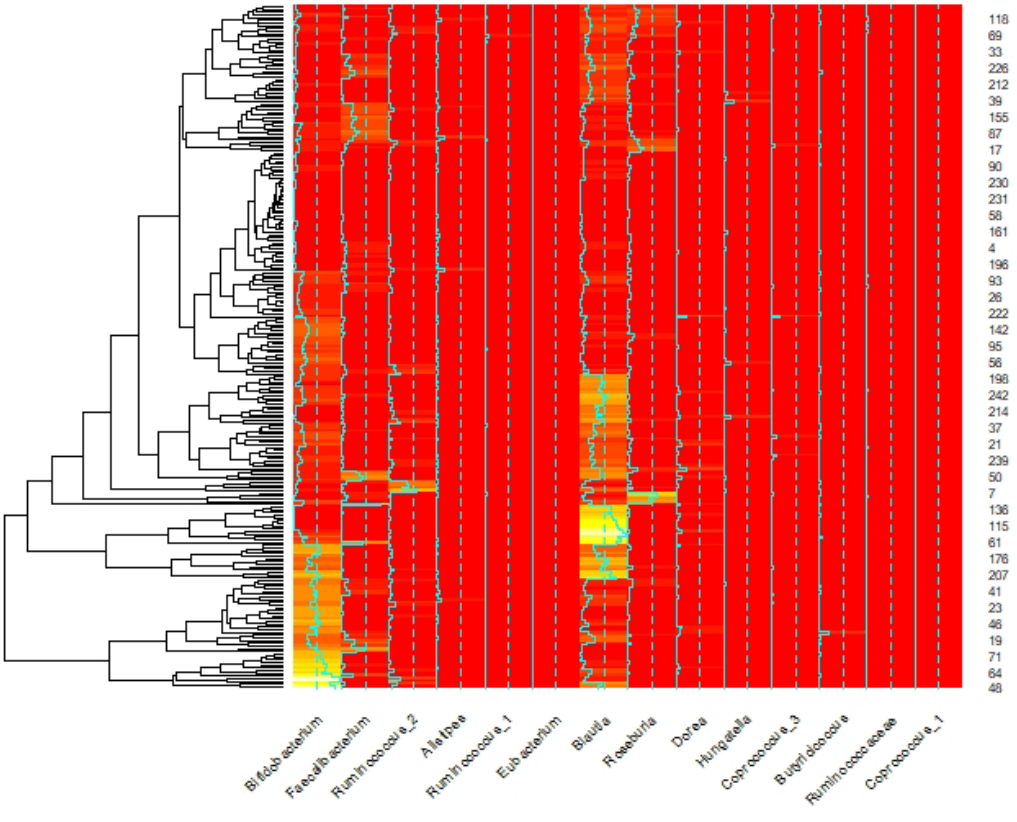
Characteristic	Inception birth cohort (n = 1074)	Study Sample: 1- month (n =289)	Study Sample: 6-month (n= 298)	Study sample: 12-month (n= 251)
Microbiota (mean ASV count)				
Bifidobacteria	-	2494.34	1239.39	554.62
Total Butyrate Producing bacteria	-	192.28	387.49	1345.89



a)



b)



c)

Figure 3.2. Heatmaps representing the ASV counts of each of the bacteria including *Bifidobacteria*, and butyrate producing bacteria of interest at a) 1-month, b) 6-months and c) 12-months of age.

that the direct measured relationship persists and is not better explained by other environmental factors. Post hoc tests indicate that Formula fed infants had significantly higher alpha diversity levels than both breastfed ($p < .001$) and mixed fed infants ($p < .05$). There is also a significant effect of breastfeeding status upon the relative abundance of butyrate producing bacteria (F (3, 214) = 13.048, $p < .001$). There were no significant covariates for this association. Formula fed infants again had significantly higher relative abundance of butyrate producing bacteria compared to both breastfed ($p > .001$) and mixed fed infants ($p > .001$).

At 12-months of age results indicated that there was a significant effect of breastfeeding status upon the relative abundance of *Bifidobacteria* (F (3, 212) = 3.286, $p < .05$), when adjusted for SEIFA scores and sample status upon arrival to the laboratory. Formula fed infants had significantly lower relative abundance levels of *Bifidobacteria* compared to mixed fed infants ($p < .05$). Finally, there was a significant effect of breastfeeding status upon the relative abundance of butyrate producing bacteria (F (3, 212) = 3.251, $p < .05$), when adjusted for maternal scores on the Edinburgh Depression scale. Breastfed infants had significantly lower relative abundance of butyrate producing bacteria than formula fed ($p < .05$).

Unique community structures of microbiota dependent upon breastfeeding status were identified at 1-month of age (F (3, 275) = 2.014, $p < .01$), measured as beta diversity using PERMANOVA applied to Euclidean distances. This was significant when adjusted for mode of birth and gender. This result indicates that at 1-month of age that the composition of the gut microbiota is distinct dependent upon the breastfeeding status, either exclusively breastfed, exclusively formula fed or mixed fed. At months 6 and 12 the results of the PERMANOVA analysis showed no unique community structures dependent upon breastfeeding status.

3.4.4 Gut microbiota and fruits, vegetable, and dietary fibre intake, hypothesis 2b

Total fruit, vegetable and dietary fibre intake was measured as a total score derived from the frequency of intake of fruit, vegetables, pulses, pasta, and nuts. Although included in the solid food introduction questionnaire, the frequency of bread, cereal and rice intake was not included in the fibre intake calculation as it was not possible to distinguish between intake of brown and white varieties. The Manual of Dietetic practice (Thomas & Bishop, 2013) distinguished brown and white bread for example as ‘good’ and ‘poor’ source of dietary fibre, respectively, and it therefore it is not possible to ascertain the benefit of bread to the diet without knowledge of variety.

Linear regression analysis showed that there was a significant positive association between total fruit, vegetable and fibre intake and SDI measured at 6-months of age (F (12, 215) = 2.371, $p < .05$, FDR = 0.04, Adj R^2 = 6.8%), when adjusted for mode of birth. At 12-months of age there was a significant

Table 3.2. Linear regression significant associations between microbiota measured at 1-, 6-, and 12- months and SDQ outcomes measured at 4-years.

Month	Microbiota Measure	SDQ	F	β	DF	DF	P value[b]	Adj R^2	FDR[a]
1	Shannon	Emotion	1.767	0.311	1	238	0.185	0.003	0.286
1	Shannon	Conduct	0.538	-0.168	12	160	0.464	-0.002	0.632
1	Shannon	Hyperactivity	1.715	-0.101	12	160	0.068	0.048	0.165
1	Shannon	Peer	0.527	0.042	12	160	0.895	-0.034	0.939
1	Shannon	Prosocial	0.518	0.131	12	160	0.901	-0.035	0.939
1	Shannon	Total	2.118	0.100	12	160	0.018	0.072	0.111 [d]
1	*Bifidobacteria*	Emotion	2.338	0.198	12	160	0.009	0.085	0.111 [c d]
1	*Bifidobacteria*	Conduct	2.283	0.288	12	160	0.011	0.082	0.111 [d h]
1	*Bifidobacteria*	Hyperactivity	1.784	-0.439	12	160	0.055	0.052	0.165
1	*Bifidobacteria*	Peer	0.554	-0.195	12	160	0.876	-0.032	0.939 [c]
1	*Bifidobacteria*	Prosocial	0.511	0.092	12	160	0.905	-0.035	0.939
1	*Bifidobacteria*	Total	2.119	-0.147	12	160	0.018	0.072	0.111 [d]
1	Butyrate Producing	Emotion	2.463	8.113	12	160	0.006	0.093	0.111 [c d]
1	Butyrate Producing	Conduct	2.225	2.986	12	160	0.013	0.079	0.111 [d h]
1	Butyrate Producing	Hyperactivity	1.851	-11.820	12	160	0.044	0.056	0.150
1	Butyrate Producing	Peer	0.525	-0.137	12	160	0.896	-0.034	0.939 [c]
1	Butyrate Producing	Prosocial	0.653	10.522	12	160	0.794	-0.025	0.939
1	Butyrate Producing	Total	2.117	-0.854	12	160	0.019	0.072	0.111 [d]
6	Shannon	Emotion	1.926	0.487	12	185	0.034	0.053	0.121
6	Shannon	Conduct	2.135	0.305	12	185	0.017	0.065	0.111
6	Shannon	Hyperactivity	1.490	0.599	12	185	0.131	0.029	0.214
6	Shannon	Peer	0.413	-0.139	12	185	0.957	-0.037	0.957
6	Shannon	Prosocial	0.946	0.124	12	185	0.502	-0.003	0.646
6	Shannon	Total	1.758	1.253	12	185	0.058	0.044	0.165 [c]
6	*Bifidobacteria*	Emotion	1.661	-0.224	12	185	0.079	0.039	0.165 [c]
6	*Bifidobacteria*	Conduct	2.001	0.161	12	185	0.026	0.057	0.121
6	*Bifidobacteria*	Hyperactivity	1.395	-0.914	12	185	0.172	0.023	0.273
6	*Bifidobacteria*	Peer	0.620	-0.884	12	185	0.823	-0.024	0.939 [e]
6	*Bifidobacteria*	Prosocial	1.001	0.594	12	185	0.450	0.000	0.632
6	*Bifidobacteria*	Total	1.632	-1.860	12	185	0.086	0.037	0.165 [d]
6	Butyrate Producing	Emotion	1.679	0.799	12	185	0.074	0.040	0.165
6	Butyrate Producing	Conduct	2.606	3.003	12	185	0.003	0.089	0.111 [d]
6	Butyrate Producing	Hyperactivity	1.526	3.132	12	185	0.118	0.031	0.212
6	Butyrate Producing	Peer	0.486	1.331	12	185	0.921	-0.032	0.939 [c]

Table 3.2. Continued

Month	Microbiota Measure	SDQ	F	β	DF	DF	P value[b]	Adj R^2	FDR[a]
6	Butyrate Producing	Prosocial	1.062	-1.902	12	185	0.394	0.004	0.592
6	Butyrate Producing	Total	1.984	8.266	12	185	0.028	0.057	0.121 [d]
12	Shannon	Emotion	1.714	0.223	12	183	0.067	0.042	0.165 [c]
12	Shannon	Conduct	1.986	0.147	12	183	0.028	0.057	0.121 [d]
12	Shannon	Hyperactivity	1.772	-0.265	12	183	0.056	0.045	0.165 [f g]
12	Shannon	Peer	0.539	0.045	12	183	0.888	-0.029	0.939
12	Shannon	Prosocial	0.982	0.156	12	183	0.468	-0.001	0.632
12	Shannon	Total	1.498	0.150	12	183	0.128	0.030	0.214
12	*Bifidobacteria*	Emotion	1.642	0.308	12	183	0.084	0.038	0.165 [c e]
12	*Bifidobacteria*	Conduct	1.963	-0.472	12	183	0.030	0.056	0.121 [d]
12	*Bifidobacteria*	Hyperactivity	1.720	-0.374	12	183	0.066	0.042	0.165 [f g]
12	*Bifidobacteria*	Peer	0.578	-0.732	12	183	0.858	-0.027	0.939
12	*Bifidobacteria*	Prosocial	0.953	0.137	12	183	0.495	-0.003	0.646
12	*Bifidobacteria*	Total	1.510	-1.270	12	183	0.124	0.030	0.214 [d]
12	Butyrate Producing	Emotion	1.639	-0.194	12	183	0.084	0.038	0.165 [c e]
12	Butyrate Producing	Conduct	3.248	1.387	12	183	0.073	0.010	0.165 [e]
12	Butyrate Producing	Hyperactivity	1.942	1.933	12	183	0.032	0.055	0.121 [f g]
12	Butyrate Producing	Peer	0.567	-0.513	12	183	0.866	-0.027	0.939
12	Butyrate Producing	Prosocial	1.002	-0.746	12	183	0.449	0.000	0.632
12	Butyrate Producing	Total	1.609	2.682	12	183	0.092	0.036	0.172

Note. Although the adjusted model may have a significant p value, only those adjusted models with significant microbiota measures have been reported in the text.
[a] FDR – False Discovery Rate

[b] P values presented are the results of 2-tailed testing.

Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 3.3. Linear regression significant associations between total fruit, vegetable and fibre scores intake, and gut microbiota measures at 6-, and 12- months.

Month	Microbiota Measure	F	β	DF	DF	P value[b]	Adj R^2	FDR[a]
6	Shannon	2.371	0.017	12	215	0.007	0.068	**0.042**
6	*Bifidobacteria*	1.362	-0.005	12	215	0.185	0.019	0.264
6	Butyrate Producing	0.972	-0.002	12	215	0.476	-0.001	0.476
12	Shannon	1.364	-0.006	12	214	0.185	0.019	0.264
12	*Bifidobacteria*	2.127	-0.006	12	214	0.017	0.056	0.051
12	Butyrate Producing	1.300	0.006	12	214	0.220	0.016	0.264

Note. Although the adjusted model may have a significant p value, only those adjusted models with significant microbiota measures have been reported in the text.
[a] FDR – False Discovery Rate

[b] P values presented are the results of 2-tailed testing.

negative association between the relative abundance *Bifidobacteria* and total fruit, vegetable, and fibre intake (F (12, 215) = 2.371, $p < .05$, FDR = 0.05, Adj R^2 = 6.8%), however this was no longer significant when applying FDR correction. SEIFA scores were positively associated with SDI and presented as a significant confounder in this model (See Table 3.3).

3.4.5 Longitudinal analysis of Microbiota, diet and SDQ using SEM, hypothesis 2c

A total of 323 participants were included in the longitudinal analysis with microbiota samples at 1 or more of the sample collection points that met the quality assessment. Using Little's MCAR test for missingness (Little, 1988), in the R package *finalfit,* using the missing_compare() function, it was possible to determine the patterns of missingness within the data (See Figure 3.4). It was determined that the pattern of missing was missing at random (MAR), and therefore it is not necessary to impute any missing data points.

Prior to conducting the SEM analysis with multiple mediation, correlations were conducted for each of the covariates, the SDQ outcomes and microbiota measures of interest (See Appendix 3 for full results). In order to streamline the SEM analyses, only those correlations that were statistically significant at p<.05, were included as part of the model. It is a minimum standard of SEM practice to include all potential covariates that are significantly associated with the variables included in the relationships of interest (Rabe-Hesketh, Skrondal, & Pickles, 2004).

Results of the SEM analysis showed that several models were able to achieve goodness of fit according to the Chi square (χ^2), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) (See Appendix 4). There were, however, no models that showed significant mediation effects of microbiota, measured as SDI, relative abundance of *Bifidobacteria,* or total butyrate producing bacteria, upon the relationship between each diet measure and SDQ outcome. From these results, the models that had good fit showed that in many instances there was an effect of diet upon behaviour, however this was not via a pathway that was mediated by the GM.

3.4.6 Moderation effect of diet upon the relationship between microbiota and SDQ.

In a follow up analysis, PCA and k-means clustering were used to create diet clusters based on the frequency of consumption of all foods included in the introduction to solid food questionnaire. Diet clusters were chosen as the best method of examining potential moderation of the relationship between GM bacteria and SDQ outcomes, because it allows for identification of specific dietary patterns of 'types' with which can then be used to examine the moderating effect of these diet types on the relationships between GM and behaviour. Additionally, these clusters allow for the multiple varying patterns of weaning foods that can be introduced. Missing data were identified and the average for each variable was imputed a method that has been approved for use with diet data

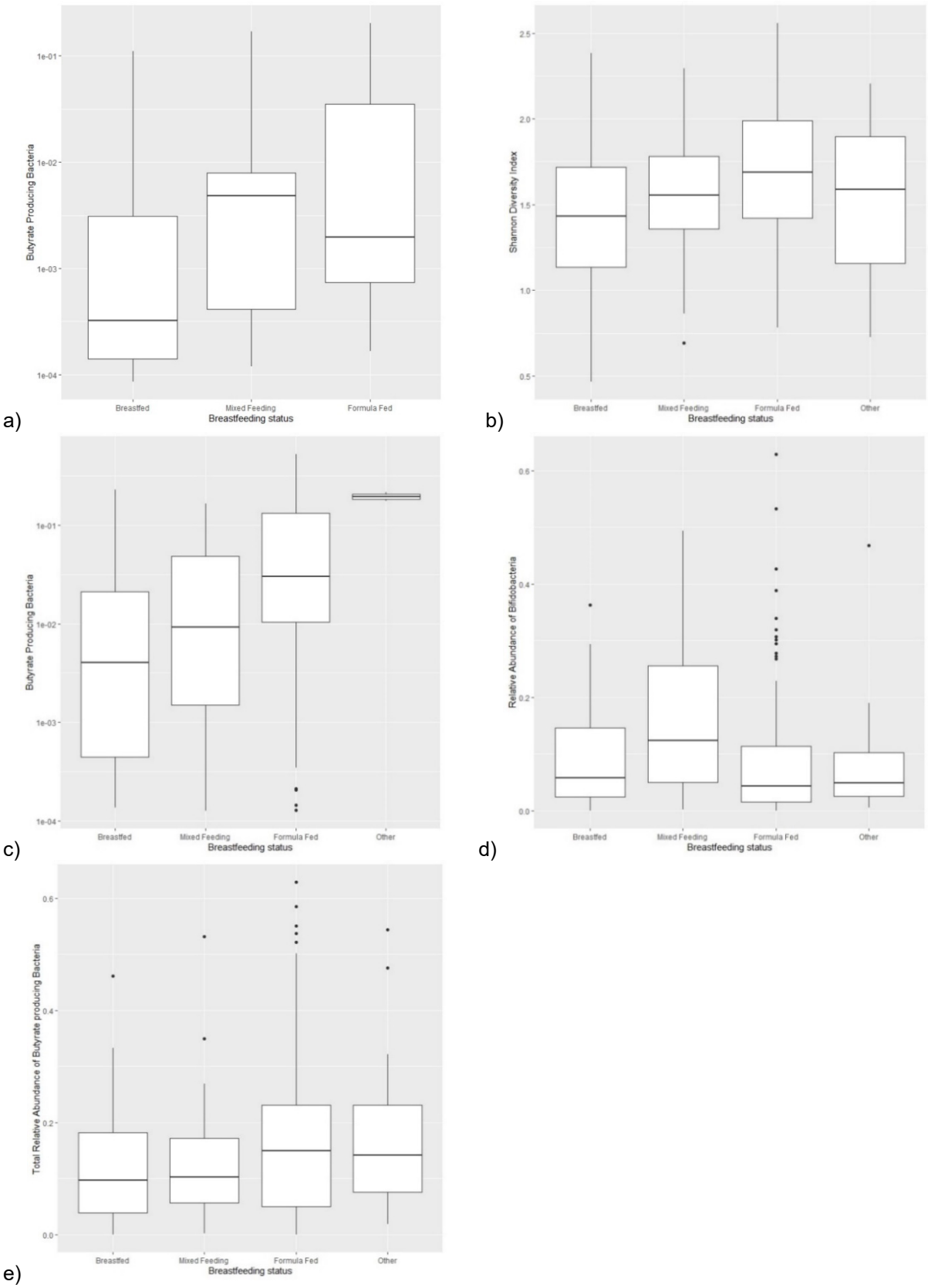


Figure 3.3. Graph to show the significant group differences relative to breastfeeding status for a) butyrate producing bacteria at 1-month, b) SDI at 6-months, c) butyrate producing bacteria at 6-months, d) *Bifidobacteria* at 12-months and, e) butyrate producing bacteria at 12-months.

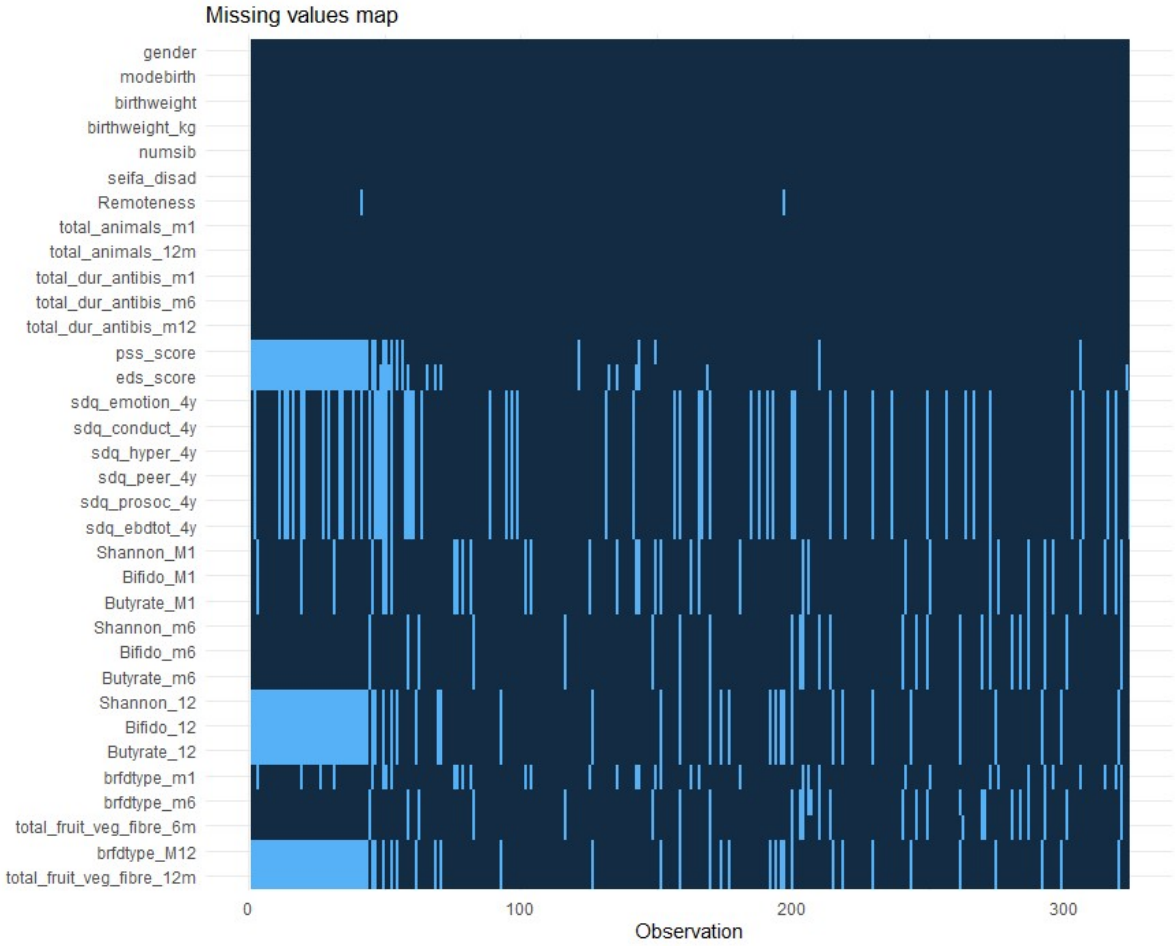


Figure 3.4. Visualisation of missingness for each of the variables included in the SEM analysis, shown to be MAR.

previously (Crozier, Robinson, Borland, & Inskip, 2006). The number of clusters for data at six and 12-month sample times was established through PCA analysis. Participants were then assigned a cluster for each time point through k-means clustering method.

At 6-months of age six clusters were identified from both the scree plot (see Figure 3.5) and the eigenvalues. There were five foods that were removed due to having coefficients <.4, thus not differentiating the clusters at this stage; these were breastfeeding status, cereal frequency, rice frequency, probiotic yoghurt frequency and pre-prepared food (See Appendix 5). Organic food frequency was kept as this was borderline at 0.39. The clusters identified were 2, 159, 75, 23, 21, and 16 participants in size at month 6 and can be described as follows:

Cluster 1 (n=2): Higher than average meat frequency, and slightly higher than average cooked food, lower than average pulses, pasta, yoghurt, fried food, raw foods, eggs, soy, and both sesame products and seeds.

Cluster 2 (n=159): Lower than average fish frequency, soy products and sesame products. Slightly higher than average raw food consumption.

Cluster 3 (n=75): Lower than average raw foods, and cooked foods. Average scores on all other food types.

Cluster 4 (n=23): Lower than average fruit and vegetable consumption. Average scores on all other food types.

Cluster 5 (n=21): Lower than average fish frequency, and much higher than average organic food consumption. Slightly lower than average meat, and pasta consumption.

Cluster 6 (n=16): Higher than average meat, fish, nut, and sesame product frequency. Slightly higher than average pulses, pasta, and cooked food consumption.

At 12-months of age there were 7 clusters identified. Vegetable, egg, and soy frequency was removed due to having coefficients <.4. Pasta and yoghurt with probiotic frequencies remained as they were borderline at 0.39. K-means clustering determined the size of the clusters to be 18, 8, 4, 38, 20, 28 and 135 participants, respectively. The cluster attributes can be described for month 12 as follows:

Cluster 1 (n=18): Much lower-than-average yoghurt frequency, slightly lower than average sesame seed product consumption. Even distribution of exclusive breast and formula fed infants.

Cluster 2 (n=8): Higher than average raw food consumption. Much lower-than-average cereal frequency, and lower than average pulses frequency. Even distribution of exclusive breast and formula fed infants.

Cluster 3 (n=4): Much lower-than-average fruit frequency, but higher than average nut frequency. Slightly higher than average cooked food. This cluster consisted of children who were receiving exclusive formula feeding at 12-months.

Cluster 4 (n=38): Higher than average sesame product, and sesame seed frequency. Slightly higher than average pulses, and organic foods consumption. Predominantly, exclusively breastfed.

Cluster 5 (n=20): Much lower-than-average frequency of yoghurts with probiotics, and lower than average frequency of pulses and nuts. Slightly higher than average frequency of cooked food. Slightly more formula feeding than breastfeeding.

Cluster 6 (n=28): Lower than average raw food, and cooked food frequency. Higher than average consumption of pre-prepared or packaged foods. Predominantly children who were exclusively formula fed or receiving other liquids.

Cluster 7 (n=135): Scores for all variables were close to average. Higher numbers of children who were exclusively formula fed.

Following the identification of the diet clusters these were use as categorical moderators to determine whether types of solid food introduced at 6 and 12-months influence the relationship between microbiota measures and SDQ outcomes.

The results of the moderation analysis showed that there was significant moderation of the relationship between relative abundance of *Bifidobacteria* measured at 6-months and conduct problems measured at 4-years when modified by diet although this was no longer significant when corrected for multiple analyses (F (14, 183) = 2.129, $p < .05$, FDR = 0.263, Adj R^2 = 7.4%), Confounding was present for mode of birth and maternal PSS score. Additionally, there was significant moderation of the relationship between relative abundance of *Bifidobacteria* measured at 12-months and hyperactivity/inattention scores measured at 4-years (F (14, 181) = 1.975, $p < .05$, FDR = 0.263, Adj R^2 = 6.5%), when adjusted for remoteness and SEIFA scores (See Figure 3.6 for interaction plot). This too did not survive correction for multiple comparisons as can be seen in the FDR values. See table 3.7 for full results.

Further exploration of these dietary clusters was necessary due to the insufficient sample sizes, specifically for month six, cluster 1, and month 12, clusters 2 and 3. Selecting for fewer number of clusters in k-means clustering variables did not allow for these small clusters to be combined with larger clusters. Exploration of a second clustering method known as Density based clustering (DBSCAN) (Ester et al., 1996), produced similar results. Manually combining the diet clusters was investigated and it was determined that none of the clusters with small samples would sufficiently fit the dietary profile of the other clusters. It was then decided to remove these individuals as outliers from the analyses. The clusters that are remaining are listed below.

At 6-months

Cluster 1 (n=159): Lower than average fish frequency, soy products and sesame products. Slightly higher than average raw food consumption.

Cluster 2 (n=75): Lower than average raw foods, and cooked foods. Average scores on all other food types.

Cluster 3 (n=23): Lower than average fruit and vegetable consumption. Average scores on all other food types.

Cluster 4 (n=21): Lower than average fish frequency, and much higher than average organic food consumption. Slightly lower than average meat, and pasta consumption.

Cluster 5 (n=16): Higher than average meat, fish, nut, and sesame product frequency. Slightly higher than average pulses, pasta, and cooked food consumption.

At 12-months

Cluster 1 (n=18): Much lower-than-average yoghurt frequency, slightly lower than average sesame seed product consumption. Even distribution of exclusive breast and formula fed infants.

Cluster 2 (n=38): Higher than average sesame product, and sesame seed frequency. Slightly higher than average pulses, and organic foods consumption. Predominantly, exclusively breastfed.

Cluster 3 (n=20): Much lower-than-average frequency of yoghurts with probiotics, and lower than average frequency of pulses and nuts. Slightly higher than average frequency of cooked food. Slightly more formula feeding than breastfeeding.

Cluster 4 (n=28): Lower than average raw food, and cooked food frequency. Higher than average consumption of pre-prepared or packaged foods. Predominantly children who were exclusively formula fed or receiving other liquids.

Cluster 5 (n=135): Scores for all variables were close to average. Higher numbers of children who were exclusively formula fed.

Moderation analyses were then run with 5 dietary clusters included at both 6 and 12 months. The significant moderation of the relationship between *Bifidobacteria* measured at 6-months and conduct problems measured at 4-years by diet was again present, although this was again not significant when corrected for multiple analyses (F (14, 183) = 2.108, $p < .05$, FDR = 0.150, Adj R^2 = 7.3%). There was also a significant moderation of the relationship between *Bifidobacteria* at 12-months and hyperactivity/inattention scores measured at 4-years (F (14, 181) = 1.965, $p < .05$, FDR = 0.150, Adj R^2 = 6.8%), this was again not significant following corrections for multiple comparisons (see table 3.8 for full results and Figure 3.7 for interaction plots).

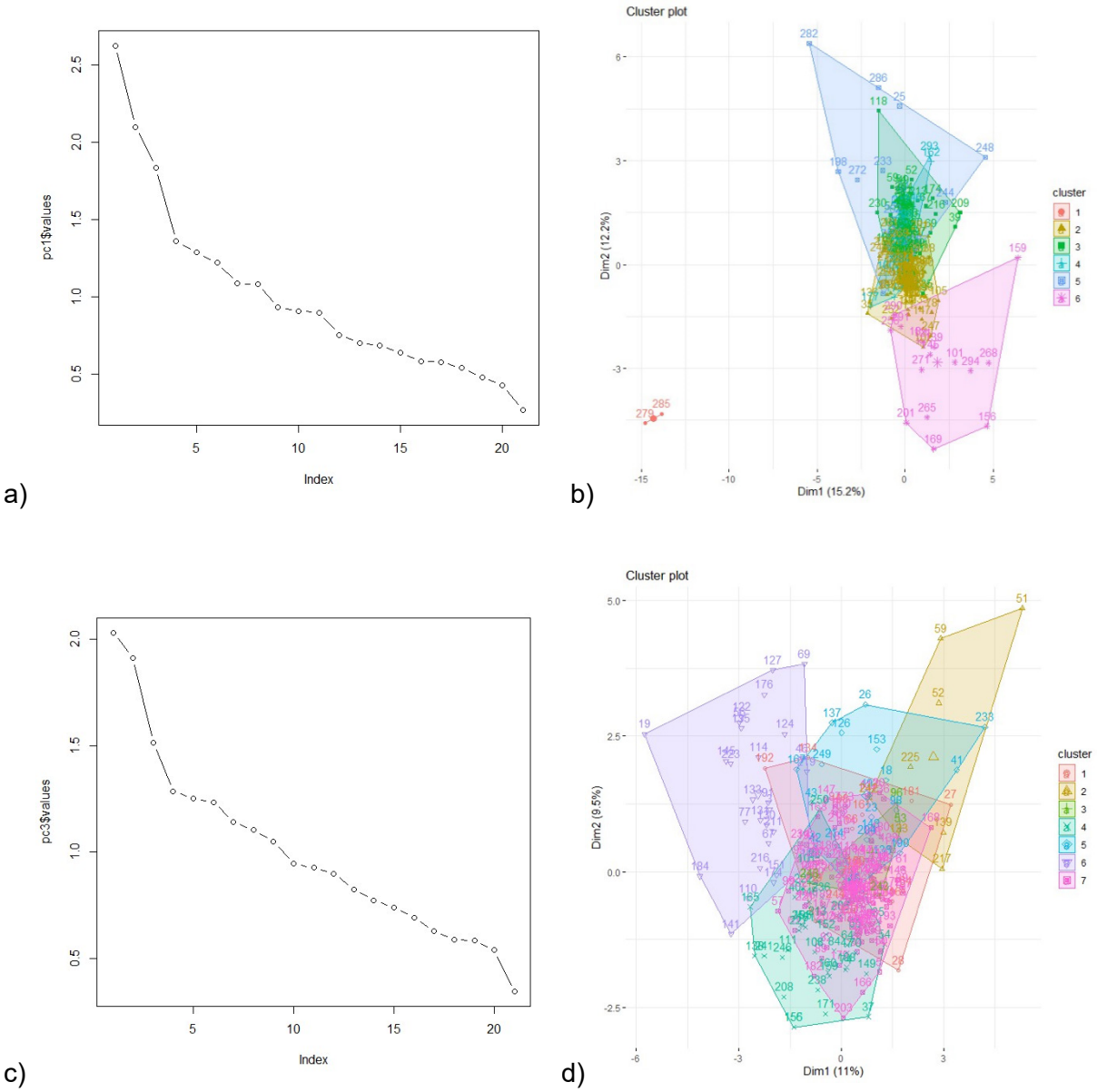
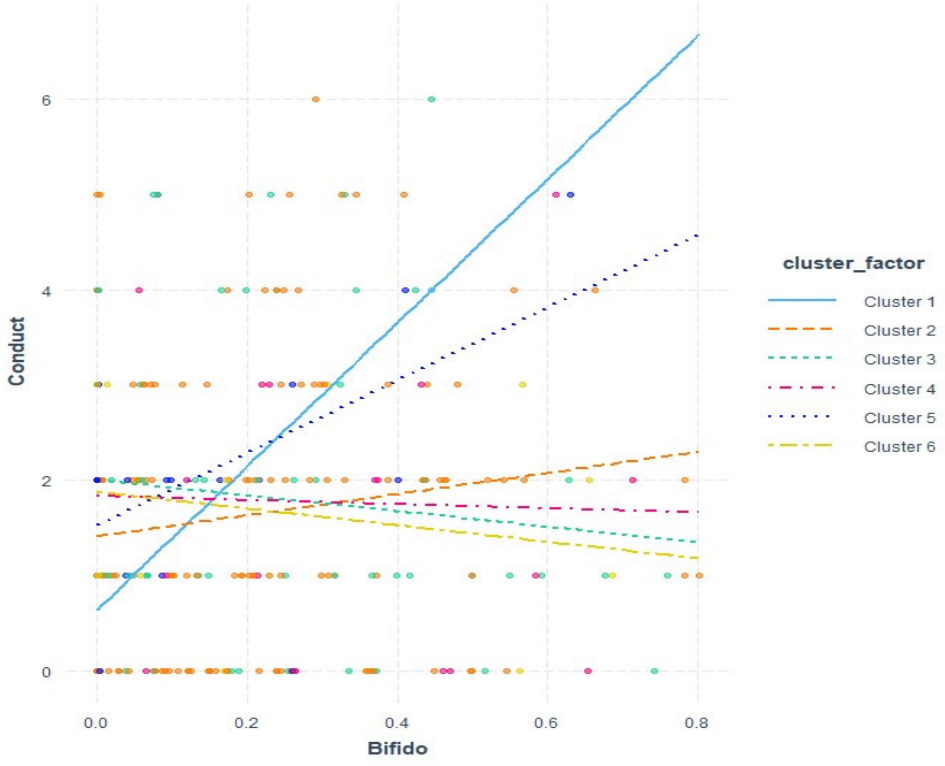
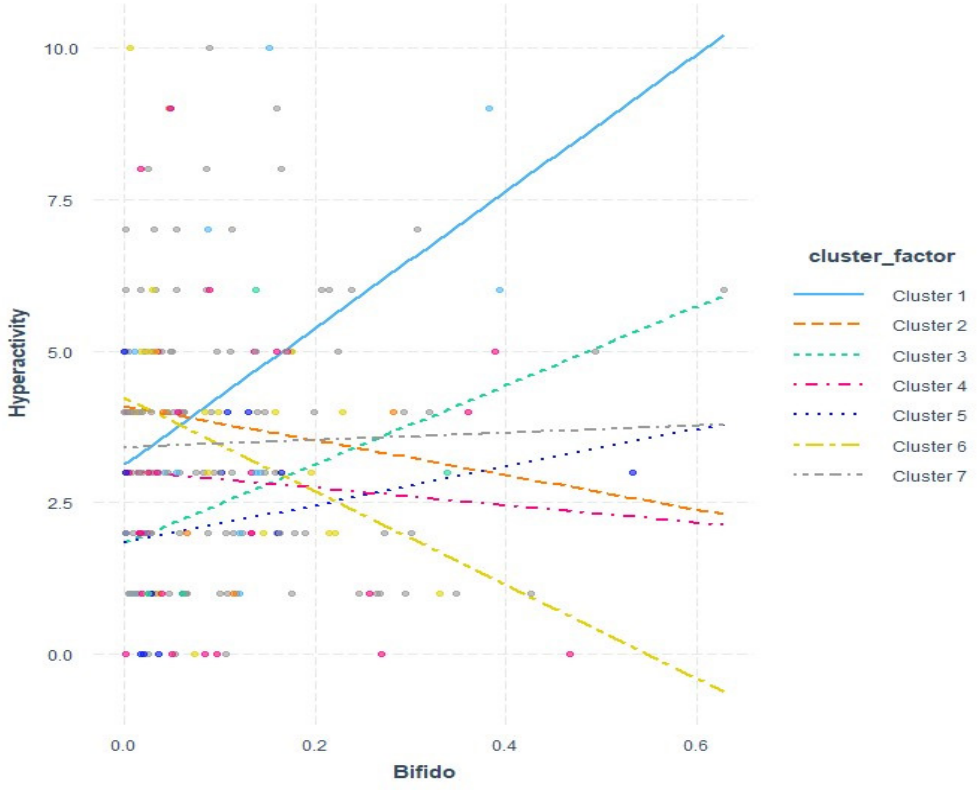


Figure 3.5. Graph to show the significant group differences relative to breastfeeding status for a) PCA Scree plot month six, b) Diet Clusters month 6, c) PCA Scree plot month 12, and d) Diet Clusters month 12.



a)



b)

Figure 3.6. Interaction plots to show the moderation effect of Diet cluster upon the relationship between a), conduct problems subscale and *Bifidobacteria* at 6-months, and b) hyperactivity/inattention and *Bifidobacteria* at 12-months.

Table 3.7. Significant moderation analyses associations between microbiota measured at 6-, and 12- months and SDQ outcomes measured at 4-years moderated by diet cluster.

Month	Microbiota Measure	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
6	Shannon	Emotion	1.783	0.281	14	183	0.044	0.053	0.265 [c]
6	Shannon	Conduct	1.889	-0.214	14	183	0.030	0.059	0.265 [d h]
6	Shannon	Hyperactivity	1.679	0.382	14	183	0.063	0.046	0.265 [g]
6	Shannon	Peer	0.351	0.122	14	183	0.986	-0.048	0.986
6	Shannon	Prosocial	0.835	-0.079	14	183	0.631	-0.012	0.841
6	Shannon	Total	1.553	0.571	14	183	0.096	0.038	0.268
6	*Bifidobacteria*	Emotion	0.977	0.539	3	244	0.404	0.000	0.633
6	*Bifidobacteria*	Conduct	2.129	-0.886	14	183	0.012	0.074	0.263 [d h]
6	*Bifidobacteria*	Hyperactivity	1.399	0.021	14	183	0.157	0.028	0.333
6	*Bifidobacteria*	Peer	0.476	0.026	14	183	0.944	-0.039	0.971
6	*Bifidobacteria*	Prosocial	0.904	-0.357	14	183	0.556	-0.007	0.770
6	*Bifidobacteria*	Total	1.303	-0.303	14	183	0.209	0.021	0.412
6	Butyrate Producing	Emotion	1.450	0.999	14	183	0.134	0.031	0.317 [c]
6	Butyrate Producing	Conduct	1.165	-1.196	14	183	0.324	0.002	0.583 [d h]
6	Butyrate Producing	Hyperactivity	1.577	0.059	14	183	0.089	0.039	0.268
6	Butyrate Producing	Peer	0.408	-0.657	14	183	0.747	-0.007	0.885
6	Butyrate Producing	Prosocial	0.973	-1.218	14	183	0.482	-0.002	0.712
6	Butyrate Producing	Total	1.681	-0.890	14	183	0.063	0.046	0.265
12	Shannon	Emotion	1.435	-0.031	14	181	0.141	0.030	0.317 [c e]
12	Shannon	Conduct	0.802	-0.159	14	181	0.494	-0.003	0.712
12	Shannon	Hyperactivity	1.612	0.097	14	181	0.080	0.042	0.268
12	Shannon	Peer	0.710	0.158	14	181	0.762	-0.021	0.885
12	Shannon	Prosocial	1.117	-0.357	3	212	0.346	0.008	0.590
12	Shannon	Total	1.290	0.065	3	212	0.217	0.020	0.412
12	*Bifidobacteria*	Emotion	0.541	-0.597	14	181	0.655	-0.006	0.842
12	*Bifidobacteria*	Conduct	1.680	0.198	14	181	0.063	0.047	0.265
12	*Bifidobacteria*	Hyperactivity	1.975	-1.821	14	181	0.022	0.065	0.263 [f g]
12	*Bifidobacteria*	Peer	0.644	-0.282	14	181	0.826	-0.026	0.885
12	*Bifidobacteria*	Prosocial	0.656	0.346	14	181	0.814	-0.025	0.885
12	*Bifidobacteria*	Total	1.529	-2.487	14	181	0.104	0.037	0.268
12	Butyrate Producing	Emotion	1.536	-0.667	14	181	0.102	0.037	0.268 [c e]
12	Butyrate Producing	Conduct	2.007	-0.505	14	181	0.019	0.047	0.263 [g]
12	Butyrate Producing	Hyperactivity	1.075	-0.892	3	212	0.361	0.001	0.590
12	Butyrate Producing	Peer	0.632	-0.342	14	181	0.836	-0.027	0.885
12	Butyrate Producing	Prosocial	0.706	-0.276	14	181	0.767	-0.022	0.885
12	Butyrate Producing	Total	1.665	-2.274	14	181	0.066	0.046	0.265

Note. Although the adjusted model may have a significant p value, only those adjusted models with significant microbiota measures have been reported in the text.

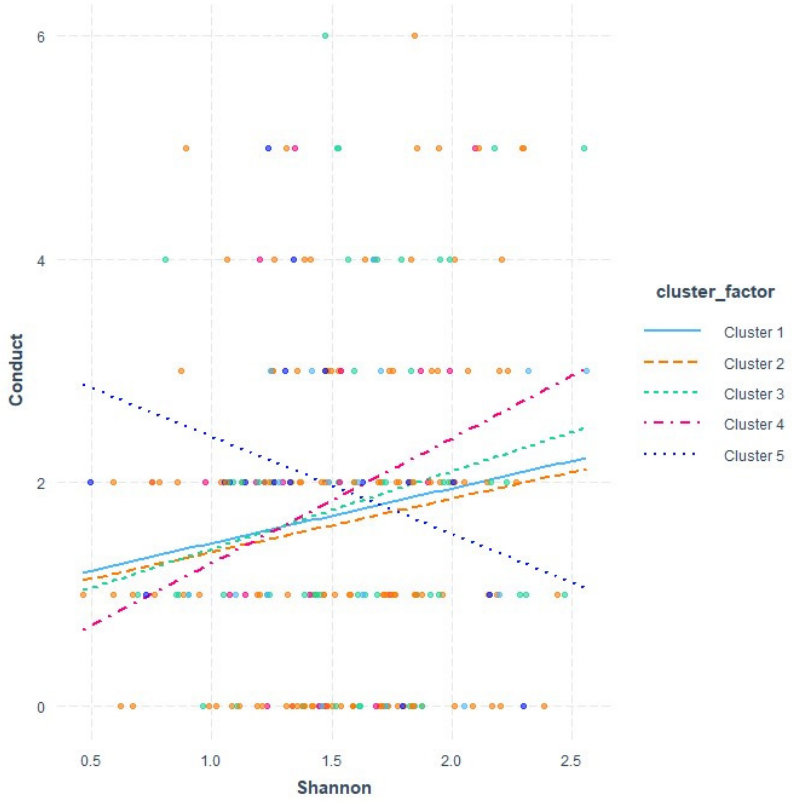
[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 3.8. Significant moderation analyses associations between microbiota measured at 6-, and 12- months and SDQ outcomes measured at 4-years moderated by diet cluster following adjustment for small sample size.

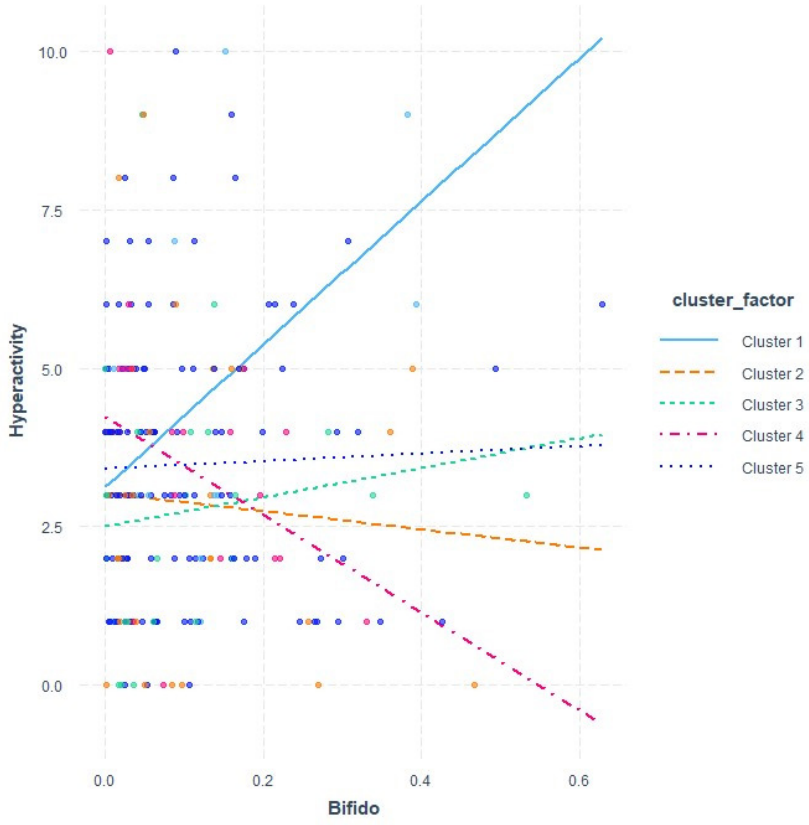
Month	Microbiota Measure	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
6	Shannon	Emotion	1.824	-0.170	14	183	0.038	0.056	0.150 [c]
6	Shannon	Conduct	1.712	0.699	3	244	0.165	0.009	0.150
6	Shannon	Hyperactivity	1.604	0.022	14	183	0.082	0.041	0.183
6	Shannon	Peer	0.455	-0.602	14	183	0.953	-0.040	0.957
6	Shannon	Prosocial	0.973	0.492	14	183	0.483	-0.002	0.621
6	Shannon	Total	1.634	0.004	14	183	0.074	0.043	0.183
6	*Bifidobacteria*	Emotion	1.038	-1.115	3	244	0.377	0.000	0.208
6	*Bifidobacteria*	Conduct	2.108	2.083	14	183	0.013	0.073	0.150
6	*Bifidobacteria*	Hyperactivity	1.327	-1.028	14	183	0.195	0.023	0.293
6	*Bifidobacteria*	Peer	0.516	-0.656	3	244	0.672	-0.006	0.957
6	*Bifidobacteria*	Prosocial	0.973	1.021	14	183	0.483	-0.002	0.621
6	*Bifidobacteria*	Total	1.395	-0.641	14	183	0.159	0.027	0.249
6	Butyrate Producing	Emotion	1.497	-0.433	14	183	0.116	0.034	0.208 [c]
6	Butyrate Producing	Conduct	2.264	5.079	14	183	0.007	0.083	0.150
6	Butyrate Producing	Hyperactivity	1.498	3.460	14	183	0.116	0.034	0.208
6	Butyrate Producing	Peer	0.447	2.516	14	183	0.957	-0.041	0.957
6	Butyrate Producing	Prosocial	1.080	-0.240	14	183	0.379	0.006	0.524
6	Butyrate Producing	Total	1.799	10.622	14	183	0.042	0.054	0.150 [d]
12	Shannon	Emotion	1.659	0.937	14	181	0.068	0.047	0.183 [c e]
12	Shannon	Conduct	1.867	0.112	14	181	0.033	0.061	0.150 [d]
12	Shannon	Hyperactivity	1.862	-1.870	14	181	0.034	0.061	0.150 [f]
12	Shannon	Peer	0.352	-0.697	3	212	0.788	-0.010	0.957
12	Shannon	Prosocial	1.137	1.619	14	181	0.329	0.010	0.474 [f]
12	Shannon	Total	1.459	-1.574	14	181	0.131	0.033	0.225 [d]
12	*Bifidobacteria*	Emotion	1.354	4.498	3	212	0.259	0.005	0.179
12	*Bifidobacteria*	Conduct	0.555	-1.126	3	212	0.645	-0.007	0.150
12	*Bifidobacteria*	Hyperactivity	1.965	8.057	14	181	0.023	0.068	0.150
12	*Bifidobacteria*	Peer	0.239	1.843	3	212	0.869	-0.011	0.957
12	*Bifidobacteria*	Prosocial	0.922	-2.166	14	181	0.536	-0.006	0.644 [f]
12	*Bifidobacteria*	Total	1.705	12.900	14	181	0.058	0.050	0.179 [d]
12	Butyrate Producing	Emotion	1.622	1.915	14	181	0.077	0.045	0.183 [c e]
12	Butyrate Producing	Conduct	1.924	0.079	14	181	0.027	0.065	0.150
12	Butyrate Producing	Hyperactivity	0.124	2.147	3	212	0.946	-0.013	0.183
12	Butyrate Producing	Peer	0.447	2.516	14	181	0.957	-0.041	0.957
12	Butyrate Producing	Prosocial	0.935	1.378	14	181	0.523	-0.005	0.644 [f]
12	Butyrate Producing	Total	1.413	-0.051	14	181	0.152	0.030	0.248 [d]

Note. Although the adjusted model may have a significant p value, only those adjusted models with significant microbiota measures have been reported in the text.

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA,



a)



b)

Figure 3.7. Interaction plots to show the moderation effect of Diet cluster upon the relationship between a), conduct problems subscale and *Bifidobacteria* at 6-months, and b) hyperactivity/inattention and *Bifidobacteria* at 12-months, following adjustment for small sample size in clusters.

3.5 Discussion

The biological pathways underlying the relationship between dietary intake from the very start of life and later behavioural outcomes are becoming an increasing focus of research. Previous literature supports the association between a lack of breastfeeding status and a western style diet, which has an impact upon behaviour, including increasing the likelihood of developing such problems as hyperactivity. This study furthers the understanding of the relationship between dietary intake and behavioural outcomes by investigating the development of the gut microbiota across the first year of life and its role in the relationship between diet and behaviour.

It was hypothesised that both alpha diversity and the composition of the gut microbiota measured at 1, 6-, and 12- months of age would be associated elevated risk of behavioural problems. Specifically, that higher levels of alpha diversity, lower relative abundance of *Bifidobacteria* and decreased total butyrate producing bacteria would be associated with increased total difficulties, hyperactivity/inattention, emotional symptoms, conduct and peer problems and lower prosocial behaviour. It was found that there were no direct relationships between SDQ outcomes and any of the microbiota measures taken at 1-, 6-, or 12- months. This is a somewhat surprising result given the previous evidence that relative abundance of *Bifidobacteria* and butyrate producing bacteria measured in children under 1-year of age has been found to be associated with behaviour such as emotional regulation, and soothability (Aatsinki et al., 2019; Christian et al., 2015) (See also Chapter 2). However, looking at the progression of the colonisation of the infant gut microbiota, as highlighted in figure 3.2, there was low level of colonisation of the gut by butyrate producing bacteria. In fact, it was only at 12 months that any of the selected bacteria were starting to colonise in numbers that would be considered anything other than small. The small number of butyrate producing bacteria present would certainly explain the lack of significant results that can be found to support this hypothesis. Although it is not possible to establish the full extent of direct relationships between the composition of the gut microbiota and behavioural outcomes as only a predetermined set of bacteria were explored. Previous literature has shown that there are several potential additional bacterium types that should be explored further, therefore, to fully explore whether there are any direct relationships between the composition of the gut microbiota over the first year of life and behavioural outcomes measured at 4-years it is necessary to include exploration of the composition the whole microbiota.

In line with hypothesis 2a, there were significant effects of breastfeeding status upon measures of alpha diversity, relative abundance of *Bifidobacteria*, and total butyrate producing bacteria. At one month of age, those receiving exclusive formula feeding had higher total butyrate producing bacteria, which is a similar pattern also found at 6 and 12-months of age. This pattern persists after introduction to solid foods and is indicative of potential accelerated maturation of the composition of gut microbiota in exclusively formula fed infants compared to those who are receiving any breastmilk. At 6-months of

age, exclusively formula fed infants also had significantly higher levels of butyrate producing bacteria compared to mixed fed infants. Additionally, at 6-months of age those receiving exclusive formula feeding also had significantly higher levels of alpha diversity than both mixed fed infants and breastfed infants. This result is in line with previous literature that has shown that as breastfeeding ceases, the microbiota moves to a more mature state and alpha diversity increases (Stewart et al., 2018b). At 12-months of age those not receiving any breastmilk and are receiving formula feeding only in combination with solid foods that have been introduced, had significantly higher levels of butyrate producing bacteria than those receiving exclusive breastfeeding. This again supports previous literature that has suggested that those receiving breastfeeding do not show maturation of the gut microbiota until breastfeeding cessation (Stewart et al., 2018b).

At 12-months, breastfeeding status was significantly associated with relative abundance of *Bifidobacteria*. Interestingly, mixed fed infants had significantly higher levels of *Bifidobacteria* than formula fed infants and, surprisingly, breastfed infants had slightly lower levels of *Bifidobacteria* than mixed fed infants although this difference was not significant. Overall, the pattern of association between breastfeeding status and relative abundance of *Bifidobacteria* was unexpected and is not in line with previous literature. The main energy source for *Bifidobacteria* is the human milk oligosaccharides found in breastmilk (Kitaoka, 2012), which should increase their relative abundance compared to both mixed fed and formula fed infants (Davis et al., 2020; Davis et al., 2017; Savage et al., 2018; Stewart et al., 2018b). Again, this could be due to the selection of the V4 hypervariable region only, as part of the pipeline for microbiota analysis in this study.

The total intake of fruits, vegetables, and dietary fibre, at 12-months of age, was not found to be associated with the total number of butyrate producing bacteria at 12-months of age, and therefore does not support our hypothesis. Again, looking at the numbers of butyrate producing bacteria that are colonising the gut at this time point, there are relatively low numbers, indicating that perhaps colonisation of this type of bacteria is not yet being influenced by fruit and vegetable intake. However, increased fruit vegetable and fibre consumption, at 6-months of age, was associated with higher levels of alpha diversity measured as SDI at 6-months of age. Again, this does not support previous literature that would suggest that increased levels of fruits, vegetables, and dietary fibre, which are a main energy source of butyrate producing bacteria, should lead to increased relative abundance of those bacteria (Knudsen 2005; Ingerslev et al 2014; Knudsen et al., 2018). A potential reason for this result was due to the dietary measure used. In this study it was not possible to ascertain all sources of dietary fibre intake that are considered good sources of fibre (Thomas & Bishop, 2013). In this instance it was not possible to include bread, rice, or cereals in this analysis, due to the ability to discriminate, as a result it might be that an important source of dietary fibre intake at this age group has been missed.

In addition to the findings highlighting the associations between microbiota and SDQ outcome, that showed no significant findings following adjustment, and the significant findings found between dietary intake and microbiota measures, there was no significant mediating effect found of microbiota upon the relationship between diet and SDQ outcomes. Previous literature has shown a significant effect of 'junk food' intake measured at 4.5 years of age and increased hyperactivity/inattention at 7 years of age. However, the measures of the current study were taken at a much younger age and at a time where the microbiota is undergoing significant change therefore it is not wholly surprising that the results are not replicated. Additionally, this study focused on a set of specific bacteria established in a priori hypotheses. These bacteria were chosen as they were associated with diets that have been previously described as healthy, such as those high in fruit, vegetable, and dietary fibre intake, and those receiving breastfeeding for longer. However, these did not include *Prevotella,* which in the same cohort decreased relative abundance was found to be associated with poor behavioural outcomes (Loughman et al., 2020). Another potential bacterium to explore is Veillonella, which has previously been associated with dietary diversity (Homann et al., 2021). Veillonella has previously also been found to have protective influence upon the GM in early life (Matsuyama et al., 2018). Further investigation should explore the complete microbiota to establish microbiota of interest associated with the outcomes measured on the SDQ subscales and total difficulties scores.

Finally, a follow up analysis explored the solid food types that were being introduced to infants at both 6- and 12-months of age. Using PCA analysis and k-means clustering it was possible to cluster the data to examine how specific diet clusters moderated the relationship between gut microbiota measures and SDQ outcomes. At 6-months of age, it was found that diet significantly moderated the relationship between relative abundance of *Bifidobacteria* and conduct problems measured at 4-years. Diets with higher-than-average frequency of meat consumption, and lower than average consumption of nearly all other food types, contributed to the significant moderation effect, which now showed a positive association between conduct problems and increased *Bifidobacteria* abundance. However, this moderation effect was driven by a cluster of only two children, therefore any conclusions drawn from this result can only be viewed with the utmost caution, and more investigation and repetition are necessary to give strength to this result. Interestingly, a second diet cluster that was characterised by lower-than-average fish consumption, much higher than average organic food consumption, and slightly lower than average meat and pasta consumption showed a similar positive association between increased conduct problems and increased *Bifidobacteria* abundance. This cluster consisted of 21 children and can therefore be viewed with more certainty when looking at the overall effect of dietary patterns. To give more confidence in this moderation result, a further investigation of the composition of dietary clusters was conducted. Reducing the number of clusters and exploring different methods of ascertaining the number of clusters to use, did not allow for these two individuals to be combined with any of the other dietary clusters in a satisfactory manner. Therefore, these

individuals were removed and excluded, and the moderation analyses were performed again. A similar pattern of results was found, with diet moderating the relationship between *Bifidobacteria* and conduct problems. The dietary cluster that was characterised by lower-than-average fish consumption, much higher than average organic food consumption and slightly lower than average meat and pasta consumption, presented the same positive association between increased conduct problems and increased *Bifidobacteria* abundance.

At 12- months of age there was a moderating effect of diet upon the relationship between relative abundance of *Bifidobacteria* and hyperactivity/inattention, which was driven predominantly by two clusters. Those with much lower-than-average consumption of yoghurts, and slightly lower than average consumption of sesame seed products had a significant positive effect upon the relationship. Higher relative abundance of *Bifidobacteria* now showed a positive association with hyperactivity/inattention scores. The second diet cluster were those individuals that consumed lower than average amounts of raw food and lower than average cooked foods, higher than average pre-processed or packaged foods, and were predominantly formula fed. This cluster then showed a negative association between relative abundance of *Bifidobacteria* and hyperactivity/inattention. This suggests that higher relative abundance of *Bifidobacteria* in this cluster has a protective element against higher levels of hyperactivity/inattention at 4-years of age. Follow up analyses removing dietary clusters with small sample sizes also found significant moderation effects of diet, with the same significant moderation relationships found at with the removal of cluster 1 at 6-months of age and the removal of clusters 2 and 3 at 12-months of age. This result is significant as pre-processed or packaged foods would often be considered to be part of an unhealthy diet, which has previously been shown to be a factor contributing to the likelihood of ADHD. However, it is not possible to ascertain whether the packaged foods recorded by the parents also included packaged weaning foods, which may not necessarily ascribe to the usual notion of a processed food, high in saturated fats, refined sugars etc., In order to explore this relationship further a more in-depth analysis of dietary intake would be beneficial to explore the types of packaged food consumed in 12-month-olds, and the nutritional value of these.

There are several strengths of this study including longitudinal design, with concurrent microbiota sampling and dietary measures, multiple collection of faecal samples, thorough consideration of potential confounders through use of DAGs, and novel use of statistical analyses with microbiota measures to investigate both mediation and moderation of the gut microbiome. Limitations include insufficient detail to establish all nutritionally good sources of fibre intake and red and white meat separation, microbiota measures limited to the first 12-months, and that this study would benefit from extending measures of microbiota to be concurrent with SDQ outcomes at 4-years. With additional information regarding fibre intake, such as separation of white from brown bread, and white from brown pastas, it would be possible to get a more accurate idea of those foods associated with good

fibre sources and those associated with poor fibre intake (Thomas, 2001). In doing so it is possible to establish a more accurate picture of how these relate to relative abundances of bacteria that metabolise dietary fibre and what this means in terms of microbiota composition. With further understanding of meat consumption, dividing this into white and red meat and fish, it would be possible to better understand the metabolic processes that are occurring and the influence that these various sources of protein have upon the microbiota composition. White meat such as chicken and other poultry is often associated with bacteria that metabolise tryptophan, which has in turn been shown to influence behaviour, temperament, and psychopathology (Gerard Clarke et al., 2013; Russo et al., 2003). Additionally, within analyses that include very early samples of infant gut microbiota there is the possibility that very low numbers of microbes are present of the specific bacteria that are being investigated, therefore it is important to fully appreciate this element and to make cautious references to the reliability of the results that we are seeing. Addressing the final limitation of extrapolating microbiota data from the first year of life and associating this with behaviour at 4-years, with the addition of further a further microbiota sample at the time of the behavioural measurement it would be possible to answer the question of whether there are any bacteria that were previously not present that can account for the current behaviour. Additionally, the microbiota composition is still undergoing changes after 12-months and right up to 36-41-months when this reaches maturity and stabilises to a more adult like microbiota (Stewart et al., 2018b). In capturing the microbiota at 4-years of age a representation of the final mature microbiota of the child would be established, and how it may have altered from that initial changeable period up to 12-months. Additionally, the field of microbiota research is moving toward inclusion of metagenomic techniques to fully understand the functional pathways associated with bacteria present as well as the taxonomic composition. Future investigations would benefit from using these techniques to fully understand the biological mechanisms and establish causal pathways associated with behavioural problems.

This is the first study to look in depth at the relationship between gut microbiota across the first year of life, behaviour outcomes and dietary intake. The results of this study show that there is no effect of the of either alpha diversity, relative abundance of *Bifidobacteria,* or total butyrate producing bacteria as a mediator of the relationship between diet and behaviour. These bacteria in particular were established as prime candidates of investigation from evaluations of previous literature; this particular analytical approach may not be the best way to investigate the relationship between diet, GM, and behaviour. It has been established here, however, that it is possible to better understand the relationship between microbiota and behavioural outcomes, through use of moderation analysis. In particular this study has identified clusters of foods being introduced during the complementary feeding stage, which specify conditions under which the direct relationship between GM and behavioural outcome is altered. For instance, where previously there were no significant associations, it can now be seen that for some patterns of diet, such as elevated levels of pre-processed food, there is a significant relationship

between GM and behaviour as measured on the SDQ. However, this is subject to replication and further expansion to the entire microbiota. Previous literature using the same data set has shown evidence of associations between *Prevotella* and problem behaviours at two years of age measured using the Child behaviour checklist (CBCL) (Loughman et al., 2020), which was established through investigation of the full microbiota. Therefore, further exploration of the full gut microbiota is warranted to fully establish this relationship. Furthermore, this was the first study to look in depth at the stage of solid food introduction, with regard to both behaviour and gut microbiota. Although diet has previously been associated with both behaviour and microbiota composition and diversity it has previously only been included as a covariate when considering the relationship between microbiota and behaviour (Aatsinki et al., 2019; Loughman et al., 2020).

CHAPTER 4

FEEDING THE MIND: MODERATION OF THE RELATIONSHIP BETWEEN GUT MICROBIOTA AND PRESCHOOL CHILDREN'S STRENGTHS AND DIFFICULTIES BY DIET.

ABSTRACT

Background: Diet during the first year of life has been shown to influence both the diversity and composition of the gut microbiota (GM). Furthermore, diet has been shown to moderate the relationship between GM composition and behaviour. The current study aims to identify bacteria of interest and determine if the relationship between GM and behaviour is moderated by diet, introduced during the first year of life.

Method: A total of 1074 infants were recruited into the Barwon Infant Study of Geelong Australia. Of these 324 participants were recruited into a microbiota subset. The SELBAL method was employed to identify candidate bacteria of interest that should be further included in the moderation analyses. Moderation analyses were performed to examine the effect of milk-based diet and frequency of solid food types measured as dietary clusters, upon the relationship between microbiota measured during the first year of life and SDQ outcome measured at 4-years.

Results: There were several significant direct relationships between the GM and SDQ outcomes. Specifically at 6-months of age the relative abundance of *Tyzzerella_4* was positively associated with peer problems, and *Gemella* was negatively associated with externalising problems, at 4-years. Moderation analyses showed a pattern of significant dietary moderation of the relationship between several candidate bacteria identified at all time points and Emotional problems subscale at 4-years.

Discussion: This study furthers our understanding of the relationship between gut microbiota, behaviour, and dietary intake during in early life, through establishing several GM – behaviour relationships that are moderated by early diet. There are indications that 6-months of age, a significant period of dietary change with the introduction of solid foods, is also a critical period of development of the GM for influence upon both internalising and externalising problems at 4-years.

4.1 Introduction

As discussed in Chapter 3, diet, both milk-based and in the form of the first solid foods introduced to an infant during the first year of life, has been shown to influence both the diversity and composition of the gut microbiota (GM). Furthermore, it has been established in chapter 3 that diet moderates the relationship between GM composition measured at 6-, and 12-months of life and behaviour measured using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) measured at 4-years. Specifically, it was found that relationship between relative abundance of *Bifidobacteria* measured at 6- and 12-months of age, conduct problems and hyperactivity/inattention problem scores respectively was moderated by dietary intake, including frequency of meat consumption. These results, however, were only established when investigating selective microbiota measures of alpha diversity, measured as Shannon Diversity Index (SDI), relative abundance of *Bifidobacteria*, and total relative abundance of butyrate producing bacteria. Further exploration of the moderation effect of diet upon the relationship between GM and behavioural outcome, through establishing candidate bacteria from all bacteria present, is necessary to establish dietary intake during the first year of life as an important component of the relationship.

The SDQ has been extensively investigated since its conception in 1997 and has been generally accepted as a validated tool for the evaluation of risk of psychopathology in children These problems can create a large amount of stress for both child and parents, and if they persist can be predictive of adolescent and adult mental health, and behavioural problems (Reef et al., 2010; Tremblay, 2010). Internalising behaviour problems, characterised by heightened anxiety or worry, depressive features, and increased fear on the emotional problem subscale, and inability to interact with peer group, and solitary tendencies on the peer problems subscale, is also a predictor for mental health outcomes such as depression. For these reasons it is essential to understand the underlying mechanisms that lead to these types of problems in early childhood.

Recently, with the emergence of next generation sequencing, there has been a large increase in research investigating the gut-brain axis (GBA), in both mental health in adults and the role it plays in the development of behaviour in children. In adults, the gut microbiota composition has been found to be associated with both symptom severity and treatment outcomes, in individuals with mental illness (A. Madan et al., 2020). Specifically, increased alpha diversity was associated with remission of depressive symptoms in patients that were treated in inpatient facilities. Severe depressive symptoms were further associated with decreased relative abundance of the bacteria *Coprococcus catus*. In children an investigation of the first 10-years of life found that increased relative abundance *Prevotella_9* was positively associated with maternal reports of externalising behaviour at age 10. Similarly, a study investigating the role of microbiota in the development of behaviour in early infancy, found normalised relative abundance of *Prevotella* to be predictive of internalising problems at 2-years

of age (Loughman et al., 2020). Internalising behaviour problems were also found to be negatively correlated with alpha diversity, measured as Shannon Diversity Index (SDI), in children aged 3-5 years of age (van de Wouw et al., 2022). In further understanding the role that the GM plays in mental health and behavioural development, it has been possible for interventions, aimed at improving mental health in adults, to start being developed. These developments establish a link between the GM and both internalising and externalising problems. However, these links have only just begun to be established, and there is a paucity of literature that investigates the potential underlying mechanisms that explain them. The interaction between environment in the form of dietary intake and microbiota, is one potential mechanism that warrants further investigation, and for this reason it is one of the aims of this study.

As diet is a major factor that influences composition and diversity of the GM (Asnicar et al., 2021), this is a primary route of interest to intervene in the underlying GBA mechanisms. One such intervention investigated the role of dietary fibre in regulating the GM and improving mental health in adults (Yan et al., 2021). It was found that in those individuals that were diagnosed with irritable bowel syndrome (IBS), increased dietary fibre intake was associated with improved self-reported anxiety state. This demonstrates the ability of dietary change to improve mental state, although the results do not show the underlying GBA mechanisms involved. Current literature is progressing towards an understanding of the GBA, for instance the influence of the interaction between bacterial metabolic by-products, and potential application towards interventions in adult populations, such as those aimed at increasing fibre intake, however, the status of the GM at these time points are thought to be relatively stable and difficult to change. For this reason, it is also important to understand developmentally early predictors that may be present in childhood, as the potential for beneficial alteration of the composition of the GM is at its highest during the period of maturation. Furthermore, the first years of life are developmentally sensitive periods for the development of both the GM, neurodevelopment, and behaviour (Cowan et al., 2020). By understanding the development of the GM, it is possible to understand its influences upon behaviour and mental health outcomes.

As mentioned in chapter 3, diet has been shown to be a key factor in the development of the GM. Diet has been shown to be associated with variation in GM composition and diversity, from as early as the very first milk-based diet given at birth (Davis et al., 2017; Stewart et al., 2018b). Furthermore, diet has been shown to be associated with a number of developmental outcomes in children such as the link between “unhealthy diets”, specifically those high in junk food consumption, and ADHD. However, in many studies investigating the relationship between GM, behaviour and mental health outcomes, the investigation of dietary intake is often limited to inclusion as a covariate. In chapter 3, it was found that it is possible to investigate in greater depth the relationship between GM composition and behaviour outcomes in early life, through moderation of the relationship by diet. Specifically, results showed that dietary clusters, defined by the first solid foods introduced at 6- and 12- months of age,

moderated the relationship between relative abundance of *Bifidobacteria* and both conduct problems and hyperactivity/inattention. Specifically, at 6-months, those children who clustered into a group that presented with higher-than-average meat consumption were associated with increased conduct problems related to elevated relative abundance of *Bifidobacteria*. Additionally at 12-months, those who consumed more processed or pre-packaged foods showed a pattern of elevated hyperactivity/inattention when the relative abundance of *Bifidobacteria* was decreased. This result aligns with previous literature that associated junk food consumption with ADHD. Following on from Chapter 3, the current study aims to expand the exploration of the relationship between the GM and SDQ outcomes, and further investigation of the moderating effect of diet. This will be achieved through exploration of the composition of the whole gut microbiota, which will identify candidate bacteria of interest. Although the previous chapter identified likely candidates of investigation through extensive review of previous literature, it was not possible to establish all potential bacterial influencers that either have a protective or a negative impact upon the relationship between GM and behaviour. For instance, in this same cohort, decreased relative abundance of *Prevotella,* measured at 12-months, has been found to be related to behaviour measured on the CBCL in 2-year-olds. They used a process to identify bacteria of interest from the whole composition of the microbiota, and then further investigated through logistic regression those bacteria that were significant predictors of risk. This study will employ a similar method to first identify candidate bacteria and then establish the significance and strength of the relationship with SDQ outcomes.

4.1.1 Aims and Hypotheses

The aims of the study and corresponding hypotheses are as follows:

Aims:

1. To investigate whether behaviour measured by SDQ, at 4-years of age is related to composition of the gut microbiota in early infancy, through investigation of the entire composition of the GM.
2. To investigate further whether diet composition during the early milk-based diet, and the complementary feeding stage, moderates the relationship between early infant GM composition and behavioural outcomes measured by the SDQ at 4-years.

Hypotheses:

1. The composition of the GM will be associated with SDQ scores. Bacteria of interest identified at 1-month, 6-months, and 12-months of age will be significantly associated with a) total problem scores, and b) subscale scores of hyperactivity/inattention, emotional symptoms, conduct problems and peer problems, and prosocial behaviour subscale scores.

2. Diet will significantly moderate the relationship between GM composition and behaviour measured using the SDQ at 4-years.

4.2 Methods

4.2.1 Study design and participants

See Chapter 3, section 3.2.1, for more information on Barwon study design and participant recruitment. Three hundred and twenty-four children, from a cohort of 1074, were recruited into a randomly recruited microbiota subset. This subset consisted of 321 families, as three pairs of twins were included. Data was collected from both mothers and infants.

4.2.2 Ethics

Consent was sought from the mothers of all infants and was given in written form prior to participation in the study. The ethics committee at Barwon Health approved the study prior to commencement (reference 10/24).

4.2.3 Measurements

For the microbiota, stool samples were collected from 324 infants in the microbiota subset at 12-months1-, 6-, and 12-months of age. Some of the samples were collected during the review appointment as fresh samples, and others were collected at home and brought to the laboratory. Of the samples that were brought into the laboratory frozen, 95% of these were received within 20 days. For an in-depth explanation of the microbiota pipeline, extraction, and preparation to Amplicon sequence variant (ASV) level please see chapter 3, section 3.2.3.

In order to measure behaviour the SDQ was completed by mothers when the child was 4-years of age (Goodman, 1997). Total difficulties as well as each of the subscales measuring the domains of hyperactivity/inattention, conduct problems, peer relationship problems, emotional symptoms, and prosocial behaviour were completed using the complete 25 item questionnaire (See Appendix 1, and chapter 3, section 3.2.4 for further description).

Diet was measured at the time of each microbiota sample, 1-, 6-, and 12-months. Milk-based diet was recorded at each time point and was measured during the child health questionnaire and denoted as 1, exclusively breastfed, 2, mixed breast and formula fed infants (any combination of breast and formula feeding in any quantity), 3, exclusively formula fed, 4, other. The measure other was an open question and was to include any other liquids given to the infant as part of their diet, examples of which include, cow's milk, rice milk, and water. Introduction to solid food was measured at 6- and 12-months sample time points. Several food types were recorded in this questionnaire including, fruit,

vegetables, nuts, cereals, proteins, fish amongst others. For each food type the participants were asked whether the child had been introduced to that food, whether they still consume the food, and the frequency of consumption over a week period, where 0 = not at all, 1 = less than once a month, 2 = at least monthly but not weekly, 3 = at least weekly but not most days, and 4 =daily/most days.

4.2.4 Data analysis

Data preparation and analysis was performed using the software R. Preparation of the data including quality assessment, removal of duplicates and sub setting was carried out in the *phyloseq* package (McMurdie & Holmes, 2013).

Hypothesis 1: In order to examine the relationship between the whole gut microbiota composition and SDQ outcomes measured at 4-years, bacteria of interest at each time point relating to each of the dimensions on the SDQ questionnaire were firstly identified. In order to complete this the SELBAL method (Rivera-Pinto et al., 2018) was used, which identifies important bacteria of interest in a microbiota sample set, based on relationships between bacterial abundance and either binary or continuous data. There are several methods that exist to identify bacteria of interest from the composition of the whole microbiota, these predominantly use a technique that looks at binary outcome variables. As our output variable of interest is the continuous scores that are generated by the SDQ, the SELBAL method was deemed the most appropriate for the analysis as it allows for this type of data. Following identification of each of the bacteria of interest, the relationship between the relative abundance of each bacterium was analysed with the SDQ scores using multiple linear regressions. These analyses are necessary in addition to the SELBAL identification method as the microbial signature that is identified in the SELBAL method uses a balance approach to identify two potential bacteria of interest, per iteration of the process until the optimal number of bacteria are identified. However, what this does not tell us is the magnitude of association, which is why it is necessary for a linear regression approach to follow bacteria identification in order to establish a significant linear relationship, confirm direction, and also establishes the magnitude of the effect that the bacteria have upon the SDQ outcomes. These regression analyses were also adjusted for potential covariates. In order to correct for the number of analyses the Benjamini-Hochberg correction was applied.

Hypothesis 2: In order to examine whether diet moderates the relationship between the bacteria identified in hypothesis 1 and SDQ scores, multiple moderation analyses were performed. For the age 1-month, Diet was categorised according to milk-based diet as either exclusively breastfed, mixed fed, or exclusively formula fed. At 6- and 12- months of age, the dietary clusters established in Chapter 3, section 3.3.6, were used as moderators (See also Appendix 5 for PCA analysis results). Benjamini-Hochberg corrections were again applied for these analyses.

4.2.5 Covariates

The covariates included in this study were previously identified using directed acyclic graphs (DAGs) using an online programme DAGitty (Textor et al., 2016) to determine suitable adjustment sets for each analysis (See Chapter 3 section 3.2.7, and Appendix 2 for further details).

4.2.6 Transformations

Microbiota data was prepared for differential abundance testing, taking into account zero-inflation, over dispersion and its compositionality. Zero counts were addressed by adding a pseudo count of one as primary analysis. A centred-log-ratio was generated after normalisation by sequencing, this allows for processing of 16S rRNA data in linear models.

4.3 Results

4.3.1 Descriptive statistics of study sample

Complete descriptive statistics of this study sample can be found in chapter 3, section 3.3.1, and Chapter 3 Table 3.1. Of the 324 infants randomly selected, a total of 289 microbiota samples were available for one-month-olds following quality control, 298 for six-month-olds, and 251 for 12-month-olds.

4.3.2 GM bacteria extracted relating to SDQ outcomes using the SELBAL method.

The SELBAL method was used to extract bacterial genera from the participant microbiota that is predictive of the response variable of interest. In this instance the response variable SDQ, measured at 4-years of age, was used to identify the number of genera with the highest prediction or classification accuracy. Initially the algorithm searches through for the first two genera, are added at each step. This continues until no more genera can be added. Finally, a cross validation procedure is performed to confirm the robustness of the balance that is identified (Rivera-Pinto et al., 2018). In this sample selection of bacteria genera was performed for each SDQ subscale, and total difficulties score, at each microbiota sample time point of 1-, 6-, and 12- months of age. The bacteria that were extracted for each time point is presented in Tables 4.1, 4.2, and 4.3. Specific to the SELBAL method, those denoted as numerators are positively associated with the SDQ subscale and denominators are negatively associated. Furthermore, Table 4.4 and Figure 4.1 presents descriptive statistics of the mean average count of bacteria for each genus of bacteria extracted at each time point. In addition to comparison of the relative abundance of each of the bacteria selected, the global balance parameters.

Table 4.1. Bacteria extracted for each SDQ subscale outcome using SELBAL method at one month of age.

Month	SDQ	ASV	Family	Genus	Numerator/ Denominator
1	Emotion	ASV00178	Staphylococcaceae	*Staphylococcus*	Den
1	Emotion	ASV00259	Micrococcaceae	*Rothia*	Den
1	Emotion	ASV00007	Lachnospiraceae	*Blautia*	Den
1	Emotion	ASV00419	Enterobacteriaceae	*Salmonella*	Den
1	Emotion	ASV00072	Lactobacillaceae	*Lactobacillus*	Den
1	Emotion	ASV00089	Coriobacteriaceae	*Collinsella*	Den
1	Emotion	ASV00001	Enterobacteriaceae	*Escherichia/Shigella*	Den
1	Emotion	ASV00768	Enterobacteriaceae	*Pseudocitrobacter*	Den
1	Emotion	ASV00273	Eubacteriaceae	*Eubacterium*	Den
1	Emotion	ASV00022	Erysipelotrichaceae	*Erysipelatoclostridium*	Num
1	Conduct	ASV00086	Ruminococcaceae	*Flavonifractor*	Den
1	Conduct	ASV00051	Peptostreptococcaceae	*Intestinibacter*	Num
1	Conduct	ASV00768	Enterobacteriaceae	*Pseudocitrobacter*	Num
1	Hyperactivity	ASV00028	Enterococcaceae	*Enterococcus*	Den
1	Hyperactivity	ASV00014	Streptococcaceae	*Streptococcus*	Den
1	Hyperactivity	ASV00043	Clostridiaceae_1	*Clostridium*	Den
1	Hyperactivity	ASV00178	Staphylococcaceae	*Staphylococcus*	Den
1	Hyperactivity	ASV00259	Micrococcaceae	*Rothia*	Den
1	Hyperactivity	ASV00011	Veillonellaceae	*Veillonella*	Num
1	Peer	ASV00066	Lachnospiraceae	*Hungatella*	Den
1	Peer	ASV00178	Staphylococcaceae	*Staphylococcus*	Den
1	Peer	ASV00060	Enterobacteriaceae	*Citrobacter*	Num
1	Peer	ASV00768	Enterobacteriaceae	*Pseudocitrobacter*	Num
1	Peer	ASV00273	Eubacteriaceae	*Eubacterium*	Den
1	Peer	ASV00165	Enterobacteriaceae	*Pluralibacter*	Den
1	Peer	ASV00005	Ruminococcaceae	*Subdoligranulum*	Den
1	Peer	ASV00461	Family_XI	*Peptoniphilus*	Den
1	Peer	ASV00282	Lachnospiraceae	*Epulopiscium*	Den
1	Peer	ASV00047	Veillonellaceae	*Megasphaera*	Den
1	ProSocial	ASV00002	Bifidobacteriaceae	*Bifidobacterium*	Den
1	ProSocial	ASV00060	Enterobacteriaceae	*Citrobacter*	Den
1	ProSocial	ASV00738	Bifidobacteriaceae	*Scardovia*	Num
1	Total Difficulties	ASV00001	Enterobacteriaceae	*Escherichia/Shigella*	Den
1	Total Difficulties	ASV00086	Ruminococcaceae	*Flavonifractor*	Den
1	Total Difficulties	ASV00014	Streptococcaceae	*Streptococcus*	Num

Table 4.2. Bacteria extracted for each SDQ subscale outcome using SELBAL method at six months of age.

Month	SDQ	ASV	Family	Genus	Numerator/ Denominator
6	Emotion	ASV00314	Coriobacteriaceae	*Eggerthella*	Den
6	Emotion	ASV00003	Verrucomicrobiaceae	*Akkermansia*	Den
6	Emotion	ASV00624	Fusobacteriaceae	*Fusobacterium*	Den
6	Emotion	ASV00054	Prevotellaceae	*Prevotella_9*	Den
6	Emotion	ASV00077	Ruminococcaceae	Ruminococcaceae_UCG.014	Den
6	Emotion	ASV00437	Desulfovibrionaceae	*Desulfovibrio*	Num
6	Emotion	ASV00020	Lachnospiraceae	*Roseburia*	Den
6	Emotion	ASV00030	Lachnospiraceae	*Fusicatenibacter*	Num
6	Emotion	ASV00085	Lachnospiraceae	*Tyzzerella_4*	Num
6	Emotion	ASV00119	Lachnospiraceae	*Lachnospira*	Den
6	Conduct	ASV00040	Lachnospiraceae	*Dorea*	Num
6	Conduct	ASV00097	Actinomycetaceae	*Actinomyces*	Den
6	Conduct	ASV00351	Family_XI	*Gemella*	Den
6	Hyperactivity	ASV00034	Peptostreptococcaceae	*Romboutsia*	Den
6	Hyperactivity	ASV00384	Carnobacteriaceae	*Granulicatella*	Den
6	Hyperactivity	ASV00351	Family_XI	*Gemella*	Den
6	Hyperactivity	ASV00066	Lachnospiraceae	*Hungatella*	Den
6	Hyperactivity	ASV00513	Prevotellaceae	*Prevotella*	Den
6	Hyperactivity	ASV00005	Ruminococcaceae	*Subdoligranulum*	Den
6	Peer	ASV00085	Lachnospiraceae	*Tyzzerella_4*	Den
6	Peer	ASV00333	Alcaligenaceae	*Sutterella*	Den
6	Peer	ASV00124	Ruminococcaceae	*Ruminiclostridium_5*	Den
6	Peer	ASV00014	Streptococcaceae	*Streptococcus*	Den
6	Peer	ASV00383	Streptococcaceae	*Lactococcus*	Den
6	Peer	ASV00041	Lachnospiraceae	*Lachnoclostridium*	Den
6	Peer	ASV00007	Lachnospiraceae	*Blautia*	Den
6	Peer	ASV00097	Actinomycetaceae	*Actinomyces*	Den
6	Peer	ASV00198	Peptostreptococcaceae	*Peptoclostridium*	Den
6	Peer	ASV00004	Enterobacteriaceae	*Enterobacter*	Den
6	ProSocial	ASV00004	Enterobacteriaceae	*Enterobacter*	Num
6	ProSocial	ASV00445	Prevotellaceae	*Paraprevotella*	Den
6	ProSocial	ASV00002	Bifidobacteriaceae	*Bifidobacterium*	Num
6	Total Difficulties	ASV00333	Alcaligenaceae	*Sutterella*	Den
6	Total Difficulties	ASV00034	Peptostreptococcaceae	*Romboutsia*	Num
6	Total Difficulties	ASV00513	Prevotellaceae	*Prevotella*	Den

Table 4.3. Bacteria extracted for each SDQ subscale outcome using SELBAL method at twelve months of age.

Month	SDQ	ASV	Family	Genus	Numerator/ Denominator
12	Emotion	ASV00273	Eubacteriaceae	*Eubacterium*	Den
12	Emotion	ASV00106	Rikenellaceae	*Alistipes*	Den
12	Emotion	ASV00119	Lachnospiraceae	*Lachnospira*	Den
12	Emotion	ASV00355	Ruminococcaceae	Ruminococcaceae	Den
12	Emotion	ASV00005	Ruminococcaceae	*Subdoligranulum*	Den
12	Emotion	ASV00029	Veillonellaceae	*Dialister*	Den
12	Emotion	ASV00450	Lachnospiraceae	*Tyzzerella*	Den
12	Emotion	ASV00378	Porphyromonadaceae	*Odoribacter*	Den
12	Emotion	ASV00233	Lachnospiraceae	Lachnospiraceae	Den
12	Emotion	ASV00470	Erysipelotrichaceae	*Catenibacterium*	Den
12	Emotion	ASV00086	Ruminococcaceae	*Flavonifractor*	Den
12	Emotion	ASV00058	Ruminococcaceae	*Anaerofilum*	Num
12	Conduct	ASV00273	Eubacteriaceae	*Eubacterium*	Num
12	Conduct	ASV00175	Alcaligenaceae	*Parasutterella*	Den
12	Conduct	ASV00178	Staphylococcaceae	*Staphylococcus*	Num
12	Conduct	ASV00399	Lachnospiraceae	Lachnospiraceae70	Den
12	Conduct	ASV00074	Ruminococcaceae	Ruminococcaceae	Den
12	Hyperactivity	ASV00056	Erysipelotrichaceae	Erysipelotrichaceae	Den
12	Hyperactivity	ASV00105	Lachnospiraceae	*Coprococcus_2*	Num
12	Peer	ASV00170	Erysipelotrichaceae	*Turicibacter*	Num
12	Peer	ASV00054	Prevotellaceae	*Prevotella_9*	Den
12	Peer	ASV00453	Lachnospiraceae	*Tyzzerella*	Den
12	Peer	ASV00085	Lachnospiraceae	*Tyzzerella_4*	Den
12	ProSocial	ASV00124	Ruminococcaceae	*Ruminiclostridium_5*	Den
12	ProSocial	ASV00043	Clostridiaceae_1	*Clostridium_sensu_stricto_1*	Den
12	ProSocial	ASV00034	Peptostreptococcaceae	*Romboutsia*	Den
12	Total Problem	ASV00273	Eubacteriaceae	*Eubacterium*	Num
12	Total Problem	ASV00178	Staphylococcaceae	*Staphylococcus*	Num
12	Total Problem	ASV00378	Porphyromonadaceae	*Odoribacter*	Den

Table 4.4. Mean average ASV count for each bacteria extracted using SELBAL method.

Month	ASV	Genus	Mean	Standard Deviation
1	ASV00002	*Bifidobacterium*	2494.34	2852.40
1	ASV00005	*Subdoligranulum*	4.16	66.56
1	ASV00007	*Bacteroides*	8.13	49.36
1	ASV00011	*Veillonella*	270.47	882.98
1	ASV00014	*Streptococcus*	156.97	426.13
1	ASV00022	*Erysipelatoclostridium*	48.63	432.71
1	ASV00028	*Enterococcus*	29.99	134.05
1	ASV00041	*Lachnoclostridium*	25.20	192.04
1	ASV00047	*Megasphaera*	9.96	133.37
1	ASV00051	*Intestinibacter*	3.17	17.67
1	ASV00060	*Citrobacter*	33.33	236.70
1	ASV00066	*Hungatella*	8.04	51.02
1	ASV00072	*Lactobacillus*	72.60	304.50
1	ASV00086	*Flavonifractor*	5.20	57.55
1	ASV00089	*Collinsella*	18.21	65.47
1	ASV00178	*Staphylococcus*	12.07	35.46
1	ASV00198	*Peptoclostridium*	0.84	11.19
1	ASV00259	*Rothia*	13.65	51.06
1	ASV00273	*Eubacterium*	1.93	16.24
1	ASV00419	*Salmonella*	2.69	32.87
1	ASV00461	*Peptoniphilus*	0.63	3.83
1	ASV00472	*Varibaculum*	0.65	4.61
1	ASV00738	*Scardovia*	0.65	3.35
1	ASV00768	*Pseudocitrobacter*	3.99	8.00
6	Asv00002	*Bifidobacterium*	1239.39	1197.70
6	Asv00003	*Akkermansia*	269.14	878.76
6	Asv00004	*Enterobacter*	74.95	333.75
6	Asv00005	*Subdoligranulum*	13.68	120.18
6	Asv00007	*Blautia*	189.02	496.03
6	Asv00014	*Streptococcus*	151.42	270.41
6	Asv00020	*Roseburia*	3.90	24.60
6	Asv00030	*Fusicatenibacter*	16.16	60.77
6	Asv00034	*Romboutsia*	4.43	27.45
6	Asv00040	*Dorea*	123.57	376.38
6	Asv00041	*Lachnoclostridium*	132.58	376.38
6	Asv00054	*Prevotella_9*	34.04	270.91
6	Asv00066	*Hungatella*	30.85	125.27

Table 4.4 Continued

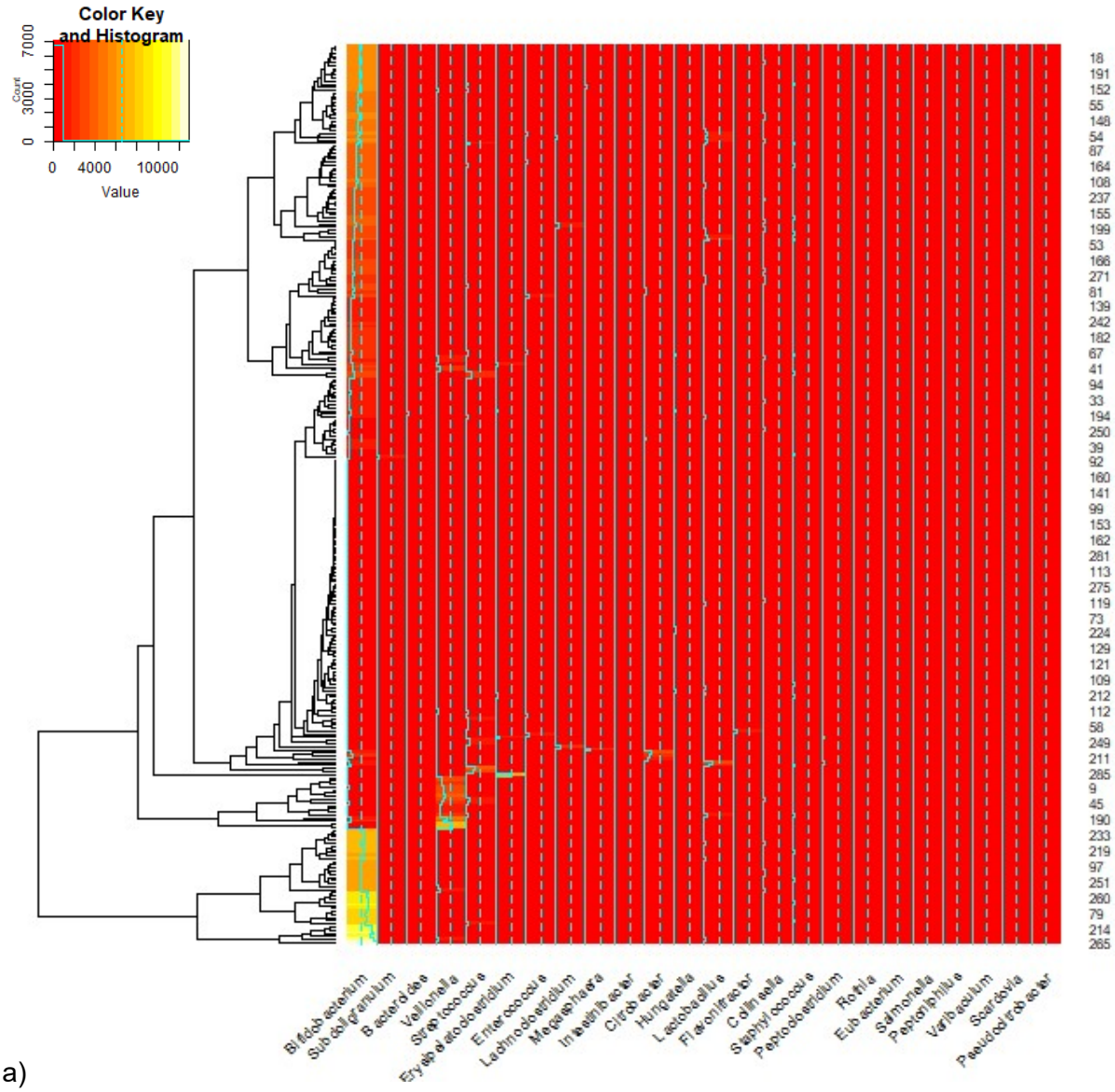
Month	ASV	Genus	Mean	Standard Deviation
6	Asv00074	Ruminococcaceae_UCG-013	3.56	25.71
6	Asv00077	Ruminococcaceae_UCG-014	1.73	27.06
6	Asv00085	*Tyzzerella_4*	16.63	73.30
6	Asv00086	*Flavonifractor*	22.05	64.36
6	Asv00097	*Actinomyces*	8.75	16.39
6	Asv00106	*Alistipes*	26.50	171.35
6	Asv00119	*Lachnospira*	7.24	41.61
6	Asv00124	*Ruminiclostridium_5*	3.55	9.33
6	Asv00198	*Peptoclostridium*	9.78	27.29
6	Asv00224	*Holdemanella*	2.56	40.33
6	Asv00263	*Sellimonas*	1.62	5.99
6	Asv00314	*Eggerthella*	2.69	5.60
6	Asv00328	*Lachnoclostridium_5*	4.65	61.33
6	Asv00333	*Sutterella*	1.95	33.73
6	Asv00351	*Gemella*	1.50	4.12
6	Asv00383	*Lactococcus*	0.87	4.30
6	Asv00384	*Granulicatella*	0.45	7.74
6	Asv00437	*Desulfovibrio*	0.64	8.41
6	Asv00445	*Paraprevotella*	1.22	15.04
6	Asv00505	*Raoultella*	2.24	22.00
6	Asv00513	*Prevotella*	1.52	6.38
6	Asv00624	*Fusobacterium*	2.03	23.57
12	Asv00005	*Subdoligranulum*	185.10	428.62
12	Asv00026	*Anaerostipes*	136.73	212.75
12	Asv00029	*Dialister*	92.45	248.47
12	Asv00034	*Romboutsia*	67.93	149.06
12	Asv00043	*Clostridium_sensu_stricto_1*	2.03	4.25
12	Asv00054	*Prevotella_9*	91.64	409.50
12	Asv00056	Erysipelotrichaceae_UCG-003	77.46	285.64
12	Asv00058	*Anaerofilum*	15.36	38.45
12	Asv00069	*Megamonas*	36.68	178.10
12	Asv00074	Ruminococcaceae_UCG-013	24.84	56.77
12	Asv00085	*Tyzzerella_4*	33.50	71.48
12	Asv00086	*Flavonifractor*	14.43	25.15
12	Asv00105	*Coprococcus_2*	4.20	21.53
12	Asv00106	*Alistipes*	19.64	55.03
12	Asv00119	*Lachnospira*	31.61	82.90
12	Asv00122	*Ruminococcus_1*	10.40	29.53
12	Asv00124	*Ruminiclostridium_5*	18.78	103.00

Table 4.4 Continued

Month	ASV	Genus	Mean	Standard Deviation
12	Asv00170	*Turicibacter*	7.98	22.91
12	Asv00175	*Parasutterella*	9.14	34.87
12	Asv00178	*Staphylococcus*	1.52	20.71
12	Asv00233	Lachnospiraceae_FCS020_group	2.79	12.44
12	Asv00273	*Eubacterium*	1.09	3.72
12	Asv00355	Ruminococcaceae_UCG-004	2.04	6.31
12	Asv00378	*Odoribacter*	1.67	7.69
12	Asv00399	Lachnospiraceae_UCG-004	3.39	8.27
12	Asv00437	*Desulfovibrio*	0.85	8.88
12	Asv00450	*Tyzzerella_3*	2.01	10.36
12	Asv00453	*Tyzzerella*	1.49	8.29
12	Asv00470	*Catenibacterium*	2.74	28.09

Table 4.5. Mean average Balance Parameter for each of the SDQ outcomes included in the SELBAL extraction at months 1, 6, and 12.

Month	SDQ Scale	Mean	Standard Deviation
1	Emotion	1.57	1.75
1	Conduct	-0.07	0.83
1	Hyperactivity/inattention	1.44	1.83
1	Peer	-0.001	0.92
1	Pro Social	1.93	1.66
1	Total Difficulties	2.04	1.63
6	Emotion	0.04	1.06
6	Conduct	0.33	0.81
6	Hyperactivity/inattention	0.23	1.32
6	Peer	1.18	1.00
6	Pro Social	3.12	1.26
6	Total Difficulties	0.66	1.25
12	Emotion	1.80	1.32
12	Conduct	0.26	0.94
12	Hyperactivity/inattention	-1.08	1.70
12	Peer	0.79	1.61
12	Pro Social	-0.29	1.74
12	Total Difficulties	0.29	0.97



a)

Figure 4.1a. Heatmaps representing the ASV counts of each of the bacteria including extracted using the SELBAL method at 1-month of age.

b)

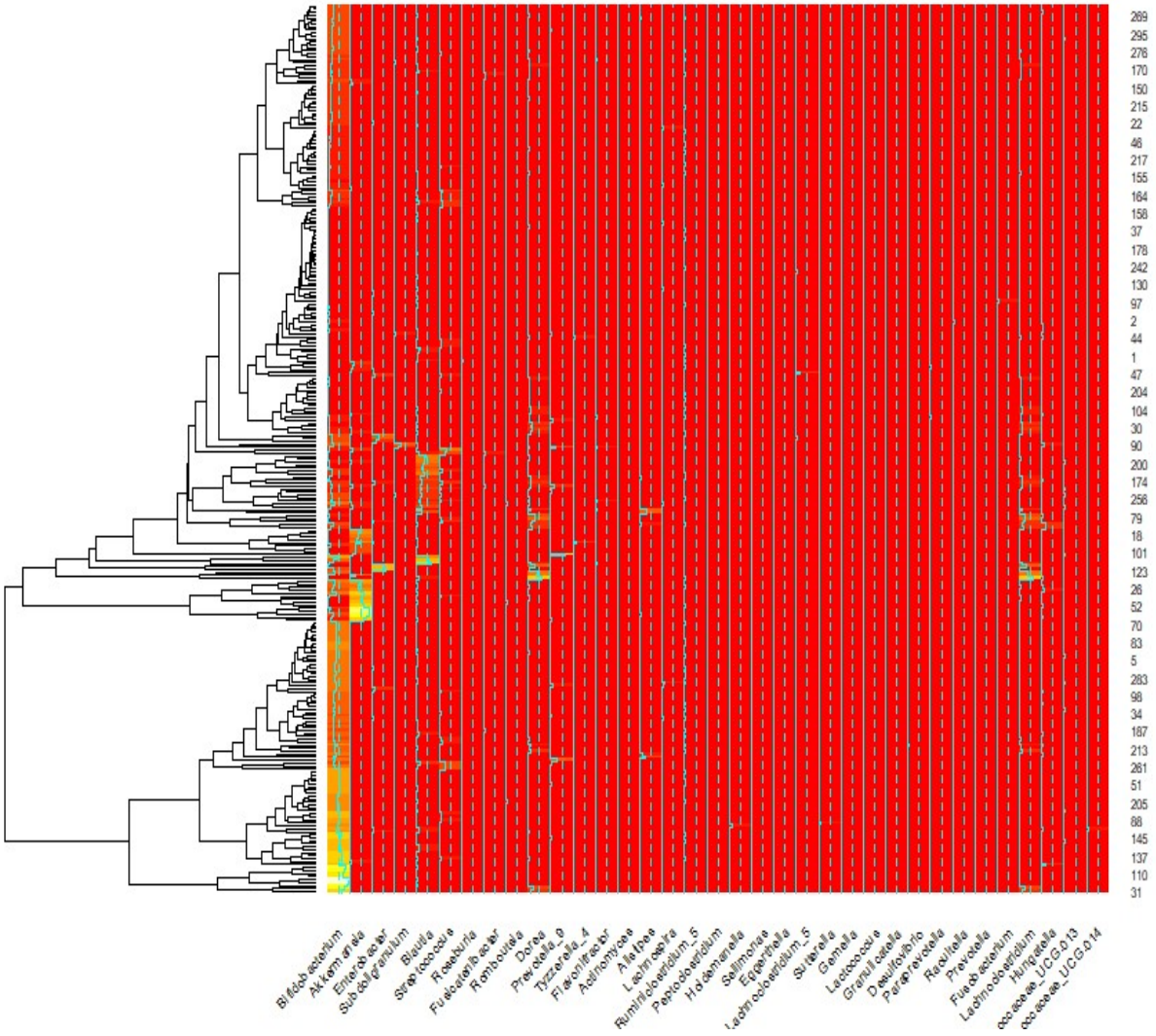
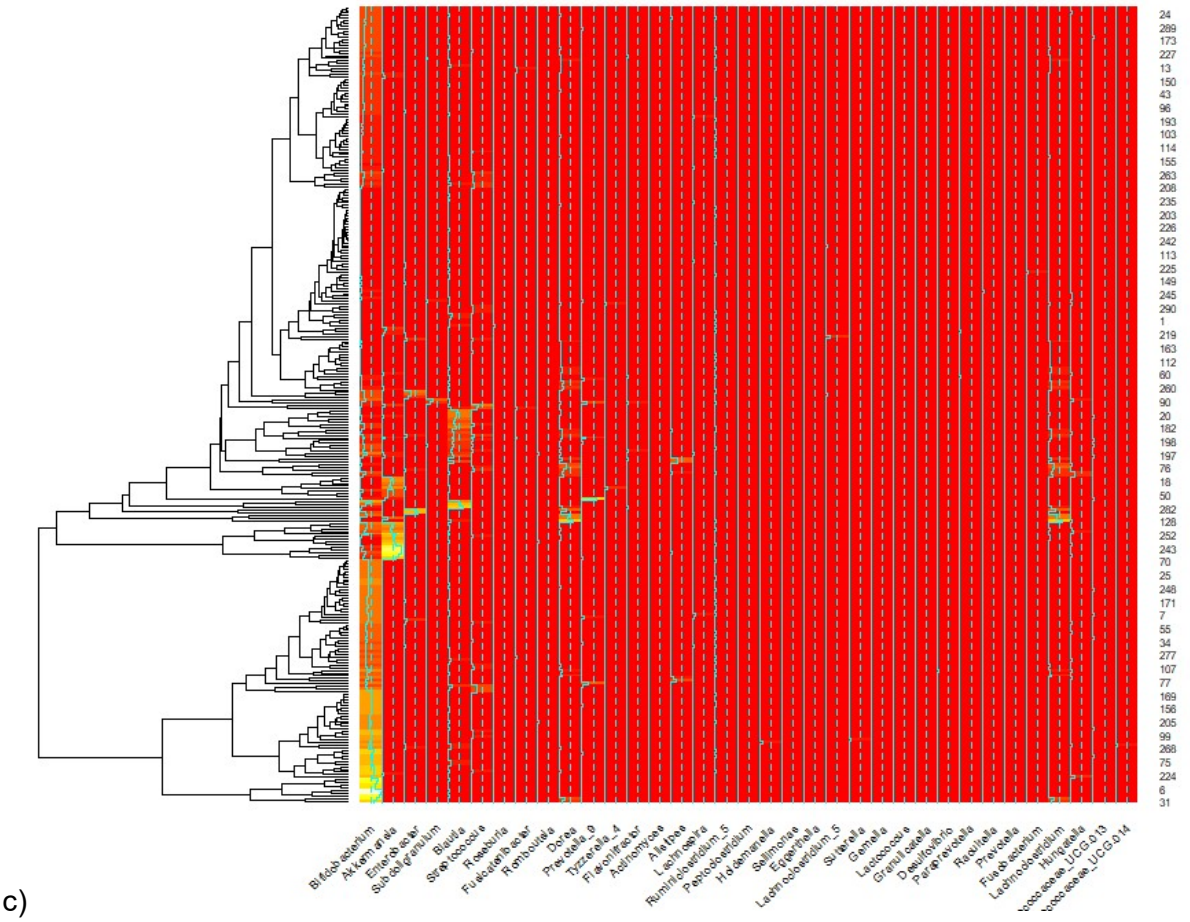


Figure 4.1b. Heatmaps representing the ASV counts of each of the bacteria including extracted using the SELBAL method at 6-months of age.



c)

Figure 4.1c. Heatmaps representing the ASV counts of each of the bacteria including extracted using the SELBAL method at 12-months of age.

were computed for each SDQ variable at both 6- and 12, months and further investigated in this chapter.

4.3.3 Extracted bacteria and SDQ outcomes, hypothesis 1.

To determine whether each of the extracted bacteria were significantly associated with the respective SDQ subscales multiple linear regressions were performed using the bacteria presented in Tables 4.4, 4.5, and 4.6, as predictors.

4.3.3.1 One month of age

At one month of age, greater relative abundance of *Intestinibacter* was significantly associated with increased conduct problems (F (12, 160) = 2.581, p < .05, FDR = 0.04, Adj R^2 =9.9%), when adjusted for mode of birth and maternal PSS scores (see Figure 4.2). However, this particular relationship appeared to be driven by a single individual as can be observed in figure 4.2. Therefore, a sensitivity check was performed by removing this outlier. Following this check the relationship was no longer significant. A second significant relationship was found at 1-month of age; hyperactivity/inattention

was positively associated with greater relative abundance of *Veillonella* (*F* (12, 160) = 2.752, *p* < .01, FDR = 0.04, Adj R^2 = 10.1%). Both conduct problems and hyperactivity/inattention subscales are indicative of increased risk of externalising problems measured on the SDQ. For this time point there were no significant relationships between identified bacteria, total difficulties, or the subscales of emotional problems, peer problems or prosocial behaviour. Results presented in table 4.4 present the p values of the entire model, with FDR corrections, where covariates were also included, only models with significant influence of the bacteria of interest have been reported here.

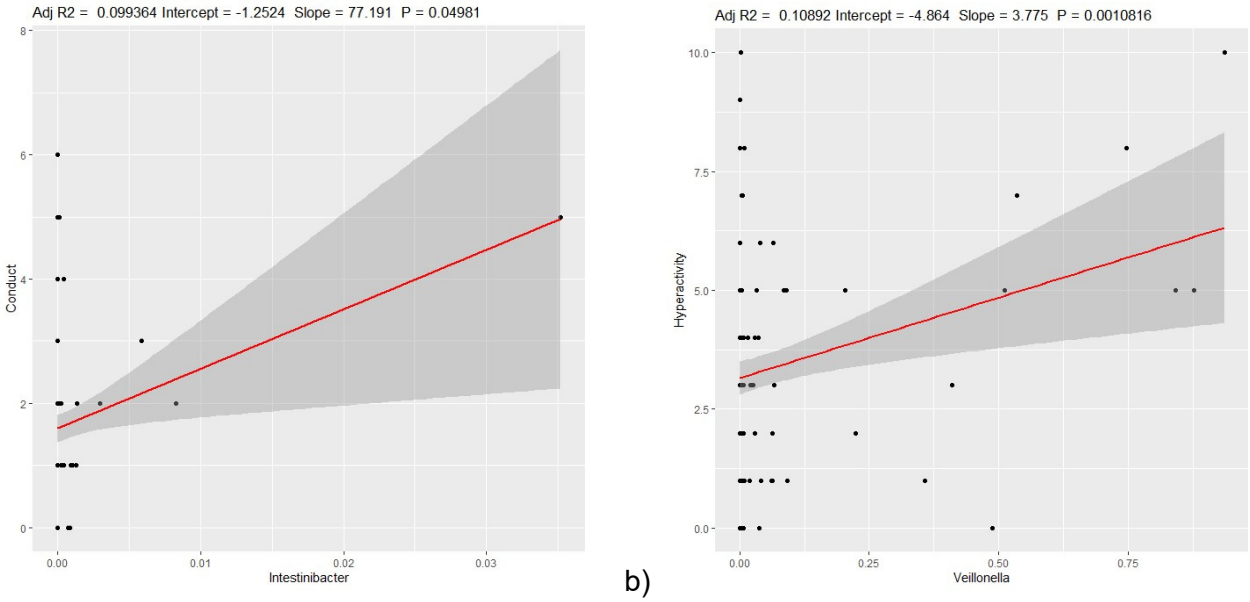


Figure 4.2. Graphs to show the linear regression between relative abundance of a) *Intestinibacter* and conduct problems at b) *Veillonella* and hyperactivity/inattention at 1-month.

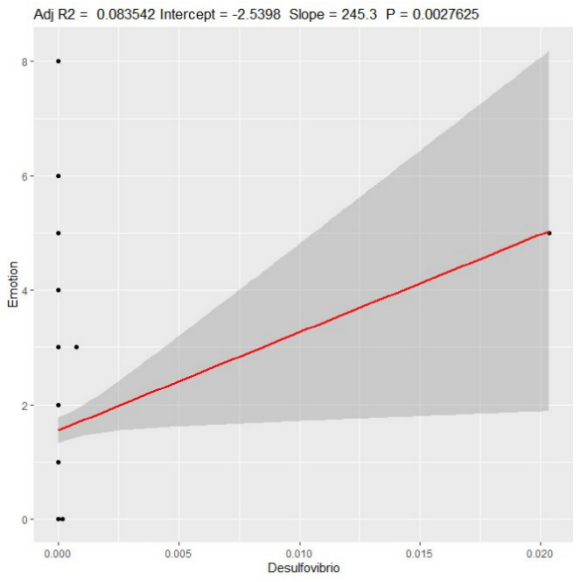
4.3.3.2 Six months of age

At 6-months of age there were several bacteria that presented with significant associations with SDQ subscales measured at 4-years of age. The subscale measuring emotional problems was positively associated with increased relative abundance of both *Desulfovibrio* (*F* (12, 185) = 2.497, *p* < .01, FDR = 0.04, Adj R^2 = 8.4%) (see Figure 4.3), which remains significant after being adjusted for birthweight and number of siblings, and *Tyzzerella* subgroup *4* (*Tyzerella_4)* (*F* (1, 248) = 20.100, *p* < .01, FDR = 0.001, Adj R^2 = 7.1%), in the unadjusted model as confounding was not present. Increased relative abundance of *Tyzzerella_4* was also significantly associated with peer problems (*F* (1, 248) = 9.802, *p* < .01, FDR = 0.04, Adj R^2 =3.4%), which together with emotional subscale scores provides an overall score for risk of internalising problems. Therefore, *Tyzzerell_4* should be viewed as a bacterium of interest in GM measured at 6-months of age, in terms of biological predictors of internalising problems. Additionally, neither model showed evidence of confounding, and therefore the variance in SDQ scores was solely explained by the relative abundance of *Tyzzerella_4.* Increased scores on the conduct problems subscale were associated with increased relative abundance of *Dorea* (*F* (12, 185)

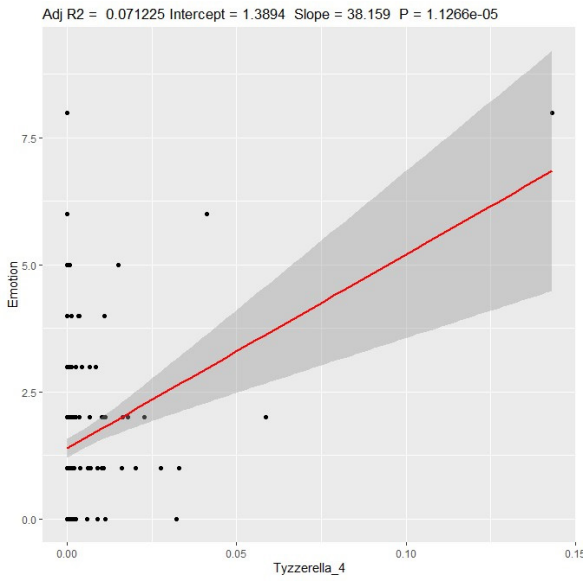
= 3.772, $p < .01$, FDR = 0.002, Adj R^2 =14.5%), adjusted for maternal PSS score. Conduct problems were also associated with lower relative abundance of *Gemella* (F (12, 185) = 2.609, $p < .05$, FDR = 0.04, Adj R^2 =8.9%), when adjusted by mode of birth. Interestingly mode of birth appears to be a significant confounder for the relationship between *Gemella* measured at both 1-, and 6- months of age, and conduct problems measured at 4-years of age. In addition to being significantly associated with increased conduct problems, lower relative abundance of *Gemella* was significantly associated with increased hyperactivity/inattention scores (F (1, 248) = 5.776, $p < .05$, FDR = 0.06, Adj R^2 =1.9%), indicating an importance of this bacteria at 6-months in predicting risk of externalising problems at 4-years. Finally, both the hyperactivity/inattention subscale (F (1, 248) = 5.223, $p < .05$, FDR = 0.08, Adj R^2 =1.7%), and total difficulties score (F (1, 248) = 3.921, $p < .05$, FDR = 0.13, Adj R^2 = 1.2%), were negatively associated with increased relative abundance of *Prevotella* at 6-months, in the unadjusted model. However, these were both non-significant when adjusted for multiple comparisons using the Benjamini-hochberg false discovery rate correction. Several of these relationships appear to have been driven by a small number of individuals, as can be seen in Figure 4.3. Therefore, sensitivity checks were performed to remove outliers for each of the bacteria. Following these sensitivity checks, *Desulfovibrio* was no longer significantly associated with emotional problems, and *Prevotella* was no longer associated with either hyperactivity/inattention or Total Difficulties. *Tyzerella_4*, however, presented with a significantly positive association with peer problems (F (1, 248) = 7.912, $p <.05$, Adj R^2 = 27.0%; See Figure 4.4), although this was no longer significantly associated with emotional problems.

4.3.3.3 Twelve months of age

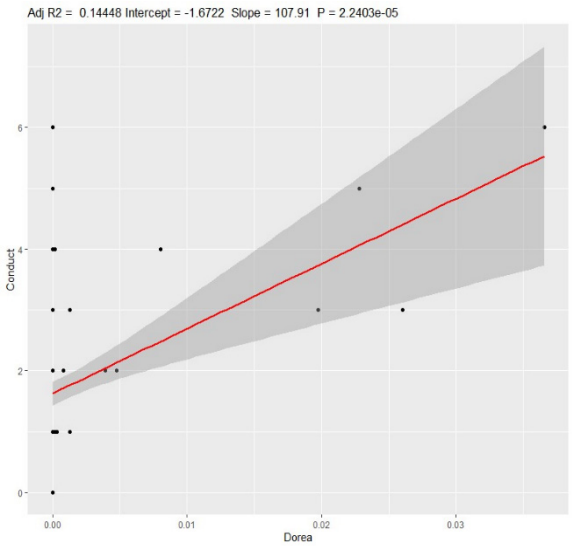
At 12-months of age results indicated that there was a significant association between increased relative abundance of both *Lachnospiraceae* (F (12, 183) = 2.133, $p < .05$, FDR = 0.07, Adj R^2 =6.5%), and *Catenibacterium* (F (12, 183) = 2.758, $p < .01$, FDR = 0.04, Adj R^2 =9.8%) with emotional problems (see Figure 4.5). Both models were adjusted for birthweight and number of siblings, *Catenibacterium* was additionally adjusted by SEIFA scores. Increased relative abundance of *Staphylococcus* associated with both increased conduct problems (F (1, 214) = 9.651, $p < .01$, FDR = 0.04, Adj R^2 =3.9%). Following sensitivity checks none of these relationships remained significant. There were, however, significant positive associations between Total difficulties measured at 4 years of age and both *Eubacterium* (F (12, 183) = 2.280, $p < .05$, Adj R^2 = 7.3%), and *Staphylococcus* (F (12, 183) = 2.197 $p < .05$, Adj R^2 = 6.9%), and a significant negative associations with *Odoribacter* (F (12, 183) = 1.904, $p < .05$, Adj R^2 = 5.3%) that remained significant following sensitivity checking (See Figure 4.6).


Adj R2 = 0.083542 Intercept = -2.5398 Slope = 245.3 P = 0.0027625
Emotion
Desulfovibrio

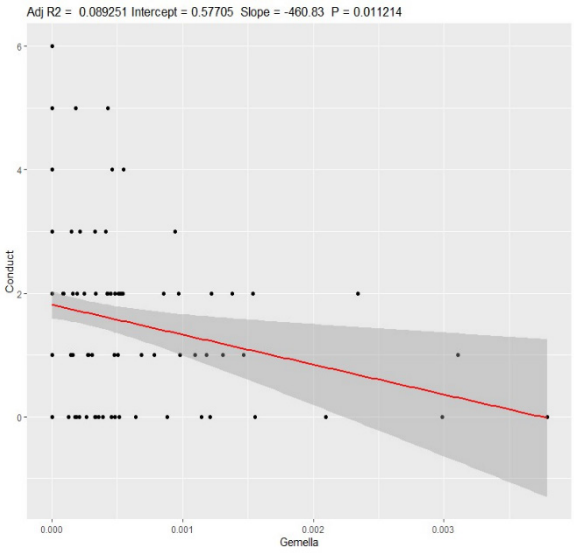
a)


Adj R2 = 0.071225 Intercept = 1.3894 Slope = 38.159 P = 1.1266e-05
Emotion
Tyzzerella_4

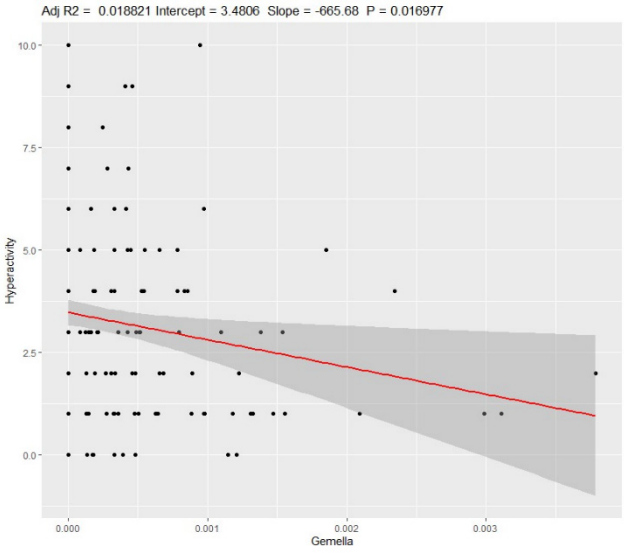
b)


Adj R2 = 0.14448 Intercept = -1.6722 Slope = 107.91 P = 2.2403e-05
Conduct
Dorea

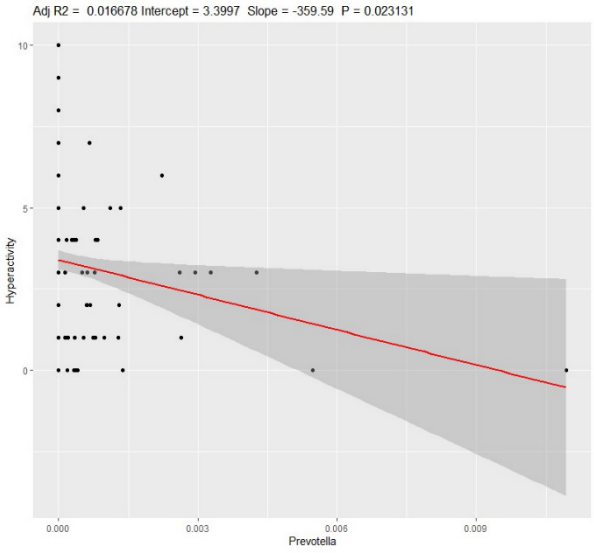
c)


Adj R2 = 0.089251 Intercept = 0.57705 Slope = -460.83 P = 0.011214
Conduct
Gemella

d)


Adj R2 = 0.018821 Intercept = 3.4806 Slope = -665.68 P = 0.016977
Hyperactivity
Gemella

e)


Adj R2 = 0.016678 Intercept = 3.3997 Slope = -359.59 P = 0.023131
Hyperactivity
Prevotella

f)

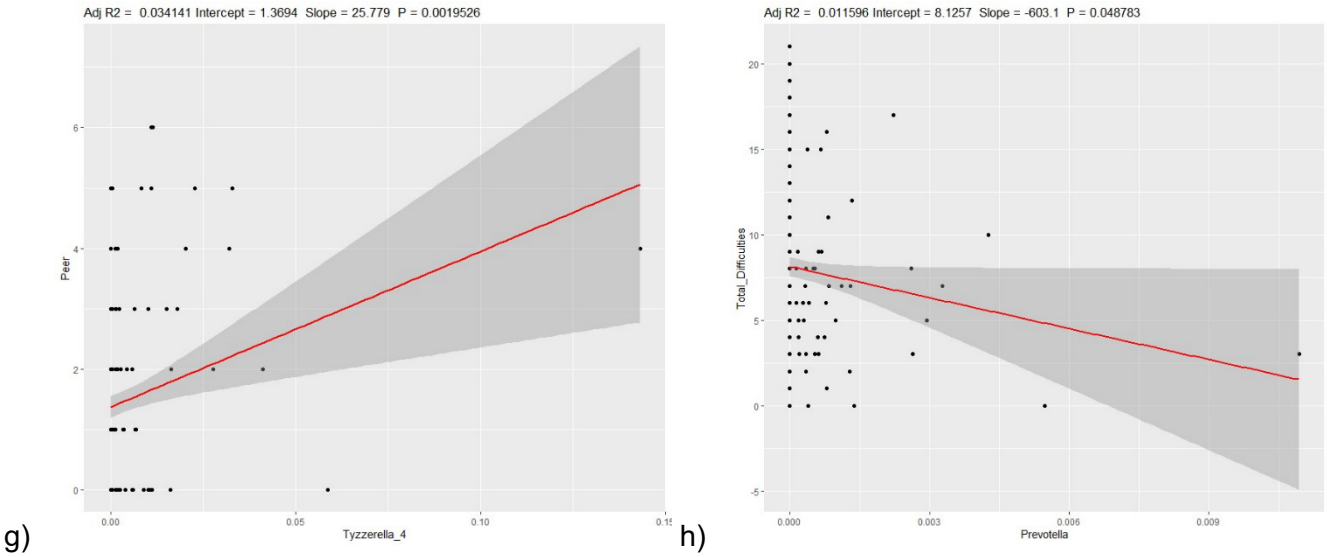


Figure 4.3. Graphs to show the linear regression between relative abundance of a) *Desulfovibrio* and emotional problems b) *Tyzzerella* and emotional problems, c) *Dorea* and conduct problems, d) *Gemella* and conduct problems, e) *Gemella* and hyperactivity/inattention, f) *Prevotella* and hyperactivity/inattention, g) *Tyzzerella* and peer problems, h) *Prevotella* and total difficulties at 6-months.

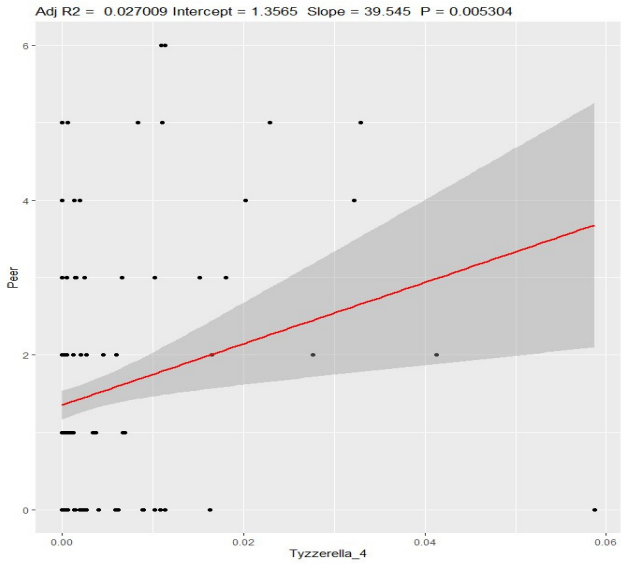


Figure 4.4. Graph to show the linear regression between relative abundance of *Tyzzerella_4* at 6 months and Peer problems at 4 years of age following sensitivity check.

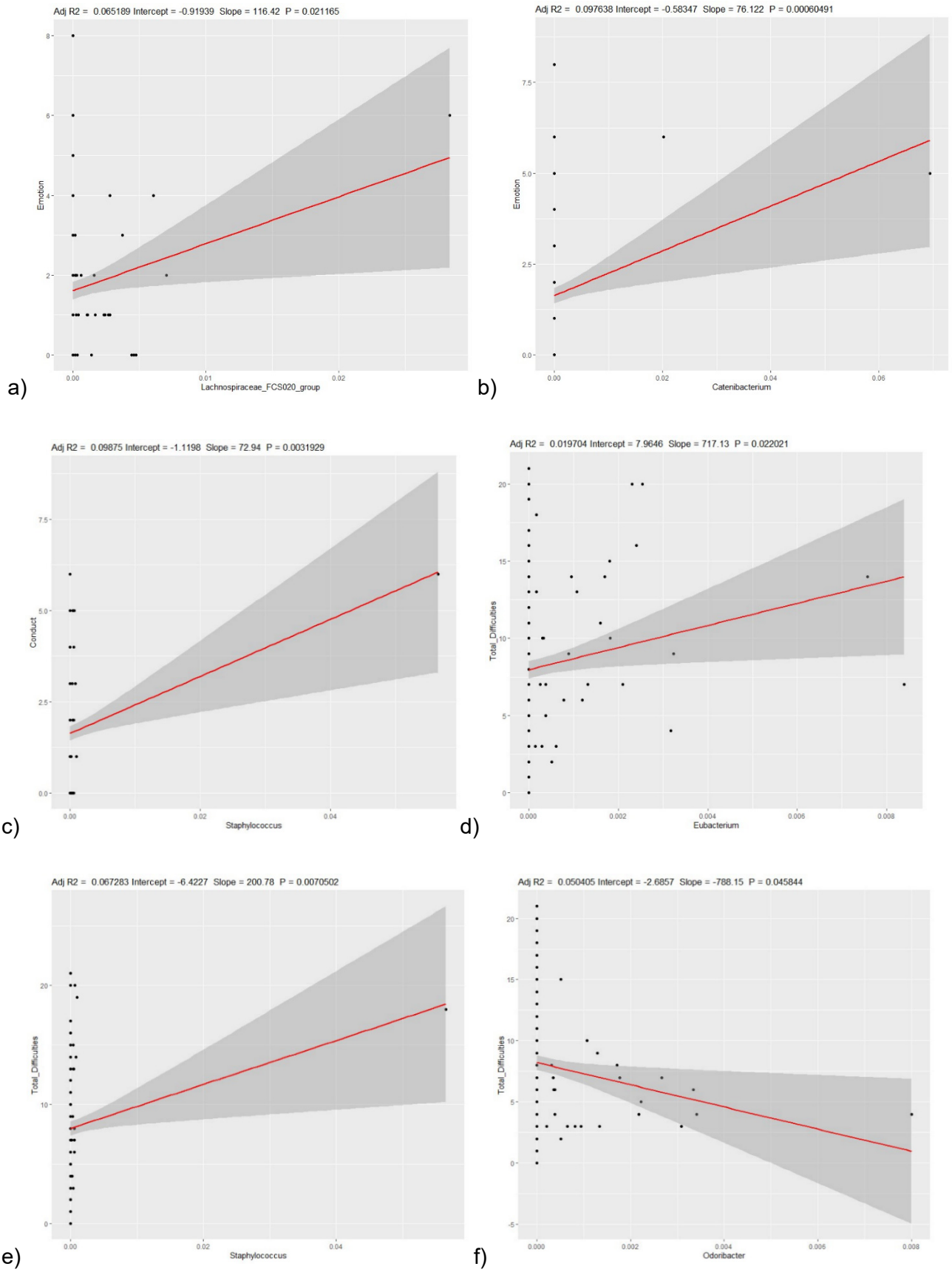


Figure 4.5. Graphs to show the linear regression between relative abundance of a) *Lachnospiraceae* and emotional problems at 12-months, b) *Catenibacterium* and emotional problems, c) *Staphylococcus* and conduct problems, d) *Eubacterium* and total difficulties, e) *Staphylococcus* and total difficulties, and f) *Odoribacter* and total difficulties at 12-months.

Table 4.6. Associations between bacteria extracted at one month of age and SDQ outcomes measured at 4-years.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
1	*Staphylococcus*	Emotion	2.476	20.375	12	160	0.005	0.093	0.043 [c d]
1	*Rothia*	Emotion	2.308	-2.407	12	160	0.010	0.084	0.052 [c d]
1	*Blautia*	Emotion	2.355	-11.220	12	160	0.008	0.086	0.052 [c d]
1	*Salmonella*	Emotion	2.338	-57.110	12	160	0.009	0.085	0.052 [c d]
1	*Lactobacillus*	Emotion	2.501	-2.566	12	160	0.005	0.095	0.043 [c d]
1	*Collinsella*	Emotion	2.454	16.829	12	160	0.058	0.092	0.131 [c d]
1	*Escherichia/Shigella*	Emotion	2.312	-0.117	12	160	0.010	0.084	0.052 [c d]
1	*Pseudocitrobacter*	Emotion	2.412	112.300	12	160	0.067	0.090	0.135 [c d]
1	*Eubacterium*	Emotion	2.425	-39.770	12	160	0.006	0.090	0.047 [c d]
1	*Erysipelatoclostridium*	Emotion	2.630	3.090	12	160	0.003	0.102	0.040 [c d]
1	*Flavonifractor*	Conduct	2.274	-16.850	12	160	0.011	0.082	0.055 [c d]
1	*Intestinibacter*	Conduct	2.581	77.191	12	160	0.004	0.099	0.042 [d h]
1	*Pseudocitrobacter*	Conduct	2.206	-21.690	12	160	0.014	0.078	0.061 [d h]
1	*Enterococcus*	Hyperactivity	1.758	5.618	12	160	0.059	0.050	0.132
1	*Streptococcus*	Hyperactivity	1.823	3.431	12	160	0.048	0.054	0.128
1	*Clostridium*	Hyperactivity	2.019	-3.785	12	160	0.026	0.066	0.082
1	*Staphylococcus*	Hyperactivity	1.711	1.945	12	160	0.069	0.047	0.135
1	*Rothia*	Hyperactivity	1.715	3.188	12	160	0.068	0.048	0.135
1	*Veillonella*	Hyperactivity	2.752	3.775	12	160	0.002	0.109	0.037
1	*Hungatella*	Peer	0.540	-3.067	12	160	0.886	-0.003	0.942
1	*Staphylococcus*	Peer	0.527	2.412	12	160	0.895	-0.034	0.942
1	*Citrobacter*	Peer	0.587	2.313	12	160	0.850	-0.030	0.942
1	*Pseudocitrobacter*	Peer	0.655	133.900	12	160	0.793	-0.025	0.942
1	*Eubacterium*	Peer	2.281	50.117	1	238	0.132	0.005	0.204
1	*Pluralibacter*	Peer	0.570	261.700	12	160	0.864	-0.031	0.942
1	*Subdoligranulum*	Peer	0.631	-6.320	12	160	0.813	-0.026	0.942
1	*Peptoniphilus*	Peer	0.526	22.730	12	160	0.896	-0.034	0.942
1	*Epulopiscium*	Peer	0.610	-25.380	12	160	0.832	-0.028	0.942
1	*Megasphaera*	Peer	0.053	-1.430	12	160	0.890	-0.034	0.942
1	*Bifidobacterium*	Prosocial	0.511	0.092	12	160	0.905	-0.035	0.942
1	*Citrobacter*	Prosocial	0.527	-1.625	12	160	0.895	-0.003	0.942
1	*Scardovia*	Prosocial	0.948	875.200	12	160	0.503	-0.004	0.666
1	*Escherichia/Shigella*	Total Problem	2.127	0.347	12	160	0.018	0.073	0.068
1	*Flavonifractor*	Total Problem	2.213	-57.920	12	160	0.013	0.078	0.061 [d]
1	*Streptococcus*	Total Problem	2.221	6.145	12	160	0.013	0.079	0.061

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.7. Associations between bacteria extracted at six months of age and SDQ outcomes measured at 4-years.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
6	*Eggerthella*	Emotion	1.679	75.800	12	185	0.074	0.040	0.136 [c]
6	*Akkermansia*	Emotion	1.653	0.194	12	185	0.081	0.038	0.138 [c]
6	*Fusobacterium*	Emotion	1.782	-177.200	12	185	0.054	0.045	0.128 [c]
6	*Prevotella_9*	Emotion	1.691	-2.973	12	185	0.072	0.040	0.136 [c]
6	Ruminococcaceae_UCG.014	Emotion	1.741	-20.310	12	185	0.061	0.043	0.418 [c]
6	*Desulfovibrio*	Emotion	2.497	245.300	12	185	0.005	0.084	0.043 [c e]
6	*Roseburia*	Emotion	1.665	8.671	12	185	0.078	0.039	0.138 [c]
6	*Fusicatenibacter*	Emotion	1.655	-2.170	12	185	0.080	0.038	0.138 [c]
6	*Tyzzerella_4*	Emotion	20.100	38.159	1	248	0.000	0.071	0.001
6	*Lachnospira*	Emotion	1.782	-15.480	12	185	0.054	0.045	0.128 [c]
6	*Dorea*	Conduct	3.772	107.900	12	185	0.000	0.145	0.002 [d]
6	*Actinomyces*	Conduct	2.039	19.746	12	185	0.023	0.060	0.076 [h]
6	*Gemella*	Conduct	2.609	-460.800	12	185	0.003	0.089	0.040 [h]
6	*Romboutsia*	Hyperactivity	1.679	21.556	12	185	0.074	0.040	0.136
6	*Granulicatella*	Hyperactivity	1.301	97.620	12	185	0.221	0.018	0.322
6	*Gemella*	Hyperactivity	5.776	-665.678	1	248	0.017	0.019	0.067
6	*Hungatella*	Hyperactivity	1.287	0.667	12	185	0.230	0.017	0.330
6	*Prevotella*	Hyperactivity	5.223	-359.591	1	248	0.023	0.017	0.076
6	*Subdoligranulum*	Hyperactivity	1.597	-13.970	12	185	0.096	0.035	0.152
6	*Tyzzerella_4*	Peer	9.802	25.779	1	248	0.002	0.034	0.037
6	*Sutterella*	Peer	0.517	-19.520	12	185	0.902	-0.030	0.942
6	*Ruminiclostridium_5*	Peer	0.241	-26.224	1	248	0.624	-0.003	0.805
6	*Streptococcus*	Peer	0.413	1.152	12	185	0.957	-0.037	0.966
6	*Lactococcus*	Peer	0.953	-125.079	1	248	0.330	0.000	0.461
6	*Lachnoclostridium*	Peer	0.418	1.130	12	185	0.955	-0.037	0.966
6	*Blautia*	Peer	0.554	1.802	12	185	0.876	-0.028	0.942
6	*Actinomyces*	Peer	0.417	-16.610	12	185	0.956	-0.037	0.966
6	*Peptoclostridium*	Peer	0.873	18.351	1	248	0.351	-0.001	0.484
6	*Enterobacter*	Peer	0.389	0.067	12	185	0.966	-0.039	0.966
6	*Enterobacter*	Prosocial	1.304	8.070	12	185	0.219	0.018	0.322
6	*Paraprevotella*	Prosocial	1.004	-90.290	12	185	0.447	0.000	0.604
6	*Bifidobacterium*	Prosocial	1.001	0.594	12	185	0.450	0.000	0.604
6	*Sutterella*	Total Problem	1.603	-48.450	12	185	0.094	0.035	0.152
6	*Romboutsia*	Total Problem	1.730	30.857	12	185	0.063	0.043	0.135
6	*Prevotella*	Total Problem	3.921	-603.099	1	248	0.049	0.012	0.128
6	*Enterobacter*	Total Problem	1.556	-6.912	12	185	0.108	0.033	0.169

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.8. Associations between bacteria extracted at twelve months of age and SDQ outcomes measured at 4-years.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
12	*Eubacterium*	Emotion	0.341	68.921	1	214	0.560	-0.003	0.732
12	*Alistipes*	Emotion	1.646	5.700	12	183	0.082	0.038	0.138 [c e]
12	*Lachnospira*	Emotion	1.797	-8.513	12	183	0.051	0.047	0.128 [c e]
12	Ruminococcaceae	Emotion	1.773	-107.400	12	183	0.006	0.045	0.043 [c e]
12	*Subdoligranulum*	Emotion	1.764	-1.774	12	183	0.057	0.045	0.131 [c e]
12	*Dialister*	Emotion	1.650	0.970	12	183	0.081	0.038	0.138 [c e]
12	*Tyzzerella_3*	Emotion	0.167	20.247	1	214	0.683	-0.004	0.871
12	*Odoribacter*	Emotion	1.802	-197.900	12	183	0.050	0.047	0.128 [c e]
12	Lachnospiraceae	Emotion	2.133	116.400	12	183	0.017	0.065	0.067 [c e]
12	*Catenibacterium*	Emotion	2.758	76.122	12	183	0.002	0.098	0.037 [c e g]
12	*Flavonifractor*	Emotion	1.714	23.837	12	183	0.067	0.042	0.135 [c e]
12	*Anaerofilum*	Emotion	1.637	-2.681	12	183	0.085	0.038	0.140 [c e]
12	*Ruminococcus_1*	Emotion	1.899	40.542	12	183	0.037	0.052	0.108 [c e]
12	*Megamonas*	Emotion	1.683	-1.616	12	183	0.074	0.040	0.136 [c e]
12	*Eubacterium*	Conduct	2.292	214.400	12	183	0.010	0.074	0.052 [d]
12	*Parasutterella*	Conduct	2.074	-19.340	12	183	0.021	0.062	0.075 [d]
12	*Staphylococcus*	Conduct	9.651	77.186	1	214	0.002	0.039	0.037
12	Lachnospiraceae	Conduct	1.942	-5.298	12	183	0.032	0.055	0.096 [d]
12	Ruminococcaceae	Conduct	1.984	8.018	12	183	0.028	0.057	0.086 [d]
12	Erysipelotrichaceae	Hyperactivity	1.724	0.969	12	183	0.065	0.043	0.135 [d]
12	*Coprococcus_2*	Hyperactivity	1.784	36.154	12	183	0.054	0.046	0.128 [f g]
12	*Turicibacter*	Peer	1.334	75.850	12	183	0.203	0.020	0.304 [f g]
12	*Prevotella_9*	Peer	0.726	-2.440	12	183	0.725	-0.017	0.890
12	*Tyzzerella*	Peer	0.718	-88.110	12	183	0.733	-0.018	0.890
12	*Tyzzerella_4*	Peer	0.535	42.980	12	183	0.890	-0.029	0.942
12	*Ruminiclostridium_5*	ProSocial	1.334	75.850	12	183	0.203	0.020	0.304
12	*Clostridium*	ProSocial	0.726	-2.440	12	183	0.725	-0.017	0.890
12	*Romboutsia*	ProSocial	0.718	-88.110	12	183	0.733	-0.018	0.890
12	*Eubacterium*	Total Problem	5.321	717.126	1	214	0.022	0.020	0.076
12	*Staphylococcus*	Total Problem	2.172	200.800	12	183	0.015	0.067	0.063
12	*Odoribacter*	Total Problem	1.863	-788.100	12	183	0.042	0.050	0.118

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.9. Associations between SELBAL balance parameter at 1-, 6-., and 12- months of age and SDQ outcomes measured at 4-years.

Month	Microbiota Genus	F	β	DF	DF	P value	Adj R^2	FDR[b]
1	Emotion	6.627	0.144	1	238	0.011	0.023	0.023
1	Conduct	2.749	0.293	12	160	0.002	0.109	0.010 [h]
1	Hyperactivity/inattention	1.757	0.071	12	160	0.060	0.050	0.087
1	Peer	0.679	0.165	12	160	0.770	-0.023	0.813
1	Pro Social	0.620	0.097	12	160	0.823	-0.027	0.823
1	Total Difficulties	2.468	0.375	12	160	0.006	0.093	0.018 [d]
6	Emotion	1.898	0.172	12	185	0.037	0.052	0.073
6	Conduct	20.450	0.513	1	248	0.000	0.072	0.000
6	Hyperactivity/inattention	7.260	0.294	1	248	0.008	0.025	0.020
6	Peer	0.681	0.195	12	185	0.768	-0.020	0.813
6	Pro Social	1.752	0.338	12	185	0.059	0.044	0.087
6	Total Difficulties	1.835	0.472	12	185	0.045	0.048	0.078
12	Emotion	1.634	-0.001	12	183	0.085	0.038	0.108 [c e]
12	Conduct	2.983	0.178	1	214	0.086	0.009	0.108
12	Hyperactivity/inattention	2.911	0.334	12	183	0.001	0.105	0.007 [g]
12	Peer	1.573	0.232	12	183	0.103	0.034	0.122
12	Pro Social	8.145	0.197	1	214	0.005	0.032	0.018
12	Total Difficulties	6.612	0.754	1	214	0.011	0.025	0.023

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

4.3.3.4 SELBAL balance parameter at 1-, 6-, and 12-months.

At one month of age there was a significant relationship between the emotional subscale of the SDQ measured at 4 years and the balance parameter derived from the SELBAL process (F (1, 283) =6.627, $p < .05$, Adj R^2 = 2.3%), this relationship was not significant at either 6-, or 12- months of age. The conduct problems subscale measured at 4 years of age was significantly related to the balance parameter at 1 months of age (F (12, 160) = 2.749, $p < .01$, Adj R^2 = 1.0%), which was also significantly related to the balance parameter at 6-months (F (12, 185) =20.450, $p < .01$, Adj R^2 = 7.2%), but not 12-months. Hyperactivity/inattention was significantly related to the balance parameter measured at 6-months (F (1, 248) = 7.260, $p < .01$, Adj R^2 = 2.5%), and 12-months (F (12, 183) = 2.911, $p < .01$, Adj R^2 = 10.5%), but not 1-month. Pro Social Behaviour was significantly associated with the balance parameter at 12-months only (F (12, 183) = 8.145, $p < .01$, Adj R^2 = 1.8%). Finally total difficulties was significantly associated at 1-month (F (12, 160) = 2.468, $p < .01$, Adj R^2 = 9.3%) and 12-months (F (12, 183) = 6.612, $p < .05$, Adj R^2 = 2.5%). All of these significant relationships were positively

associated and remained significant following correction for multiple comparisons. Full results and significant confounders for each relationship can be found in table 4.9.

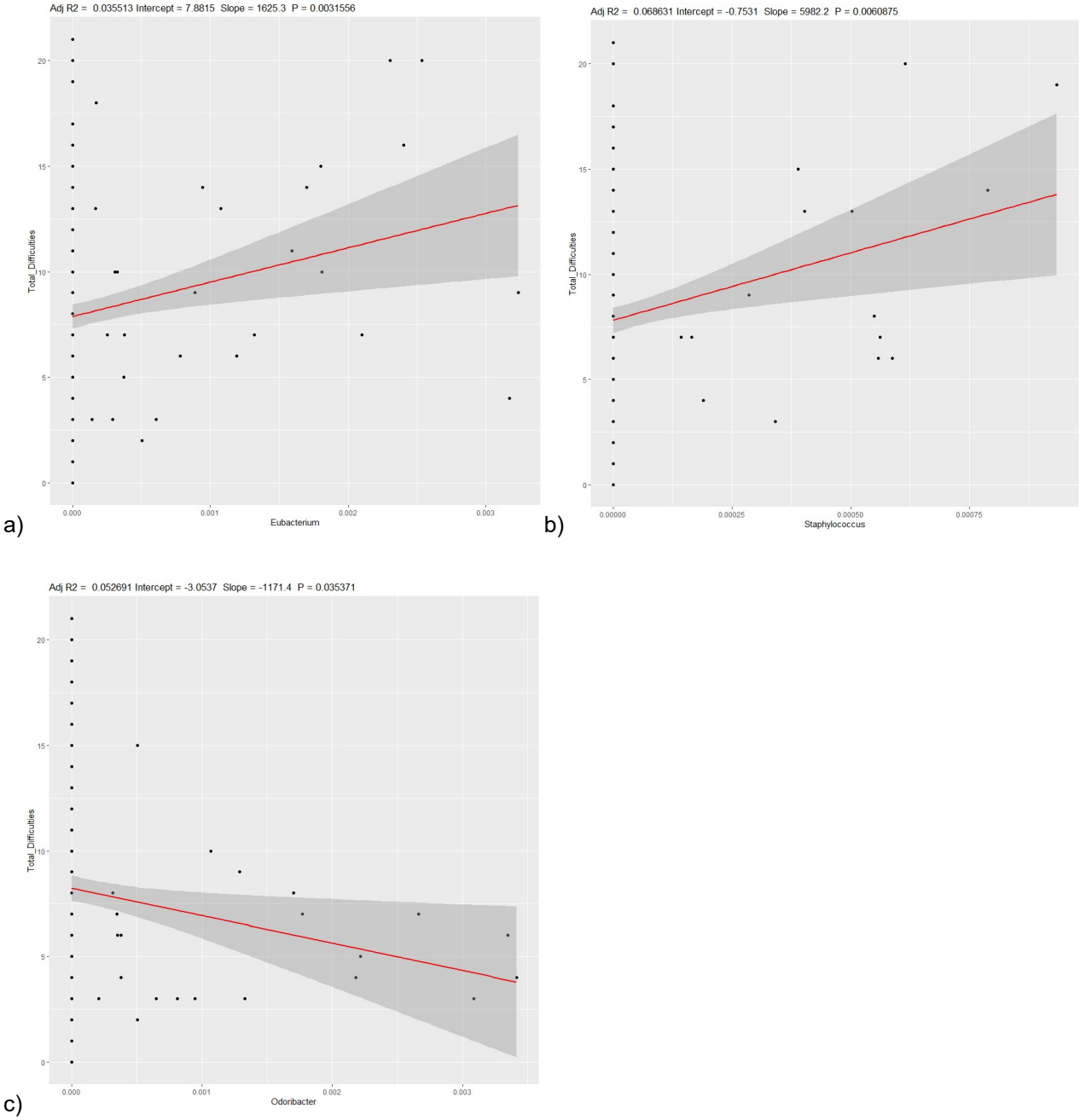


Figure 4.6. Graphs to show the linear regression between relative abundance of a) *Eubacterium* and total difficulties, b) *Staphylococcus* and total difficulties, and c) *Odoribacter* and total difficulties at 12-months, following sensitivity checks.

4.3.4 Moderation of the of relationship between extracted bacteria and SDQ by diet.

In order to establish if the relationship between the GM and SDQ was significantly influenced by diet, the milk-based diet, measured at 1-month of age, measured as exclusively breastfed, mixed feeding or exclusively formula fed, and dietary intake at 6-, and 12-months of age, measured using diet clusters, were added to the analyses using a moderation technique. Dietary clusters were established using PCA analysis and k means clustering as described in chapter 3.

For month 6 the clusters can be described as follows:

Cluster 1 (n=2): Higher than average meat frequency, and slightly higher than average cooked food, lower than average pulses, pasta, yoghurt, fried food, raw foods, eggs, soy, and both sesame products and seeds.

Cluster 2 (n=159): Lower than average fish frequency, soy products and sesame products. Slightly higher than average raw food consumption.

Cluster 3 (n=75): Lower than average raw foods, and cooked foods. Average scores on all other food types.

Cluster 4 (n=23): Lower than average fruit and vegetable consumption. Average scores on all other food types.

Cluster 5 (n=21): Lower than average fish frequency, and much higher than average organic food consumption. Slightly lower than average meat, and pasta consumption.

Cluster 6 (n=16): Higher than average meat, fish, nut, and sesame product frequency. Slightly higher than average pulses, pasta, and cooked food consumption.

At 12-months of age there were 7 clusters identified. The cluster attributes can be described for month 12 as follows:

Cluster 1 (n=18): Much lower-than-average yoghurt frequency, slightly lower than average sesame seed product consumption. Even distribution of exclusive breast and formula fed infants.

Cluster 2 (n=8): Higher than average raw food consumption. Much lower-than-average cereal frequency, and lower than average pulses frequency. Even distribution of exclusive breast and formula fed infants.

Cluster 3 (n=4): Much lower-than-average fruit frequency, but higher than average nut frequency. Slightly higher than average cooked food. This cluster consisted of children who were receiving exclusive formula feeding at 12-months.

Cluster 4 (n=38): Higher than average sesame product, and sesame seed frequency. Slightly higher than average pulses, and organic foods consumption. Predominantly, exclusively breastfed.

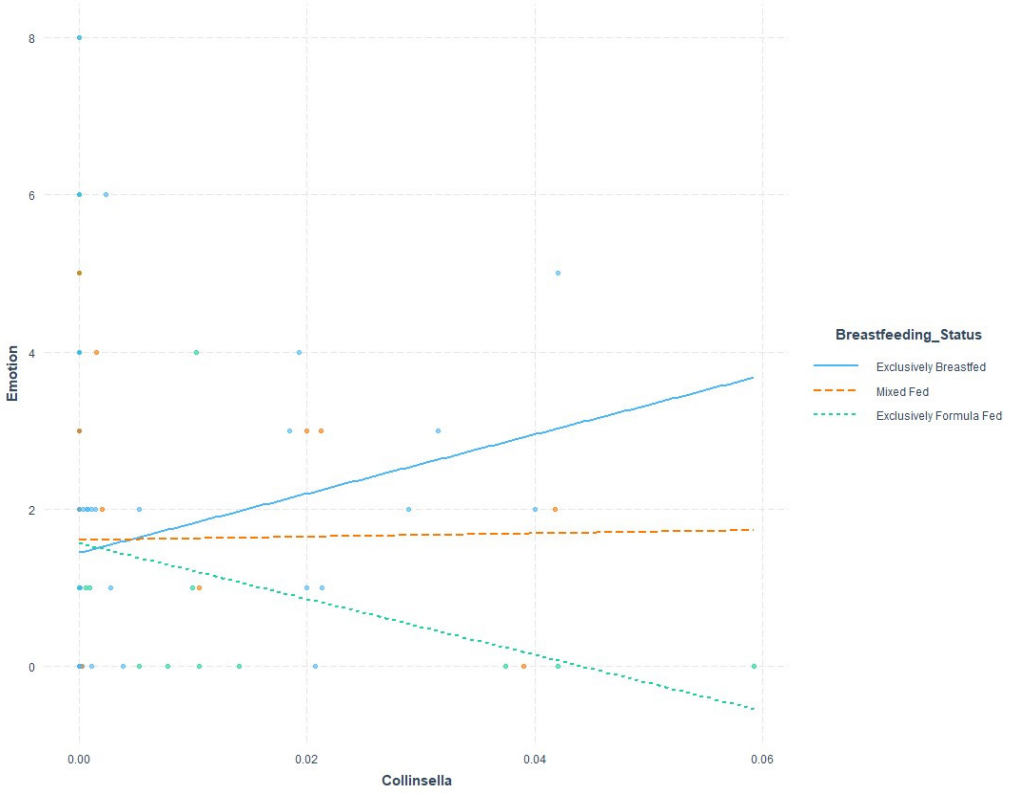
Cluster 5 (n=20): Much lower-than-average frequency of yoghurts with probiotics, and lower than average frequency of pulses and nuts. Slightly higher than average frequency of cooked food. Slightly more formula feeding than breastfeeding.

Cluster 6 (n=28): Lower than average raw food, and cooked food frequency. Higher than average consumption of pre-prepared or packaged foods. Predominantly children who were exclusively formula fed or receiving other liquids.

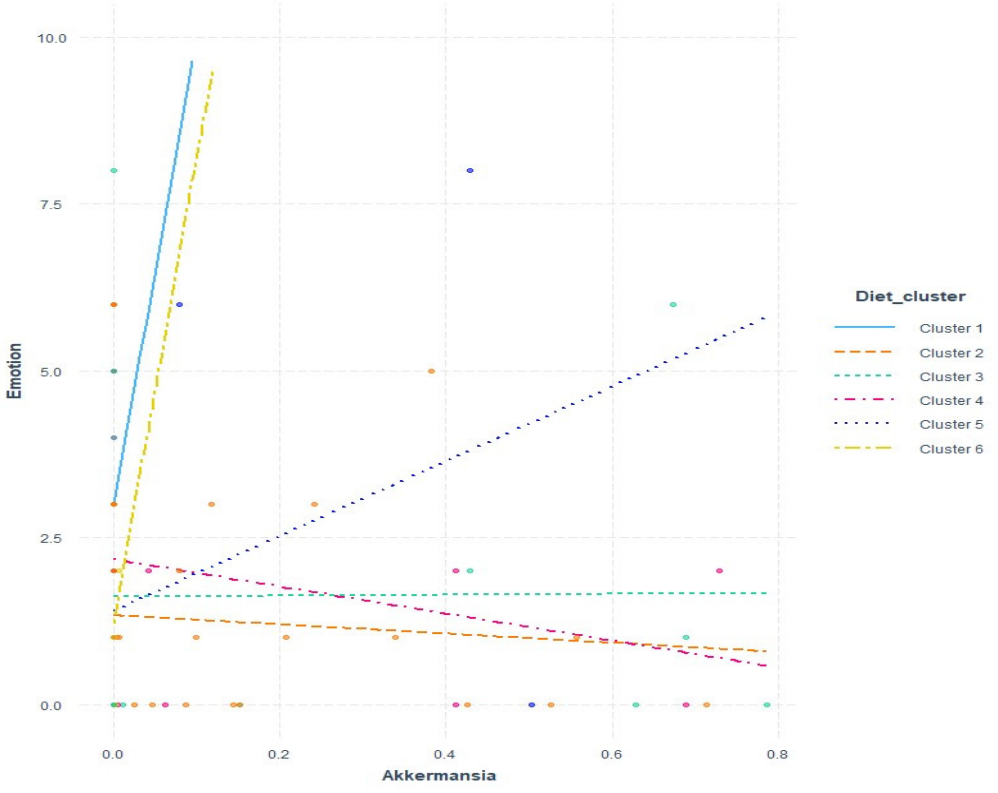
Cluster 7 (n=135): Scores for all variables were close to average. Higher numbers of children who were exclusively formula fed.

The results of the moderation analysis showed that the relationship between relative abundance of *Collinsella* measured at 1-month of age and emotional problems measured at 4-years was modified by milk-based diet (F (14, 158) = 2.900, $p < .01$, FDR = 0.02, Adj R^2 = 13.4%). Exclusive breastfeeding moderated the relationship between relative abundance of *Collinsella* and emotional problems, shown in Figure 4.4. The moderation positively affected the relationship, with increased relative abundance of *Collinsella* being associated with increased emotional problems. Conversely, when infants were exclusively formula fed, increased relative abundance of *Collinsella* was associated with decreased emotional problems. Confounding was present for both the number of siblings and maternal PSS score, which indicates that in addition to the relative abundance of *Collinsella,* both of these confounders influence the emotional problems measured at 4-years. There were no further significant associations between identified bacteria and SDQ outcomes moderated by diet at one month of age (see table 4.7).

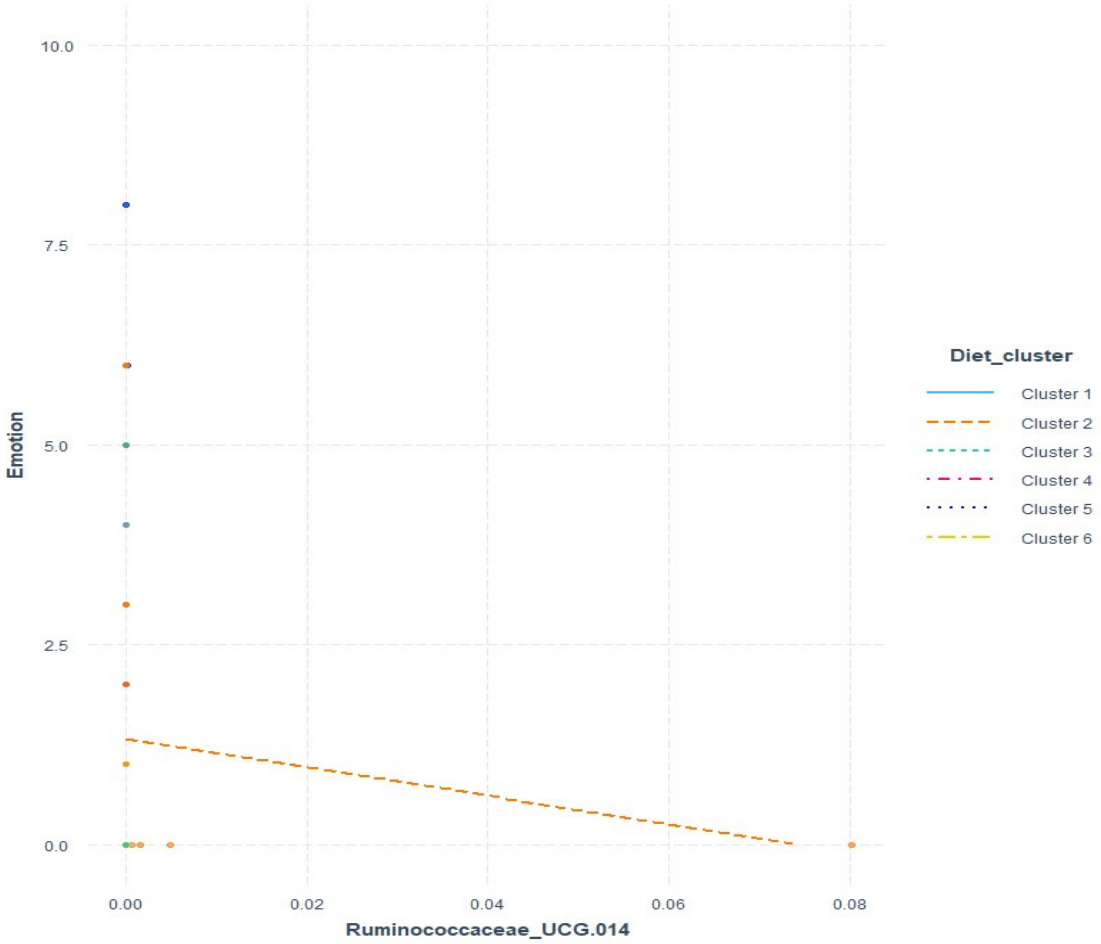
At 6-months of age the relationship between the relative abundance of both *Akkermansia* (F (14, 183) = 1.815, $p < .05$, FDR = 0.13, Adj R^2 = 0.05%), and an unclassified genus in the family Ruminococcaceae (F (14, 183) = 1.938, $p < .05$, FDR = 0.10, Adj R^2 = 6.3%), and the emotional problems subscale were significantly moderated by dietary cluster. Both models presented confounding by the number of siblings, in the model between relative abundance of *Akkermansia* and emotional problems there was additional adjustment by birthweight. This indicates that in addition to the influence of GM upon emotional problems, increases in emotional problems can also be explained in part by increased number of siblings and birthweight. The significant moderation of the relationship between the relative abundance of *Akkermansia* and emotional problems subscale was driven predominantly by dietary clusters 1, 5, and 6. In clusters 1, and 6, the relationship between relative abundance of *Akkermansia* and emotional problems was significantly positive, the same pattern can be seen for cluster 5 but with a less strong effect. Only cluster 2 had an effect upon the relative


Emotion
Collinsella
Breastfeeding_Status
Exclusively Breastfed
Mixed Fed
Exclusively Formula Fed

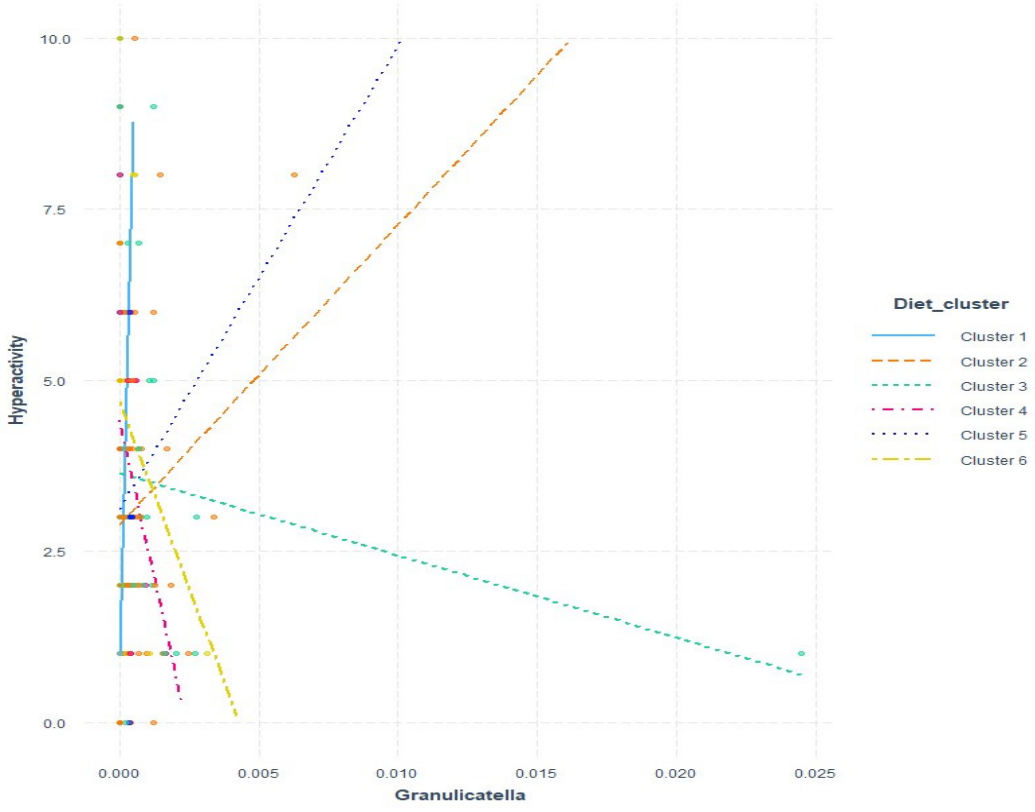
a)


Emotion
Akkermansia
Diet_cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

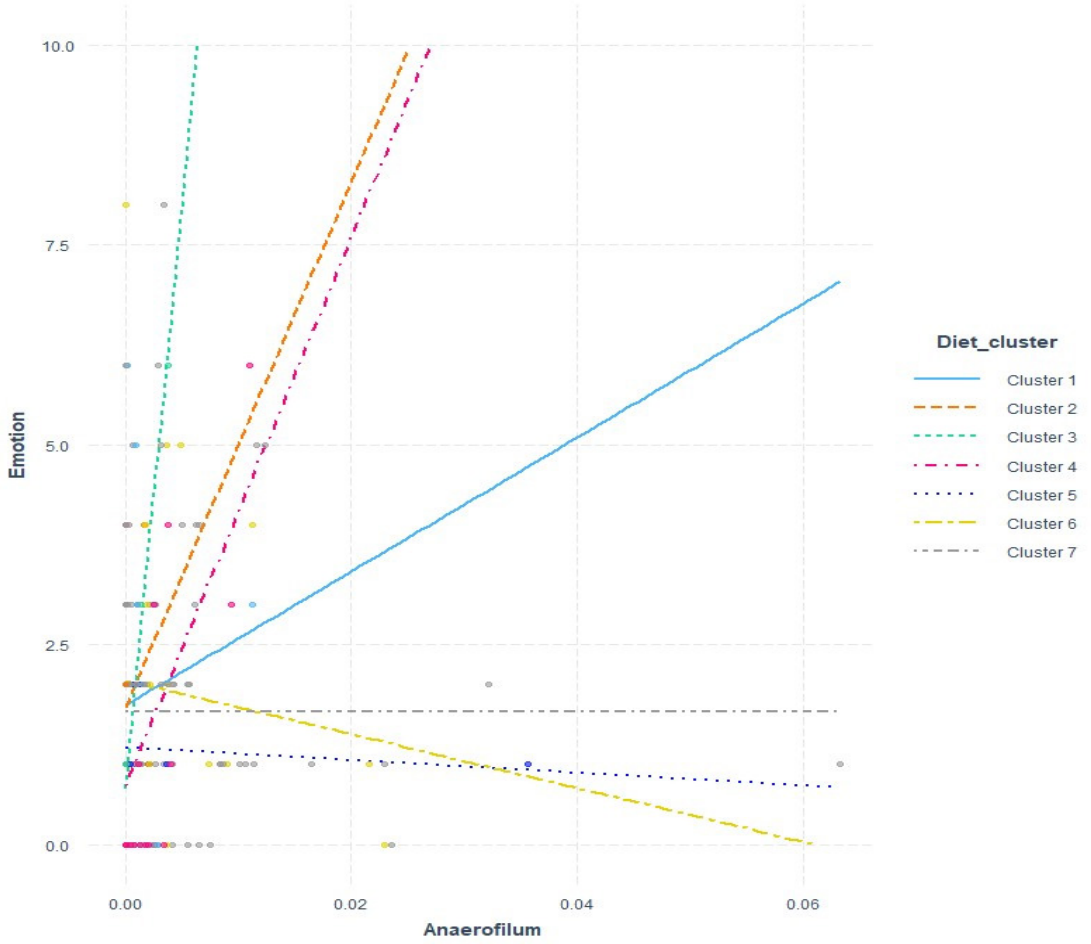
b)


10.0
7.5
5.0
2.5
0.0
Emotion
0.00
0.02
0.04
0.06
0.08
Ruminococcaceae_UCG.014
Diet_cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

c)


10.0
7.5
5.0
2.5
0.0
Hyperactivity
0.000
0.005
0.010
0.015
0.020
0.025
Granulicatella
Diet_cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

d)



e)

Figure 4.5. Interaction plots to show the moderation effect of milk-based diet and Diet cluster upon the relationship between a), emotional problems subscale and *Collinsella* at 1-month, b) emotional problems and *Akkermansia* at 6-months, c) emotional problems and Ruminococcaceae at 6-months, d) hyperactivity/inattention and *Granulicatella* at 6-months, and e) emotional problems and *Anaerofilum* at 12 -months.

Table 4.10. Moderation analyses of microbiota measured at 1-month and SDQ outcomes measured at 4-years moderated by milk-based diet.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
1	*Staphylococcus*	Emotion	2.261	-115.500	14	158	0.008	0.093	0.061 [c d]
1	*Rothia*	Emotion	2.291	-140.100	14	158	0.007	0.095	0.061 [c d]
1	*Blautia*	Emotion	0.235	-78.599	3	234	0.872	-0.010	0.978
1	*Salmonella*	Emotion	2.158	-133.400	14	158	0.012	0.086	0.061 [c d]
1	*Lactobacillus*	Emotion	2.326	-5.488	14	158	0.006	0.097	0.061 [c d]
1	*Collinsella*	Emotion	2.900	93.260	14	158	0.001	0.134	0.021 [c d]
1	*Escherichia/Shigella*	Emotion	2.124	-0.288	14	158	0.013	0.084	0.069 [c d]
1	*Pseudocitrobacter*	Emotion	2.264	66.410	14	158	0.008	0.093	0.061 [c d]
1	*Eubacterium*	Emotion	2.211	-21.080	14	158	0.009	0.090	0.064 [c d]
1	*Erysipelatoclostridium*	Emotion	2.495	6.286	14	158	0.003	0.109	0.061 [c d]
1	*Flavonifractor*	Conduct	2.321	508.000	14	158	0.006	0.097	0.061 [d h]
1	*Intestinibacter*	Conduct	2.379	-3406.000	14	158	0.005	0.101	0.061 [d h]
1	*Pseudocitrobacter*	Conduct	2.112	-82.330	14	158	0.014	0.083	0.069 [d h]
1	*Enterococcus*	Hyperactivity	1.732	-15.470	14	158	0.054	0.056	0.149 [f]
1	*Streptococcus*	Hyperactivity	1.634	4.413	14	158	0.075	0.049	0.171
1	*Clostridium*	Hyperactivity	1.812	-2.282	14	158	0.041	0.062	0.135 [f]
1	*Staphylococcus*	Hyperactivity	1.653	312.200	14	158	0.071	0.050	0.164 [f]
1	*Rothia*	Hyperactivity	1.678	-164.200	14	158	0.065	0.052	0.245
1	*Veillonella*	Hyperactivity	2.419	4.708	14	158	0.004	0.104	0.061
1	*Hungatella*	Peer	0.475	5.132	14	158	0.944	-0.045	0.978
1	*Staphylococcus*	Peer	0.645	331.000	14	158	0.824	-0.030	0.978
1	*Citrobacter*	Peer	0.544	-5.110	14	158	0.904	-0.039	0.978
1	*Pseudocitrobacter*	Peer	0.559	177.700	14	158	0.893	-0.037	0.978
1	*Eubacterium*	Peer	0.601	37.897	14	158	0.861	-0.034	0.978
1	*Pluralibacter*	Peer	0.502	-213.800	14	158	0.930	-0.042	0.891
1	*Subdoligranulum*	Peer	0.586	-6.253	14	158	0.863	-0.032	0.978
1	*Peptoniphilus*	Peer	0.515	764.000	14	158	0.922	-0.041	0.978
1	*Epulopiscium*	Peer	0.570	-25.640	14	158	0.875	-0.034	0.978
1	*Megasphaera*	Peer	0.524	147.900	14	158	0.917	-0.040	0.978
1	*Bifidobacterium*	Prosocial	0.518	0.103	14	158	0.920	-0.041	0.978
1	*Citrobacter*	Prosocial	0.542	2.446	14	158	0.905	-0.039	0.978
1	*Scardovia*	Prosocial	0.881	2165.000	14	158	0.580	-0.098	0.789
1	*Escherichia/Shigella*	Total Problem	1.880	-0.953	14	158	0.032	0.067	0.121 [h]
1	*Flavonifractor*	Total Problem	1.936	421.500	14	158	0.026	0.071	0.103 [h]
1	*Streptococcus*	Total Problem	1.917	4.839	14	158	0.028	0.069	0.318 [d h]

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted[c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.11. Moderation analyses of microbiota measured at 6-months and SDQ outcomes measured at 4-years moderated by diet cluster.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
6	*Eggerthella*	Emotion	1.451	-63.810	14	183	0.134	0.031	0.240
6	*Akkermansia*	Emotion	1.815	-6.485	14	183	0.039	0.055	0.133 [c e]
6	*Fusobacterium*	Emotion	1.702	187.000	14	183	0.058	0.048	0.150 [e]
6	*Prevotella_9*	Emotion	1.449	-3.883	14	183	0.135	0.031	0.240 [c]
6	Ruminococcaceae	Emotion	1.938	-16730.000	14	183	0.025	0.063	0.102 [c]
6	*Desulfovibrio*	Emotion	2.318	-6480.000	14	183	0.006	0.086	0.061 [c]
6	*Roseburia*	Emotion	1.108	-78.288	3	244	0.346	0.001	0.527
6	*Fusicatenibacter*	Emotion	1.429	32.740	14	183	0.143	0.030	0.245 [c]
6	*Tyzzerella_4*	Emotion	2.886	-10.100	14	183	0.001	0.118	0.021 [c]
6	*Lachnospira*	Emotion	0.995	-7.937	3	244	0.396	0.000	0.585
6	*Dorea*	Conduct	3.368	-69.020	14	183	0.000	0.144	0.008 [d]
6	*Actinomyces*	Conduct	1.731	32.435	14	183	0.053	0.049	0.149 [h]
6	*Gemella*	Conduct	2.215	-420.800	14	183	0.009	0.079	0.064 [h]
6	*Romboutsia*	Hyperactivity	1.747	-21.990	14	183	0.050	0.050	0.149
6	*Granulicatella*	Hyperactivity	3.696	1147.726	3	244	0.012	0.032	0.069
6	*Gemella*	Hyperactivity	1.702	-803.100	14	183	0.058	0.048	0.150
6	*Hungatella*	Hyperactivity	1.297	-7.543	14	183	0.213	0.021	0.345
6	*Prevotella*	Hyperactivity	3.525	-152.871	3	244	0.016	0.030	0.069
6	*Subdoligranulum*	Hyperactivity	3.598	49.344	3	244	0.014	0.031	0.069
6	*Tyzzerella_4*	Peer	1.106	101.100	14	183	0.355	0.008	0.532
6	*Sutterella*	Peer	0.465	-58.450	14	183	0.949	-0.040	0.978
6	*Ruminiclostridium_5*	Peer	0.512	-234.400	14	183	0.924	-0.036	0.978
6	*Streptococcus*	Peer	0.371	1.640	14	183	0.982	-0.047	0.984
6	*Lactococcus*	Peer	0.466	-617.000	14	183	0.948	-0.039	0.978
6	*Lachnoclostridium*	Peer	0.390	2.360	14	183	0.976	-0.045	0.984
6	*Blautia*	Peer	0.588	2.087	3	244	0.623	-0.005	0.815
6	*Actinomyces*	Peer	0.573	-161.900	14	183	0.884	-0.031	0.978
6	*Peptoclostridium*	Peer	0.502	-30.100	14	183	0.930	-0.037	0.978
6	*Enterobacter*	Peer	0.360	-2.246	14	183	0.984	-0.048	0.984
6	*Enterobacter*	Prosocial	1.224	-5.640	14	183	0.261	0.016	0.404 [f]
6	*Paraprevotella*	Prosocial	0.861	583.500	14	183	0.602	-0.010	0.799
6	*Bifidobacterium*	Prosocial	0.895	1.798	14	183	0.565	-0.008	0.789
6	*Sutterella*	Total Problem	1.481	543.800	14	183	0.122	0.033	0.234
6	*Romboutsia*	Total Problem	1.528	27.318	14	183	0.105	0.036	0.214
6	*Prevotella*	Total Problem	1.682	-304.200	14	183	0.063	0.046	0.152
6	*Enterobacter*	Total Problem	1.500	22.288	14	183	0.115	0.034	0.225

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.12. Moderation analyses of microbiota measured at 12-months and SDQ outcomes measured at 4-years moderated by diet cluster.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
12	*Eubacterium*	Emotion	1.419	119.000	14	181	0.148	0.029	0.247 [c e]
12	*Alistipes*	Emotion	1.446	-63.230	14	181	0.136	0.031	0.240 [c e]
12	*Lachnospira*	Emotion	1.617	-28.710	14	181	0.078	0.042	0.173 [c e]
12	Ruminococcaceae	Emotion	1.566	133.800	14	181	0.093	0.039	0.196 [c e]
12	*Subdoligranulum*	Emotion	1.511	-3.299	14	181	0.111	0.035	0.222 [c e]
12	*Dialister*	Emotion	1.447	-5.421	14	181	0.136	0.031	0.240 [c e]
12	*Tyzzerella_3*	Emotion	1.468	202.500	14	181	0.127	0.033	0.240 [c e]
12	*Odoribacter*	Emotion	1.561	-808.700	14	181	0.094	0.039	0.196 [c e]
12	Lachnospiraceae	Emotion	1.835	220.100	14	181	0.037	0.057	0.130 [c e]
12	*Catenibacterium*	Emotion	3.540	-668.600	3	212	0.016	0.034	0.069
12	*Flavonifractor*	Emotion	1.701	-80.110	14	181	0.059	0.048	0.150 [c e]
12	*Anaerofilum*	Emotion	1.749	220.100	14	181	0.050	0.051	0.149 [c e]
12	*Ruminococcus_1*	Emotion	1.686	-24.650	14	181	0.062	0.047	0.152 [c]
12	*Megamonas*	Emotion	0.721	8.960	3	212	0.541	-0.004	0.766
12	*Eubacterium*	Conduct	1.583	9.249	14	181	0.088	0.040	0.766 [d]
12	*Parasutterella*	Conduct	2.118	-32.900	14	181	0.013	0.074	0.069 [d]
12	*Staphylococcus*	Conduct	1.831	-41.850	14	181	0.037	0.056	0.130 [d]
12	Lachnospiraceae	Conduct	2.407	-1515.000	14	181	0.004	0.092	0.061 [d]
12	Ruminococcaceae	Conduct	1.255	-46.200	14	181	0.043	0.054	0.137 [d]
12	Erysipelotrichaceae	Hyperactivity	0.726	-63.590	14	181	0.051	0.050	0.149 [f g]
12	*Coprococcus_2*	Hyperactivity	0.888	-36.874	3	212	0.496	-0.003	0.712
12	*Turicibacter*	Peer	0.612	107.200	14	181	0.240	0.018	0.383
12	*Prevotella_9*	Peer	0.860	-2.377	14	181	0.746	-0.020	0.952
12	*Tyzzerella*	Peer	1.225	186.600	14	181	0.572	-0.008	0.789
12	*Tyzzerella_4*	Peer	1.010	12.980	14	181	0.862	-0.029	0.978
12	*Ruminiclostridium_5*	ProSocial	1.676	-21.320	14	181	0.603	-0.010	0.799
12	*Clostridium*	ProSocial	1.962	10.754	14	181	0.261	0.016	0.404 [f]
12	*Romboutsia*	ProSocial	1.582	-4.911	14	181	0.445	0.001	0.649
12	*Eubacterium*	Total Problem	1.255	937.300	14	181	0.064	0.046	0.152
12	*Staphylococcus*	Total Problem	0.726	-11660.000	14	181	0.023	0.065	0.098
12	*Odoribacter*	Total Problem	0.888	-499.900	14	181	0.088	0.040	0.191

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

Table 4.13. Moderation analyses of microbiota measured at 6-months and SDQ outcomes measured at 4-years moderated by diet cluster following adjustment for small sample size.

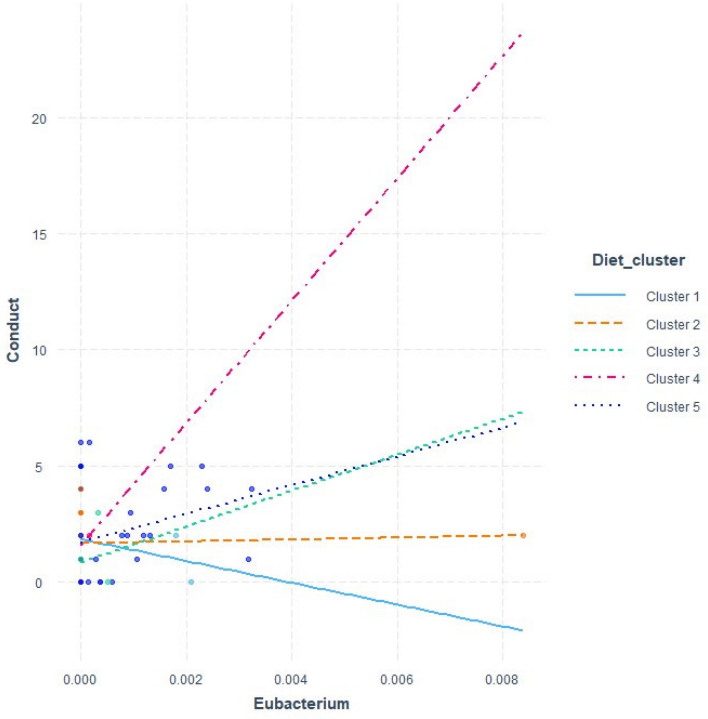
Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
6	*Eggerthella*	Emotion	1.489	8.001	14	183	0.119	0.034	0.199 [c]
6	*Akkermansia*	Emotion	1.853	-4.258	14	183	0.034	0.057	0.126 [c e]
6	*Fusobacterium*	Emotion	1.771	73.780	14	183	0.046	0.052	0.134 [c]
6	*Prevotella_9*	Emotion	1.488	-3.199	14	183	0.119	0.034	0.199 [c e]
6	Ruminococcaceae	Emotion	1.961	-8250.000	14	183	0.023	0.064	0.096 [c]
6	*Desulfovibrio*	Emotion	2.349	-4773.000	14	183	0.005	0.088	0.070 [c e g]
6	*Roseburia*	Emotion	1.261	-51.987	3	242	0.289	0.003	0.430
6	*Fusicatenibacter*	Emotion	1.464	3.975	14	183	0.129	0.032	0.210 [c]
6	*Tyzzerella_4*	Emotion	2.941	-6.397	14	183	0.000	0.122	0.016 [c]
6	*Lachnospira*	Emotion	1.559	-10.190	14	183	0.094	0.038	0.173 [c]
6	*Dorea*	Conduct	3.332	-1.090	14	183	0.000	0.143	0.006 [d]
6	*Actinomyces*	Conduct	0.279	30.053	3	242	0.841	-0.009	0.980
6	*Gemella*	Conduct	2.185	-435.000	14	183	0.010	0.078	0.087
6	*Romboutsia*	Hyperactivity	1.668	-4.518	14	183	0.066	0.046	0.163
6	*Granulicatella*	Hyperactivity	3.568	724.772	3	242	0.015	0.030	0.090
6	*Gemella*	Hyperactivity	1.629	-768.400	14	183	0.075	0.043	0.163
6	*Hungatella*	Hyperactivity	1.224	-5.665	14	183	0.261	0.016	0.397
6	*Prevotella*	Hyperactivity	3.455	-245.271	3	242	0.017	0.029	0.096
6	*Subdoligranulum*	Hyperactivity	1.451	20.400	14	183	0.134	0.031	0.213
6	*Tyzzerella_4*	Peer	0.992	61.720	14	183	0.464	-0.001	0.609
6	*Sutterella*	Peer	0.490	-47.570	14	183	0.937	-0.038	0.980
6	*Ruminiclostridium_5*	Peer	0.510	-153.400	14	183	0.925	-0.036	0.980
6	*Streptococcus*	Peer	0.402	2.003	14	183	0.973	-0.045	0.980
6	*Lactococcus*	Peer	0.472	-40.730	14	183	0.946	-0.039	0.980
6	*Lachnoclostridium*	Peer	0.271	2.325	3	242	0.846	-0.009	0.980
6	*Blautia*	Peer	0.524	2.035	14	183	0.917	-0.035	0.980
6	*Actinomyces*	Peer	0.598	-113.300	14	183	0.865	-0.030	0.980
6	*Peptoclostridium*	Peer	0.518	-6.925	14	183	0.921	-0.036	0.980
6	*Enterobacter*	Peer	0.378	-0.918	14	183	0.980	-0.046	0.980
6	*Enterobacter*	Prosocial	1.341	-2.963	14	183	0.187	0.024	0.292 [f]
6	*Paraprevotella*	Prosocial	0.954	357.000	14	183	0.503	-0.003	0.635
6	*Bifidobacterium*	Prosocial	0.963	1.119	14	183	0.493	-0.003	0.635
6	*Sutterella*	Total Problem	1.574	344.900	14	183	0.090	0.039	0.173 [d]
6	*Romboutsia*	Total Problem	1.631	31.682	14	183	0.074	0.043	0.163
6	*Prevotella*	Total Problem	1.773	-401.000	14	183	0.046	0.052	0.134
6	*Enterobacter*	Total Problem	1.601	15.966	14	183	0.082	0.041	0.167 [d]

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth

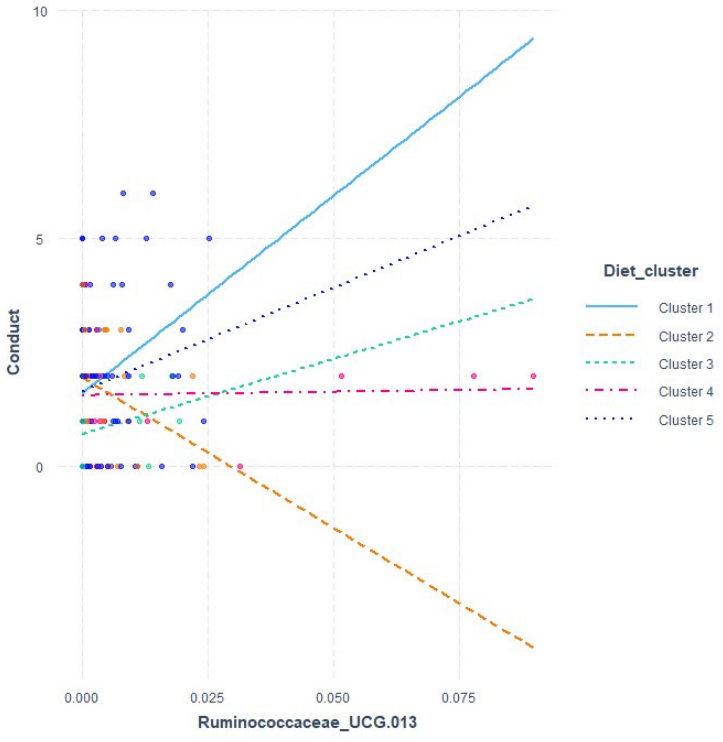
Table 4.14. Moderation analyses of microbiota measured at 12-months and SDQ outcomes measured at 4-years moderated by diet cluster following adjustment for small sample size.

Month	Microbiota Genus	SDQ[a]	F	β	DF	DF	P value	Adj R^2	FDR[b]
12	*Eubacterium*	Emotion	0.410	-170.988	3	200	0.746	-0.009	0.909
12	*Alistipes*	Emotion	1.543	-22.460	14	172	0.100	0.039	0.177 [c e]
12	*Lachnospira*	Emotion	1.654	-12.600	14	172	0.070	0.047	0.163 [c e]
12	Ruminococcaceae	Emotion	1.648	104.300	14	172	0.071	0.046	0.163 [c e]
12	*Subdoligranulum*	Emotion	1.604	0.261	14	172	0.082	0.044	0.167 [c e]
12	*Dialister*	Emotion	1.573	-3.447	14	172	0.091	0.041	0.173 [c e]
12	*Tyzzerella_3*	Emotion	1.559	-406.000	14	172	0.095	0.040	0.173 [c e]
12	*Odoribacter*	Emotion	1.695	-503.000	14	172	0.060	0.050	0.162 [c e]
12	Lachnospiraceae	Emotion	2.089	209.100	14	172	0.015	0.076	0.090 [c e]
12	*Catenibacterium*	Emotion	3.879	-417.245	3	200	0.010	0.041	0.087
12	*Flavonifractor*	Emotion	2.330	-41.946	3	200	0.076	0.019	0.163
12	*Anaerofilum*	Emotion	1.829	155.600	14	172	0.038	0.059	0.126 [c e]
12	*Ruminococcus_1*	Emotion	1.901	60.914	14	172	0.029	0.064	0.115 [c e]
12	*Megamonas*	Emotion	1.161	6.626	3	200	0.326	0.002	0.473
12	*Eubacterium*	Conduct	3.247	-383.465	3	200	0.023	0.032	0.093 [d]
12	*Parasutterella*	Conduct	2.084	-20.830	14	172	0.015	0.075	0.090 [d]
12	*Staphylococcus*	Conduct	2.708	-335.300	14	172	0.001	0.114	0.028 [d]
12	Lachnospiraceae	Conduct	1.988	-8.707	14	172	0.021	0.069	0.096 [d]
12	Ruminococcaceae	Conduct	2.476	-127.700	14	172	0.003	0.100	0.054 [f g]
12	Erysipelotrichaceae	Hyperactivity	2.009	-57.070	14	172	0.020	0.071	0.096 [f g]
12	*Coprococcus_2*	Hyperactivity	1.801	-178.200	14	172	0.042	0.057	0.133
12	*Turicibacter*	Peer	1.052	167.500	14	172	0.405	0.004	0.543
12	*Prevotella_9*	Peer	0.575	-1.993	14	172	0.882	-0.033	0.980
12	*Tyzzerella*	Peer	0.756	-365.100	14	172	0.715	-0.019	0.887
12	*Tyzzerella_4*	Peer	0.457	-7.669	14	172	0.952	-0.043	0.980
12	*Ruminiclostridium_5*	ProSocial	1.062	143.600	14	172	0.395	0.005	0.541 [f]
12	*Clostridium*	ProSocial	1.134	6.378	14	172	0.332	0.010	0.473 [f]
12	*Romboutsia*	ProSocial	1.120	-10.070	14	172	0.344	0.009	0.480
12	*Eubacterium*	Total Problem	1.710	-675.700	14	172	0.057	0.051	0.160
12	*Staphylococcus*	Total Problem	2.177	-5877.000	14	172	0.010	0.081	0.087
12	*Odoribacter*	Total Problem	1.844	-71.160	14	172	0.036	0.060	0.126 [d]

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.
[b] FDR – False Discovery Rate
Confounders adjusted: [c] Number of Siblings, [d] PSS score, [e] Birthweight, [f] Remoteness, [g] SEIFA, [h] Mode of Birth



a)



b)

Figure 4.7. Interaction plots to show the moderation effect of Diet cluster upon the relationship between conduct problems subscale and a), *Eubacterium,* and b) Ruminococcaceae at 12 -months

Table 4.15. Moderation analyses of SELBAL balance parameter at 1-, 6-., and 12- months of age and SDQ outcomes measured at 4-years, moderated by milk-based diet and diet cluster following adjustment for small sample size.

Month	SDQ[a]	F	β	DF	DF	P value	Adj R^2
1	Emotion	2.513	0.030	14	158	0.003	0.110
1	Conduct	2.495	0.109	14	158	0.003	0.109
1	Hyperactivity/inattention	1.838	-0.251	14	158	0.037	0.064
1	Peer	0.597	0.043	14	158	0.865	-0.034
1	Pro Social	0.597	0.064	14	158	0.865	-0.034
1	Total Difficulties	2.259	-0.149	14	158	0.008	0.093
6	Emotion	1.693	0.303	14	182	0.060	0.047
6	Conduct	3.248	0.518	14	182	0.000	0.138
6	Hyperactivity/inattention	2.037	0.677	14	182	0.017	0.069
6	Peer	0.639	0.313	14	182	0.830	-0.027
6	Pro Social	1.596	0.254	14	182	0.084	0.041
6	Total Difficulties	1.849	-0.169	14	182	0.035	0.057
12	Emotion	1.581	-0.075	14	182	0.089	0.042
12	Conduct	2.071	-0.019	14	182	0.016	0.075
12	Hyperactivity/inattention	3.050	0.537	14	182	0.000	0.134
12	Peer	1.291	0.095	14	182	0.217	0.021
12	Pro Social	2.901	0.318	3	200	0.036	0.027
12	Total Difficulties	2.277	-0.009	14	182	0.007	0.088

[a] SDQ – Strengths and Difficulties Questionnaire including Emotional Problems, Conduct problems, Hyperactivity/inattention, Peer problems, Prosocial behaviour, and Total Difficulties scales.

abundance of an unclassified genus in the family Ruminococcaceae and emotional problems, which now showed a moderately negative relationship. Additionally, at 6-months of age the relationship between relative abundance of *Granulicatella* and subscale scores of hyperactivity/inattention were significantly moderated by diet cluster, in the unadjusted model (F (3, 244) = 3.696, $p < .05$, FDR = 0.07, Adj R^2 = 3.2%). All dietary clusters influence the relationship between the relative abundance of *Granulicatella* and scores of hyperactivity/inattention, with clusters 1, 2, and 5 showing increasing hyperactivity/inattention scores with increased relative abundance of *Granulicatella,* and clusters 3, 4, and 6, showing the opposite effect. Following correction for small cluster sizes, using the same method described in Chapter 3, section 3.4.6, the same relationships remain significant (See table 4.13). The clusters that remained following assessment for small cluster size are listed below.

At 6-months

Cluster 1 (n=159): Lower than average fish frequency, soy products and sesame products. Slightly higher than average raw food consumption.

Cluster 2 (n=75): Lower than average raw foods, and cooked foods. Average scores on all other food types.

Cluster 3 (n=23): Lower than average fruit and vegetable consumption. Average scores on all other food types.

Cluster 4 (n=21): Lower than average fish frequency, and much higher than average organic food consumption. Slightly lower than average meat, and pasta consumption.

Cluster 5 (n=16): Higher than average meat, fish, nut, and sesame product frequency. Slightly higher than average pulses, pasta, and cooked food consumption.

At 12-months

Cluster 1 (n=18): Much lower-than-average yoghurt frequency, slightly lower than average sesame seed product consumption. Even distribution of exclusive breast and formula fed infants.

Cluster 2 (n=38): Higher than average sesame product, and sesame seed frequency. Slightly higher than average pulses, and organic foods consumption. Predominantly, exclusively breastfed.

Cluster 3 (n=20): Much lower-than-average frequency of yoghurts with probiotics, and lower than average frequency of pulses and nuts. Slightly higher than average frequency of cooked food. Slightly more formula feeding than breastfeeding.

Cluster 4 (n=28): Lower than average raw food, and cooked food frequency. Higher than average consumption of pre-prepared or packaged foods. Predominantly children who were exclusively formula fed or receiving other liquids.

Cluster 5 (n=135): Scores for all variables were close to average. Higher numbers of children who were exclusively formula fed.

At 12-months of age the relationship between relative abundance of *Anaerofilum* and emotional problems was significantly moderated by diet cluster (F (14, 181) = 1.749, $p < .05$, FDR = 0.15, Adj R^2 = 5.1%), when adjusted for the number of siblings and birthweight. Dietary clusters 1, 2, 3, and 4 moderated the relationship between relative abundance of *Anaerofilum* significantly, so that it was now positively associated with emotional problem scores, whereas cluster 6 presented with a negative relationship between *Anaerofilum* and emotional problem scores. This relationship was no longer significant following correction for small cluster sizes. Similar to Chapter 3, clusters 2 (n=8), and 3 (n=4), were removed from the dietary clusters and the moderation analyses were rerun. It can be seen

that there are now three significant moderation relationships at 12-months of age, although the relationship between the bacteria in in the Erysipeloctrichaceae unclassified group were no longer significant following sensitivity checks for outliers. Of the two remaining significant relationships, the relative abundance *Eubacterium* measured at 12-months of age, and the conduct subscale of the SDQ measured at 4 years of age was significantly moderated by diet cluster (F (3, 200) = 3.247, p < .05, FDR = 0.10, Adj R^2 = 3.2%). Clusters 3, 4, and 5 moderated the relationship so that *Eubacterium* is positively associated with conduct problems, and cluster 1 moderated the relationship in a negative direction. Relative abundance of an unclassified group of Ruminococcaceae was also significantly associated with conduct problems at 4-years when modified by diet (F (14, 172) = 2.480, p < .01, FDR = 0.54, Adj R^2 = 10.0%). Clusters 1, 3, and 5, show a positive association and cluster 2 shows a negative association.

Finally, when investigating the relationships between SDQ subscales and total difficulties and the balance parameters derived from the SELBAL analyses, there were no significant moderation of the relationship between balance parameters and SDQ scores by diet at either 1-, 6-, or 12-months of age (See table 4.15). Corrections for multiple comparisons were not conducted on these analyses as there were no significant interactions.

4.4 Discussion

Dietary intake has been shown to be closely related to early child health outcomes from the very beginning of life (Alimujiang et al., 2018; Ness et al., 2005), including the parental choice of milk-based diet and first solid foods introduced to an infant (Hair et al., 2016). These parental dietary choices influence the child during a sensitive period of development of both the GM and brain, influencing further both cognitive and behavioural development (Aatsinki et al., 2019; Carlson et al., 2018; Clarke, O'Mahony, Dinan, & Cryan, 2014). These sensitive periods of development have been extensively investigated in literature exploring associations between composition and diversity of GM and the influence upon the gut-brain axis (GBA) (Cowan et al., 2020) However, when considering the influence of diet upon the relationship between GM and behavioural outcomes, diet has previously only been considered as a confounder. This study furthers the understanding of the influence of dietary intake during the first year of life, upon the composition of the GM, and the role that it plays in the relationship between GM and behaviour.

It was hypothesised that bacteria identified as candidate bacteria of interest would be significantly associated with SDQ outcomes measured at 4-years. Several bacteria presented with significant associations at all three time points, with SDQ subscales measured at 4-years. However, it can be seen that 6-months of age showed the most frequent significant relationships between GM and behavioural outcome, indicating that this might be a critical period of change for GM composition and the influence upon behaviour. Specifically, relative abundance of the bacteria *Tyzzerella_4*, measured

at 6-months, was positively associated with both emotional and peer problem subscales at 4-years. These two subscales of the SDQ are added together to create a score for risk of internalising problems, thus, at 6-months of age, increased relative abundance of *Tyzzerella_4* may be a risk factor for these problems. These results were also in the unadjusted model, and therefore were not confounded by any of the covariates measured, which suggests that the direct relationship explains all of the variance in SDQ scores and are not additionally explained by any other environmental factors measured. However, when investigating the dispersion of *Tyzzerella_4* within the population, it was apparent that the relationship could potentially be driven by a small number of individuals. Following sensitivity checks to remove the relative abundance of these from the regression model, the relationship between *Tyzzerella_4* at 6-month of age and emotional problems at 4-years was no longer significant. The relationship between *Tyzzerella_4* at 6-month of age and peer problems remained significant. This highlights the need to carefully investigate relative abundance and the need to look at the overall population dispersion when considering significant results of microbes present in small numbers. Interestingly, greater relative abundance of *Tyzzerella_4* in mothers has been associated with Postpartum Depressive Disorder (PPD) (Chen et al., 2020; Zhou, Chen, Yu, & Yang, 2020) and is positively associated with both repeated use of antibiotics and childhood obesity associated with a high fat diet (Chen et al., 2020). Furthermore, *Tyzzerella*, specifically the sub-group 4, has been associated with gut inflammation and is also often found in the intestine of children diagnosed with autism spectrum disorder (ASD) (Ma et al., 2019). Overall, the pattern of results for *Tyzzerella_4* found in this study aligns with previous findings that associate this bacteria group with increased risk of internalising problems. Furthermore, this adds to the evidence indicating the first year of life as a critical window of microbiota change, which is associated with behavioural development into later childhood.

Again at 6-months of age, there was a significant relationship between increased relative abundance of *Gemella* and lower scores on both conduct problems and hyperactivity/inattention subscales at 4-years. These two subscales form the risk of externalising problems subscale of the SDQ, thus increased relative abundance of *Gemella* at 6-months could be a protective factor that protect against developing externalising problems. This bacterium is a Gram-negative cocci, typically associated with the oral microbiome or upper digestive system (Ramanathan, Gordon, & Shrestha, 2020). Most notably, this bacterium is associated with increased risk of endocarditis in the aging population, whereby the effects of age allow for the bacteria to cross the gut barrier, enter the blood stream, and cause infection to the heart (Ramanathan et al., 2020). To date there is little evidence to show influence of this bacterium upon human behaviour, however one study investigating rats that present with autistic-like traits induced by valproate found a positive association between anxiety behaviours and increased relative abundance of *Gemella* (Kong et al., 2021). Due to the lack of evidence supporting the relationship between this bacterium and behavioural outcomes, further investigation

and replication is necessary to understand whether this genus of bacteria plays an important role in human behavioural development.

Finally, decreased relative abundance of *Prevotella,* measured at 6-months, was associated with increased scores on both the hyperactivity/inattention subscale and total difficulties, measured at 4-years. This result is consistent with previous investigations of the same cohort, measuring behaviour outcomes using the child behaviour checklist (CBCL) at 2-years. This indicates that the effect of *Prevotella,* which can be seen behaviourally at 2-years, is consistent up until 4-years of age. However, this again was driven by a small number of individuals within the population, and both relationships were no longer significant following removal of outliers during sensitivity checks.

The current picture is that there is emerging evidence which gives strength to the predictive value of microbiota measures taken at 6-months of age, although concurrent microbiota measures taken at the time of the behavioural measure would lend further credence to this assertation, and also help to understand the mechanisms, whether this is an effect of concurrent metabolic processes or an effect of metabolites on the brain at a critical period earlier in development. Overall, at all age points there were several bacteria genera identified that were directly associated with SDQ outcomes. These results were significant in both unadjusted models and models that were adjusted for several potentially confounding variables, showing that the composition of the gut microbiota is a significant predictor of behaviour in 4-year-olds. The most significant confounders associated with emotional problems were birthweight and the increased number of siblings. Likewise, the most significant confounders associated with conduct problems were mode of birth, caesarean was associated with increased problems, and increased maternal PSS scores. The confounding present in these analyses' highlights that whilst the relationship between relative abundance of bacteria and SDQ outcomes is significant, these are other variables that add to the model and additionally explain the variance in SDQ scores, highlighting the complexity of these relationships.

Further investigation of the balance parameter extracted during the SELBAL process, showed that there were several significant relationships between the balance parameter and the subscales and total difficulties of the SDQ measured at 4 years. This represents the compositional balance of the microbiota identified using the SELBAL method, relating to the outcome variable. This can be calculated for each sample included in the analysis and therefore the balance variable that is generated represents how well the compositional balance of each of the samples relate to the bacteria that are identified. The compositional balance of the microbiota at 1-month were positively associated with emotion and conduct problems and Total difficulties. At 6-months, the compositional balance was also associated with conduct problems as well as hyperactivity/inattention. At 12-months of age the compositional balance was positively associated with hyperactivity/inattention, pro social behaviour and the total difficulties scale. This would indicate that the bacteria that were extracted at each of

these time points, relative to each of the outcome variables measuring SDQ, where there is elevated relative abundance, there is an increase in each of the SDQ scores.

The second aim of this study was to investigate the effect of diet upon the relationship between GM, measured at 1-, 6-, and 12-months, and SDQ outcomes measured at 4-years, through moderation analyses. For all time points it was found that the direct relationships that were identified between GM and SDQ outcomes were not moderated by diet, and that many of the direct relationships that were established persisted (See figure 4.5 for synthesis of significant direct relationships). The exceptions to this were the relationship between *Eubacterium* and *Odoribacter* measured at 12-months, and Total difficulties scores. *Odoribacter* was also no longer significantly related to emotional problems when measured at 12-months. However, several relationships that were previously non-significant showed significance following moderation by diet (See figure 4.6 for synthesis of indirect relationships moderated by diet). In particular it can be seen that the relationship between GM and the emotional subscale showed a pattern of significance at each time point. These results are interesting as there is a wealth of literature that explores, at all ages, the influence of dietary consumption upon emotional regulation (Aparicio, Canals, Arija, De Henauw, & Michels, 2016; Haycraft, Farrow, Meyer, Powell, & Blissett, 2011; Jarman, Edwards, & Blissett, 2022; Ling & Zahry, 2021; Rogers & Blissett, 2019). This bidirectional relationship is one that may start from the very first milk-based diet that is consumed and develops throughout life being influenced by many factors such as stress exposure (Ulrich-Lai, Fulton, Wilson, Petrovich, & Rinaman, 2015) and depression (Paans et al., 2019). The results of these analyses allow us insight into the underlying biological mechanisms between diet, GM, and emotional regulation. Firstly, at 1-month of age, the relationship between relative abundance of *Collinsella* and emotional problems measured at 4-years was significantly moderated by the type of milk-based diet an infant was receiving. Those receiving exclusive breastfeeding showed a positive association between emotional problems and relative abundance of *Collinsella*, whereas those exclusively formula fed, showed a negative association between emotional problems and relative abundance of *Collinsella*. This result contradicts our expectations that breastfeeding would always moderate the relationship between GM bacteria and developmental outcomes in a positive manner. Interestingly, *Collinsella,* is generally associated with infants from vaginal births (Dogra et al., 2015) and is transferred from the maternal gut microbiota to the infant through breastmilk (Jost, Lacroix, Braegger, Rochat, & Chassard, 2014). Furthermore, increased relative abundance of *Collinsella* is associated with a diet low in fibre, and associated with overweight and obesity in pregnant women. It is therefore plausible that the diet quality of the mother is indirectly influencing the behavioural outcomes of the child through this route. Further investigation should consider measuring the microbiota composition and quality of breastmilk in addition to the exclusivity

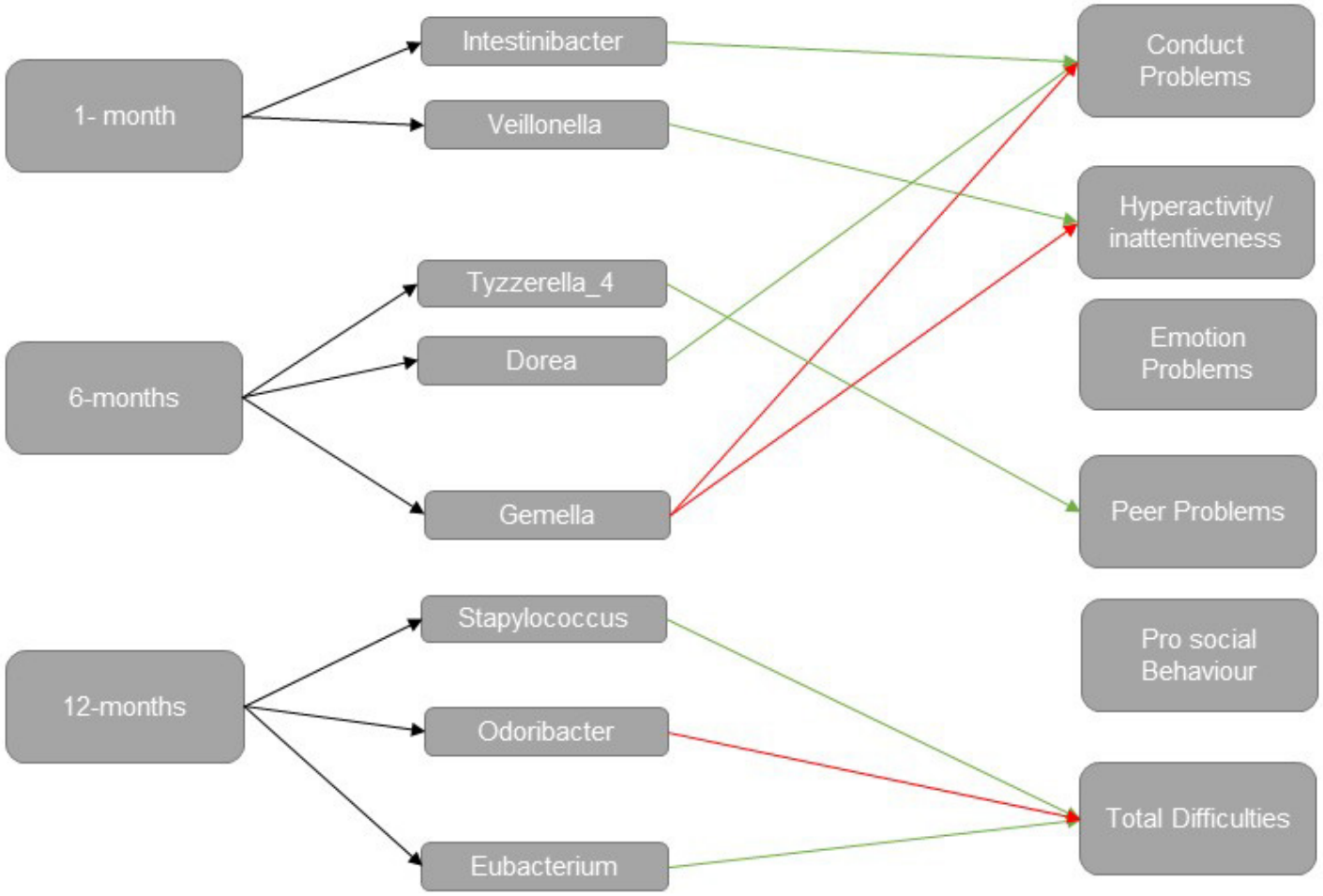


Figure 4.8 Synthesis figure illustrating the significant direct relationships between bacteria extracted at 1-, 6-, and 12 -months and SDQ outcomes measured at 4-years following sensitivity analyses
Note. Green lines denote direct positive relationships and red lines denote direct negative relationships.

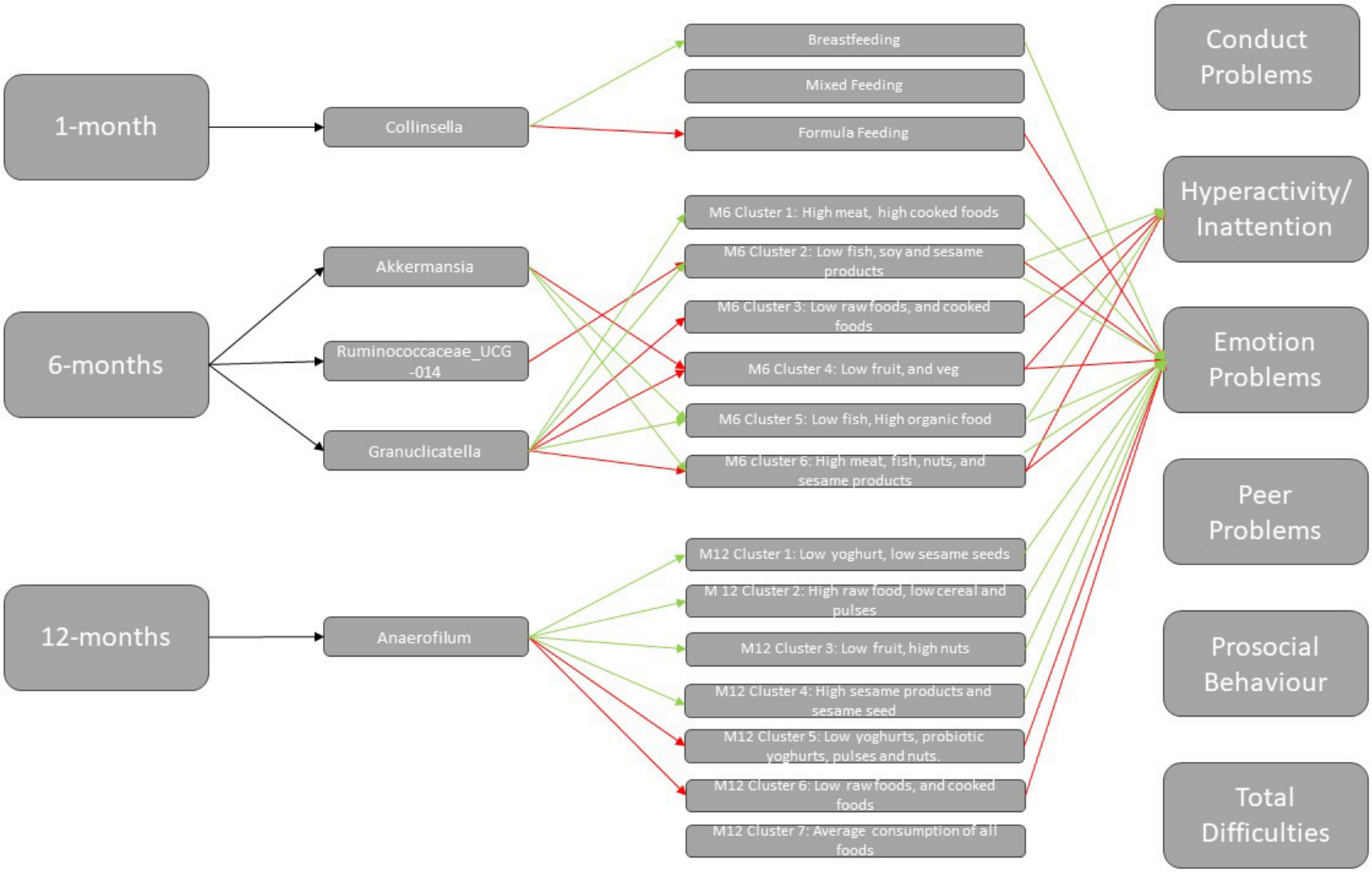


Figure 4.9 Synthesis figure illustrating the significant direct and indirect relationships between bacteria extracted at 1-, 6-, and 12 -months and SDQ outcomes measured at 4-years moderated by diet.
Note. Green lines denote direct positive relationships and red lines denote direct negative relationships.

of breastmilk to ascertain the extent of this influence. Maternal diet from pregnancy throughout the first year of life should also be considered as a confounder of this relationship. At 6-months of age, the relationship between relative abundance of *Akkermansia* and emotional problems measured at 4-years was significantly moderated by diet, specifically clusters 1, 5 and 6. Clusters 1 and 6, showed the most significant influence having a strongly positive association between higher relative abundance of *Akkermansia* and emotional problems. These two clusters are both characterised by increased meat consumption, with higher-than-average fish consumption also present in cluster 6. Interestingly Cluster 5, which also showed a positive association between *Akkermansia,* and emotional problem score is characterised by lower-than-average frequency of fish consumption. The relative abundance of *Akkermansia* is typically associated with dietary polyphenols resulting from consumption of variety of foods including nuts, berries, vegetables, olive, and olive oil (Verhoog et al., 2019). Given that these food types are typically associated with a healthy diet it would then follow that consumption would result in colonisation of bacteria that are beneficial to health and development. However, the results presented in this study do not align with this assumption and might actually indicate that increased relative abundance of *Akkermansia* may predict *more* emotional problems. This result may arise for one of two reasons, either *Akkermansia* does not have beneficial properties as part of the GM, or that increased relative abundance of this bacteria is not beneficial at 6-months of age suggesting that timing of colonisation is also imperative to the impact that GM has upon behavioural development. Of the three clusters, only cluster 6 provides a clear explanation for increased levels of *Akkermansia,* with individuals in this cluster consuming higher than average frequency of nuts and sesame products. Both clusters 1 and 6 are characterised by slightly higher consumption of cooked foods, however, it is not possible to ascertain how these meals are prepared, whether there is increased use of olive oil for example, which again would provide an explanation for the increase in relative abundance of *Akkermansia*. Furthermore, upon closer inspection of the diet clusters it was seen that cluster 1 was made up of only 2 individuals, however, in a follow up analysis that took into consideration the small cluster sizes, cluster 1 was removed, but the significant relationship remained, presenting the same positive associations with clusters 5 and 6 above (now labelled 4 and 5). It would appear that the original significant moderation relationships were therefore not driven by the very small cluster, and the overall model can be trusted. Removal of the small cluster does however alleviate doubt that is caused by clusters of n=2. One of the limitations of investigating diet through use of a food frequency questionnaire (FFQ), and, in particular, a FFQ that is not specifically designed with effect of diet on GM in mind is the lack of depth in understanding all of the elements relative to microbiota composition. Further research should explore a more in-depth investigation of diet by using, for instance, a diary technique. By using a diet diary, it is possible to record more specifically the food quantities consumed, to parse out more specifically consumption of different food types such as red and white meat, soluble and insoluble fibre types, and to understand

the influence of preparation techniques fully, and to eventually look further to align diet measures with metabolites and metabolic processes of the microbiome.

Relative abundance of an unclassified genus of the family Ruminococcaceae, measured at 6-months of age, and its association with emotional problems was significantly moderated by dietary cluster 2. This cluster, which was the largest cluster at this time point, is characterised by lower-than-average fish consumption, soy products, and sesame products, and slightly higher than average raw food consumption. The moderation effect drives a negative association between relative abundance of the unclassified genus of the family Ruminococcaceae, and emotional problems. In those individuals within this dietary cluster, higher relative abundance of the unclassified genus of the family Ruminococcaceae at 6-months has a protective influence upon the development of emotional problems at 4-years. Further analyses performed to adjust for small cluster sizes, did not influence the significant results that were found at 6 months of age. Although cluster 1 was no longer included in the analyses due to being a cluster of 2 individuals only, this did not significantly alter the findings, the results that were previously significant with the inclusion of cluster 1, remained significant when it was removed from the analysis.

The final dietary moderation of the relationship between GM and emotional problems relates to the relative abundance of *Anaerofilum* measured at 12-months. There is significant moderation influenced strongly by clusters 1, 2, 3, 4 and 6. Cluster 3, characterised by lower-than-average fruit and vegetable consumption, but higher than average nut frequency shows the strongest positive relationship. Clusters 2 and 4 are similar in their effect, however they show very different dietary characteristics, with cluster 2 having higher than average raw food but lower than average pulse frequency and cluster 4 having higher than average pulse frequency. Cluster 1 also shows a positive association, with lower-than-average yoghurt intake and lower than average sesame seed product consumption. Participants whose diet clustered into cluster 6, which was characterised by lower-than-average raw food consumption, show a strong negative relationship between relative abundance of *Anaerofilum* and emotional problems measured at 4-years. When viewing the patterns of results in clusters 2, higher than average raw food consumption and positive relationship between relative abundance of *Anaerofilum* and emotional problem scores, and cluster 6 with lower than average raw food consumption showing a negative relationship, it is tempting to say that there is a pattern forming, which would suggest that increased consumption of raw food negatively impacts emotional problems in those individuals with elevated relative abundance of *Anaerofilum*. This relationship between relative abundance of *Anaerofilum,* and emotional problems did not remain significant following adjustment for the small sample sizes in dietary clusters, which highlights the importance of evaluating the process of determining the number of clusters and determining when the cluster sizes are too small. Although this relationship is no longer significant there was a significant moderation of the relationship between *Eubacterium* and conduct problems and also an unclassified group of

Ruminococcaceae. The relationship between *Eubacterium* and Conduct problems was primarily driven by cluster 4 (previously cluster 6), characterised by lower-than-average raw food consumption, which now showed a positive association, and by cluster 1, which now shows a negative association, and is characterised by lower-than-average yoghurt intake and lower- than-average sesame seed product consumption. The relationship between an unclassified group of Ruminococcaceae measured at 12-months and conduct problems measured at 4-years was positively associated with diet clusters 1, 3, and 5, and negatively associated with cluster 2 (previously cluster 4). Clusters 1, and 2, as described previously, were strong drivers of this relationship. Finally, there were no significant moderation effects of the relationship between any of the scores on the SDQ subscales or total difficulties with the compositional balance parameters identified by the SELBAL method.

This study does show several strengths including novel use of moderation analyses to explore the relationship between early dietary intake, the whole gut microbiota, and the influence upon behavioural outcomes. Use of robust techniques such as the SELBAL method to identify and extract relative abundance of bacteria of interest with respect to the SDQ outcomes has resulted in several novel findings with regards to the direct relationship between composition of GM and behavioural outcomes. Previously unexplored associations such as the influence of *Gemella* at 6-months of age on both conduct and hyperactivity/inattention subscales have come to light. There are, however, limitations that are present within this study, as mentioned previously in Chapter 3, there is insufficient detail to establish all dietary measures that can influence GM composition, A potential way to overcome this limitation is to record the dietary intake in the form of a diet diary. This method has been trialled in the Alberta Pregnancy Outcomes and Nutrition (APrON) study and has successfully used a diary method to establish dietary patterns (Kaplan et al., 2014). It is also necessary to be cautious with results that are derived from compositional data with relatively low number of microbes, and also from cluster data with small sample sizes. However, in the present study effort was made to correct for the small cluster sample sizes by rerunning the analyses reducing the number of dietary clusters. Additionally, this study would benefit from concurrent behavioural and microbiota measures taken at 4-years of life. This would potentially bring to light new associations with the microbiota at a point after maturation of the GM, which occurs at approximately 41-months, or would strengthen the results that are currently presented by confirming it is the period of change that influences behaviour. Finally, the behavioural outcomes mapped here are the continuous scores measured using each subscale of the SDQ, and not the clinical predictors of behavioural problems designed by the SDQ. It is possible that when investigating caseness, using clinical cut offs for this same scale that different patterns of significant bacteria may emerge.

The results of this study present findings that show significant direct associations between composition of gut microbiota measured at 1-, 6-, and 12- months of age and SDQ outcomes measured at 4-years of age. A first important finding is the two patterns that emerge, after sensitivity

analyses showing that relative abundance of *Tyzzerella_4* measured at 6-months is associated with difficulties measured on the peer problems subscales. Secondly, increased relative abundance of *Gemella* is associated with decreased hyperactivity/inattention problems. Finally, through use of moderation analyses further evidence of the influence of dietary intake upon the relationship between GM and SDQ outcomes was established. In particular, it can be seen that some genera of bacteria, such as *Collinsella*, exert their effects upon behavioural outcomes, specifically the emotional problem subscale of the SDQ measured at 4-years of age, dependent on diet, whereas some, such as *Gemella*, have a more direct effect. However, in order to identify why these patterns are emerging it is necessary to have a greater understanding of the metabolic mechanisms at play. This can be achieved through functional analysis of the microbiome, conducted using whole genome sequencing. Future research should explore in greater depth dietary intake, establish ways of conducting longitudinal analyses of these moderating effects of diet upon the relationship between GM composition and behavioural outcomes in order to reflect the maturation process, and further explore the potential of identifying bacteria that lead to clinical cases.

CHAPTER 5
PREDICTION OF CLINICALLY SIGNIFICANT BEHAVIOURAL OUTCOMES BY COMPOSITION OF GUT MICROBIOTA

ABSTRACT

Background: It has been established previously that there is a relationship between gut microbiota (GM) composition and behavioural outcomes. However, it is possible that different patterns may emerge when limited to determining clinical caseness. The current study aims to identify clinical indicators of behavioural abnormality and the GM markers that predict increased likelihood of these behaviour outcomes.

Method: A total of 1074 infants were recruited into the Barwon Infant Study of Geelong Australia. Of these 324 participants were recruited into a microbiota subset. Logistic regression analyses were performed to examine the likelihood that relative abundance of identified gut bacteria, measured at 1-, 6-, and 12-months would predict caseness on the SDQ measured at 4-years.

Results: Several significant direct relationships were demonstrated between GM and SDQ caseness for each time point. Relative abundance of *Tyzzerella_4,* and *Dorea* measured at 6-months of age were confirmed to predict clinically significant scores on the internalising and externalising problems scale of the SDQ at 4-years respectively.

Discussion: The composition of the GM in infants is of importance in predicting subsequent clinically significant behavioural outcomes. Future research should explore underlying mechanisms using whole genome sequencing to explore bacterial metabolic processes that contribute to increased likelihood of clinical cases.

5.1 Introduction

As established in both Chapters 3 and 4, it is possible to identify bacteria from the composition of the whole gut microbiota that are associated with each of the scales of the Strengths and Difficulties questionnaire (SDQ) (Goodman, 1997). To further our understanding of the relationship between GM composition and behaviour, the next step is to investigate the potential clinically significant impact that these identified bacteria have on behavioural development.

In the UK alone approximately 10% of children develop mental health disorders that result in distress and further impairments (Meltzer, Gatward, Goodman, & Ford, 2003). These disorders often persist into adolescence and into adulthood. Additionally, between 13-20% show a developmental trajectory that may lead to mental health disorders later in life (Avenevoli et al., 2013; Ghandour et al., 2019). It is estimated that approximately 24.5% of the adult population of the UK experiences mental health problems, although in recent years this figure has been shown to have increased due to the COVID-19 pandemic. However, the long-term effect of this is yet to be fully investigated (Daly, Sutin, & Robinson, 2022). Current research has shown that there is a strong link between mental health disorders and the composition and diversity of the gut microbiota in adults. A meta-analysis carried out by Nikolova et al. (2021) investigated compositional alterations in the GM of adults. They found that there are decreased levels of *Faecalibacterium* and *Coprococcus* and increased levels of *Eggerthella,* in those diagnosed with depression, anxiety, bipolar disorder, psychosis, and schizophrenia. This review concluded that there is a reduction of butyrate producing bacteria, which has also been shown to be an important bacteria in children relating to temperament (Alving-Jessep, Botchway, Wood, Hilton, & Blissett, 2022). These bacteria are responsible for anti-inflammatory actions within the gut. A further review by Borkent, Ioannou, Laman, Haarman, and Sommer (2022) found very similar results showing that in patients with mental health disorders there is a reduced relative abundance of butyrate producing bacteria, Additionally, it was found that in patients diagnosed with mental health disorders there were higher levels of *Streptococcus, Lactobacillus* and *Eggerthella* that was consistent across the studies included in the review. What is unclear from these reviews is whether the relationships between relative abundance of bacteria and mental health disorder are causal, or whether environmental factors such as poor diet, leads to altered GM following diagnosis. Given evidence of the link between early behaviour problems in childhood that lead to later mental health disorders, it may be possible to increase the understanding of the causal relationship between GM, and the development of mental health disorders through investigation of the childhood period.

In chapters 3 and 4, significant direct relationships between GM composition, and behavioural outcomes during early childhood were established, utilising continuous scores of the SDQ subscales and total problems scores. However, investigation of clinically significant scores, using the SDQ established cut off points, may establish a different pattern of significance between GM composition

and behavioural outcome. The continuous scores are important as they establish GM measures that increase problem behaviours, but also indicate those that may have a protective effect. However, they do not provide information on those scores that reach the threshold of clinical significance. The SDQ was also developed to predict those whose behaviour measure as clinically significant for developing clinically significant mental health and behavioural problems. This was achieved through establishing cut off points, that were developed and validated from investigating large populations of children within the UK, which based on a large study of approximately 10% of the population. Through investigating whether the composition of the GM may predict clinically significant scores based on these cut off points it is possible to further refine the current understanding of the composition of the GM that not only indicates increased behavioural problems, but those that are clinically significant and predictive of future mental health disorders.

Previous investigation of the same cohort found that there was a potential association between decreased relative abundance of the bacteria *Prevotella* in samples collected at 12-months of age and elevated risk of internalising problem scores measured on the Child Behaviour Checklist (CBCL) at 2-years of age (Loughman et al., 2020). Further recent investigations have found that internalising problems, such as those that lead to the development of severe depressive symptoms, measured at age 10 on the SDQ, were associated with decreased relative abundance of the bacteria *Akkermansia* measured from microbiota samples taken from children aged 6 and 10 (Ou et al., 2022). This study further found that both *Phascolarctobacterium and Prevotella_9*, measured at 6 and 10-years, were positively associated with maternal reports of externalising behaviour at age 10 (Ou et al., 2022). Internalising behaviour problems were also found to be negatively correlated with alpha diversity, measured as Shannon Diversity Index (SDI), in children aged 3-5 years of age (van de Wouw et al., 2022).

Childhood, specifically the period from birth to 41-months of life is the period of maturation of the gut microbiome, in which the composition and diversity of species within the gut undergo the greatest changes. This period also aligns with periods of sensitivity for brain development, a period where alterations in the composition of the GM may have the greatest influence, through the gut-brain axis (Cowan et al., 2020). For these reasons there has been an increase in focus upon this period in the development of both gut and behaviour. It is therefore the aim of this study to explore the relationship between clinical indicators of behavioural abnormality, measured on the SDQ at 4-years, and the GM markers that predict increased likelihood of these behaviour outcomes.

Previous investigation of the same cohort found that there was a potential association between decreased relative abundance of the bacteria *Prevotella* in samples collected at 12-months of age and elevated risk of internalising problem scores measured on the Child Behaviour Checklist (CBCL) at 2-years of age (Loughman et al., 2020). Further recent investigations have found that internalising

problems, such as those that lead to the development of severe depressive symptoms, measured at age 10 on the SDQ, were associated with decreased relative abundance of the bacteria *Akkermansia* measured from microbiota samples taken from children aged 6 and 10 . This study further found that both *Phascolarctobacterium and Prevotella_9*, measured at 6 and 10-years, were positively associated with maternal reports of externalising behaviour at age 10. Internalising behaviour problems were also found to be negatively correlated with alpha diversity, measured as Shannon Diversity Index (SDI), in children aged 3-5 years of age.

Childhood, specifically the period from birth to 41-months of life is the period of maturation of the gut microbiome, in which the composition and diversity of species within the gut undergo the greatest changes (Stewart et al., 2018b). This period also aligns with periods of sensitivity for brain development, a period where alterations in the composition of the GM may have the greatest influence, through the gut-brain axis. For these reasons there has been an increase in focus upon this period in the development of both gut and behaviour. It is therefore the aim of this study to explore the relationship between clinical indicators of behavioural abnormality, measured on the SDQ at 4-years, and the GM markers that predict increased likelihood of these behaviour outcomes.

5.1.1 Aims and Hypotheses

The aims of the study and corresponding hypotheses are as follows:

Aim:

1. To investigate whether composition of the gut microbiota measured at 1-, 6-, and 12-months of age predicts the risk of clinically significant behavioural scores, measured as internalising problems, externalising problems, and total difficulties on the SDQ at 4-years of age.

Hypothesis:

1. The composition of the GM will significantly predict risk of clinically significant scores on the scales of internalising problems, externalising problems, and total difficulties measured at 4-years of age on the SDQ.

5.2 Methods

5.2.1 Study design and participants

Three hundred and twenty-four children, from a cohort of 1074, were recruited into a randomly recruited microbiota subset. This subset consisted of 321 families, as three pairs of twins were

included. Data was collected from both mothers and infants (see Chapter 3, section 3.2.1, for more information on study design and participant recruitment).

5.2.2 Ethics

Consent was sought from the mothers of all infants and was given in written form prior to participation in the study. The ethics committee at Barwon Health approved the study prior to commencement (reference 10/24).

5.2.3 Measurements

For the microbiota, stool samples were collected from 324 infants in the microbiota subset at 12-months1-, 6-, and 12-months of age. Some of the samples were collected during the review appointment as fresh samples, and others were collected at home and brought to the laboratory. Of the samples that were brought into the laboratory frozen, 95% of these were received within 20 days. For an in-depth explanation of the microbiota pipeline, extraction, and preparation to Amplicon sequence variant (ASV) level please see chapter 3, section 3.2.3.

In order to determine risk of clinically significant behaviour the SDQ was completed by mothers when the child was 4-years of age (Goodman, 1997). Scores were determined for each of the subscales measuring the domains of hyperactivity/inattention, conduct problems, peer relationship problems, emotional symptoms, and prosocial behaviour were completed using the complete 25 item questionnaire (See Appendix 1, and chapter 3, section 3.2.4 for further description). Furthermore, scores were determined for risk of Externalising behaviour problems, risk of internalising behaviour problems, and total difficulties, as per the standard scoring for the SDQ (See Appendix 6 for gull scoring guide). In a study conducted by Goodman, Lamping, and Ploubidis (2010), it was found that in low-risk samples it is advantageous to use the broader internalising, and externalising scales to measure clinical cases. As recruitment was not carried out selectively to increase numbers of those at risk, the relative number of cases is likely to be low, and therefore the broader scales will be used rather than all five sub-scales. The cut-points of the SDQ were determined based on a UK community sample, which generates four categories of close to average, slightly raised, abnormal high, and abnormal very high for each of the scales. These cut-points represent 80%, 10%, 5% and 5% respectively have been validated for use in children under the age of 18 (Goodman & Goodman, 2009).

5.2.4 Data analysis

Data preparation and analysis was performed using the software R (Team, 2018). Preparation of the data including quality assessment, removal of duplicates and sub setting was carried out in the *phyloseq* package (McMurdie & Holmes, 2013).

Hypothesis 1: In order to examine whether it is possible to predict clinically significant scores that are associated with clinical cases, bacteria that were extracted using the SELBAL method in Chapter 4, for each of the subscales of SDQ were examined. Furthermore, the scope of extraction was extended to include internalising and externalising scales for each of the three time points of 1-, 6-, and 12-months of microbiota samples. To establish the predictive relationship, logistic regressions for each of the extracted bacteria and related SDQ scores were performed. These regression analyses were also adjusted for potential covariates.

The covariates included in this study were previously identified using directed acyclic graphs (DAGs) using an online programme DAGitty (Textor et al., 2016) to determine suitable adjustment sets for each analysis (See Chapter 3 section 3.2.7, and Appendix 2 for further details).

5.2.5 Transformations

Microbiota data was prepared for differential abundance testing, taking into account zero-inflation, over dispersion and its compositionality. Zero counts were addressed by adding a pseudo count of one as primary analysis. A centred-log-ratio was generated after normalisation by sequencing, this allows for processing of 16S rRNA data in linear models.

5.3 Results

5.3.1 Descriptive statistics of study sample

Complete descriptive statistics of this study can be found in chapter 3, section 3.3.1, and Chapter 3 Table 1. Of the 324 infants randomly selected, a total of 289 microbiota samples were available for one-month-olds following quality control, 298 for six-month-olds, and 251 for 12-month-olds. Using the standard scoring system that accompanies the SDQ cases were classified as those scoring in the “abnormal high” or “abnormal very high” range (Goodman & Goodman, 2009), which is indicative of those presenting with patterns of clinically significant behavioural problems. These two classifications represent scores found in approximately 10% of the total population (See Appendix 6 for score ranges). Of the 289 participants with complete samples at one-month of age, 13 scored as internalising problems cases, 19 scored as externalising problem cases, and 12 scored as cases for total difficulties. These numbers were the same for the 298 participants included at 6-months of age. For those with complete samples at 12-months there were 12 internalising cases, 14 externalising cases, and 8 total difficulties cases. When inspecting the participants that presented with clinically significant scores at 4-years, from the group with complete samples at each time point it was observed that for the Total difficulties, and externalising problems scales, the same participants were identified across all time points. For the internalising problems cases the participants were predominantly the

same, however there was one additional child that presented at 6-, and 12-months that was not present in the 1-month group and one participant that was present only in the 6-month group.

5.4 GM bacteria extracted relating to SDQ outcomes using the SELBAL method.

Bacteria of interest were extracted using the SELBAL method as described in Chapter 4, section 4.3.2. The bacteria of interest were extracted for each of the subscales of the SDQ, and Total difficulties and can be seen for each time point in Chapter 4, Table 1. Extraction was further extended to those bacteria of interest relating to internalising scores and externalising scores for each time point (See Table 5.1).

5.5 Prediction of cases, abnormal SDQ scores by extracted bacteria, hypothesis 1.

To determine whether each of the extracted bacteria significantly predicted cases of total difficulties, and internalising and externalising problems, multiple logistic regressions were performed. At one month of age, increased relative abundance of the *Pseudocitrobacter* was associated with clinically significant scores on the internalising problems scale of the SDQ at 4-years of age (OR: 575.50 [-0.338 - -0.071], $p = 0.041$), when adjusted for number of siblings. Similar patterns are seen for increased relative abundance of *Collinsella* (OR: 88.6 [-0.325 - -0.069], $p = 0.025$) and risk of clinically significant scores on the internalising problems scale. Finally, increased relative abundance of *Veillonella* at one month of age was associated with clinically significant scores on the externalising problems scale at 4-years (OR: 3.59 [-0.416 - -0.155], $p = 0.043$) (See Table 5.2 and Figure 5.1 for results).

At 6-months of age increased relative abundance of *Tyzzerella_4* was associated with clinically significant scores on the internalising problems scale at 4-years of age (OR: 177.10 [-0.350 - -0.081], $p = 0.004$), and increased relative abundance of *Dorea* was associated with elevated risk of externalising problems at 4-years of age (OR: 136.20 [-0.363 - -0.176], $p = 0.010$) (See Table 5.3 and Figure 2 for results). Finally, at 12-months of age increased relative abundance of *Staphylococcus* was associated with increased risk of clinically significant scores on the total difficulties scale (OR: 3803.00 [-0.246 - -0.085], $p = 0.019$). Clinically significant scores on the internalising problems scale at 4-years was also associated with increased relative abundance of both *Catenibacterium* (OR: 151.40 [-0.284 - -0.034], $p = 0.004$), and *Turicibacter* (OR: 126.70 [-0.334 - -0.080], $p = 0.016$). At both 6 and 12-months of age there were no significant confounders (See table 5.4 and Figure 5.3 for results).

Further analysis of the SELBAL balance parameters shows that there were significant associations between the balance parameter identified at month 6 and internalising problems, showing that increases in this balance parameter were related to increased risk of internalising problems (OR:

0.595 [-0.364 - -0.106], $p < .05$). Similarly increases in balance parameter measures at month 12 were associated with elevated risk of clinically significant scores of total difficulties at 4 years of age, (OR: 1.181 [-0.281 - -0.056], $p < .05$) (See Table 5.5 for results).

5.6 Discussion

It was previously demonstrated, in chapters 3 and 4, that it is possible to ascertain significant relationships between both composition and diversity of the GM and behavioural outcomes measured on SDQ subscales. Identification of these relationships is important in understanding the underlying mechanisms between GM and behaviour. Furthermore, it has been established that these relationships can be moderated by diet. However, it is not possible from investigation of the continuous scores on the SDQ whether these GM factors are predictive of clinical caseness. This study furthers the understanding of the relationship between GM, behaviour, and the role of early childhood dietary patterns by establishing GM indicators of clinical caseness, measured on the SDQ, and whether this relationship can be moderated by diet.

It was hypothesised that bacteria identified using the SELBAL method measured at 1-, 6-, and 12-months of age would significantly predict abnormal behavioural scores, which can be classified as cases, on the scales of internalising problems, externalising problems, and total difficulties measured at 4-years of age on the SDQ. From these results it can be seen that at one-month of age increased relative abundance of the genus *Pseudocitrobacter,* was associated with increased risk of clinically significant internalising problems, and *Collinsella, and Veillonella,* were all associated with increased risk of clinically significant externalising problems at 4-years of age. The finding relating relative abundance of *Pseudocitrobacter* with increased risk of clinically significant internalising problems is a novel one. In previous chapters there has been neither direct nor indirect relationships between this bacterium and behaviour at any time point. Likewise, there is no previous literature that has previously shown a link between this bacterium and behaviour. A potential explanation for this is that the classification of the genus *Pseudocitrobacter* is relatively recent. This bacterium is a Gram-negative facultative anaerobe, which was previously classified as *Citrobacter* species phenotypically, Since the availability of 16S rRNA sequencing it has been found that this specific bacterium shares only 97% similarity with *Citrobacter* and would cluster into it on phylogenetic tree during identification, resulting in the reclassification. Characteristically this bacterium behaves in a similar manner to *Citrobacter* (Kämpfer, Glaeser, Raza, Abbasi, & Perry, 2014). The novelty of this result, therefore, is one that warrants further investigation and repetition in order to establish its importance as a genus of interest in GM behaviour investigations.

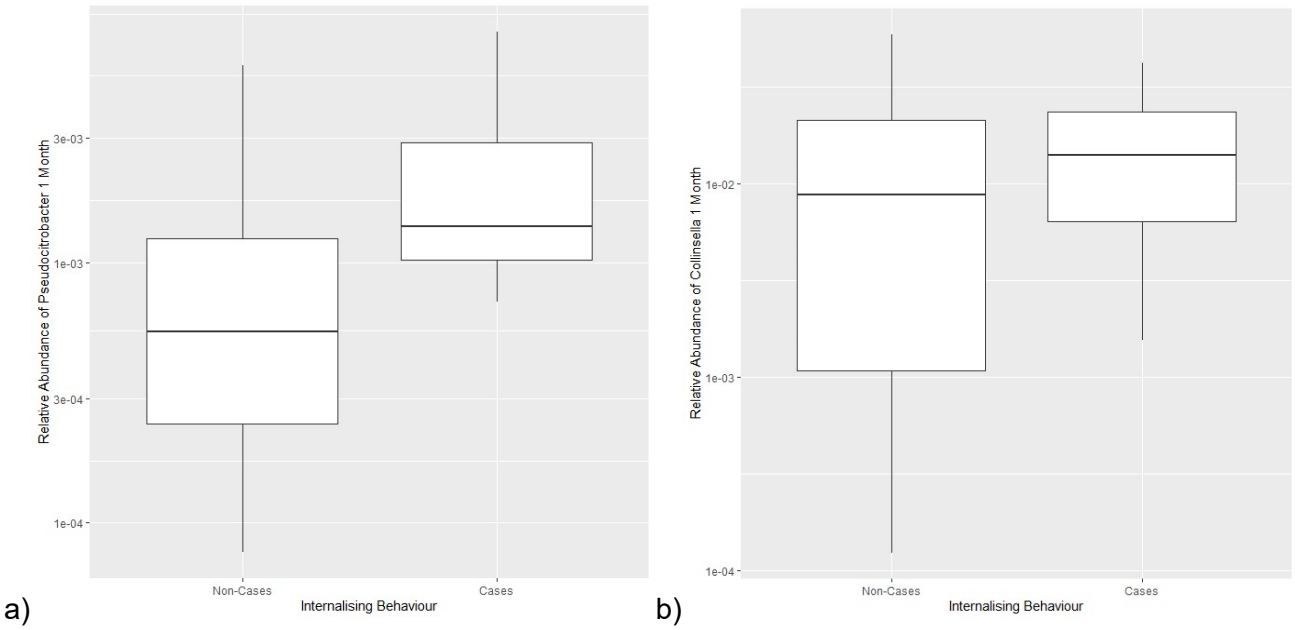
Table 5.1. Bacteria extracted for each Internalising and Externalising problems outcome using SELBAL method.

Month	SDQ	ASV	Family	Microbiota Genus	Numerator/ Denominator
1	Internalising	ASV00259	Micrococcaceae	*Rothia*	Den
1	Internalising	ASV00041	Lachnospiraceae	*Lachnoclostridium*	Den
1	Internalising	ASV00768	Enterobacteriaceae	*Pseudocitrobacter*	Num
1	Internalising	ASV00014	Streptococcaceae	*Streptococcus*	Den
1	Internalising	ASV00001	Enterobacteriaceae	*Escherichia/Shigella*	Den
1	Internalising	ASV00198	Peptostreptococcaceae	*Peptoclostridium*	Den
1	Internalising	ASV00051	Peptostreptococcaceae	*Intestinibacter*	Den
1	Internalising	ASV00472	Actinomycetaceae	*Varibaculum*	Den
1	Internalising	ASV00419	Enterobacteriaceae	*Salmonella*	Den
1	Externalising	ASV00014	Streptococcaceae	*Streptococcus*	Den
1	Externalising	ASV00086	Ruminococcaceae	*Flavonifractor*	Den
1	Externalising	ASV00001	Enterobacteriaceae	*Escherichia/Shigella*	Num
6	Internalising	ASV00085	Lachnospiraceae	*Tyzzerella_4*	Den
6	Internalising	ASV00333	Alcaligenaceae	*Sutterella*	Den
6	Internalising	ASV00314	Coriobacteriaceae	*Eggerthella*	Den
6	Internalising	ASV00007	Lachnospiraceae	*Blautia*	Den
6	Internalising	ASV00328	Lachnospiraceae	*Lachnoclostridium_5*	Den
6	Internalising	ASV00074	Ruminococcaceae	Ruminococcaceae_UCG-013	Den
6	Internalising	ASV00005	Ruminococcaceae	*Subdoligranulum*	Den
6	Internalising	ASV00106	Rikenellaceae	*Alistipes*	Den
6	Internalising	ASV00086	Ruminococcaceae	*Flavonifractor*	Den
6	Internalising	ASV00505	Enterobacteriaceae	*Raoultella*	Den
6	Internalising	Asv00263	Lachnospiraceae	*Sellimonas*	Den
6	Externalising	ASV00351	Family_XI	*Gemella*	Den
6	Externalising	ASV00034	Peptostreptococcaceae	*Romboutsia*	Num
6	Externalising	Asv00224	Erysipelotrichaceae	*Holdemanella*	Num
12	Internalising	ASV00378	Porphyromonadaceae	*Odoribacter*	Den
12	Internalising	ASV00437	Desulfovibrionaceae	*Desulfovibrio*	Num
12	Internalising	ASV00054	Prevotellaceae	*Prevotella_9*	Den
12	Externalising	ASV00056	Erysipelotrichaceae	Erysipelotrichaceae_UCG-003	Den
12	Externalising	ASV00026	Lachnospiraceae	*Anaerostipes*	Den
12	Externalising	ASV00273	Eubacteriaceae	*Eubacterium*	Num

Table 5.2. Logistic regression between relative abundance of bacteria measured at 1-month, and clinically significant SDQ outcomes measured at 4-years.

Microbiota Genus	SDQ	OR[a]	DF	Q1	Q2	P value
Escherichia/Shigella	Total Problem	-3.325	171	-0.325	-0.078	0.198
Flavonifractor	Total Problem	164.400	171	-0.312	-0.108	0.098
Streptococcus	Total Problem	4.432	171	-0.322	-0.128	0.441
Rothia	Internalising	-95.920	171	0.334	-0.079	0.553
Lachnoclostridium	Internalising	-0.785	238	-0.337	-0.337	0.758
Pseudocitrobacter	Internalising	575.500	171	-0.338	-0.071	0.041
Streptococcus	Internalising	-45.500	171	-0.333	-0.051	0.290
Escherichia/Shigella	Internalising	1.280	238	-0.332	-0.280	0.102
Peptoclostridium	Internalising	-50120.000	238	-0.339	-0.339	0.993
Intestinibacter	Internalising	74.430	171	-0.356	-0.088	0.462
Varibaculum	Internalising	-137500.000	171	-0.346	-0.047	0.994
Salmonella	Internalising	21.110	171	-0.353	-0.090	0.986
Staphylococcus	Internalising	5.235	171	-0.357	-0.089	0.929
Blautia	Internalising	10.390	171	-0.364	-0.091	0.777
Lactobacillus	Internalising	1.882	171	-0.366	-0.087	0.679
Collinsella	Internalising	88.600	171	-0.325	-0.069	0.025
Eubacterium	Internalising	-100500.000	238	-0.340	-0.340	0.994
Erysipelatoclostridium	Internalising	6.192	238	-0.315	-0.315	0.081
Hungatella	Internalising	-471.200	171	-0.343	-0.069	0.786
Citrobacter	Internalising	-2.028	171	-0.354	-0.091	0.907
Pluralibacter	Internalising	-131800.000	238	-0.346	-0.346	0.990
Subdoligranulum	Internalising	-8760.000	171	-0.357	-0.085	0.993
Peptoniphilus	Internalising	-154700.000	171	-0.348	-0.037	0.996
Epulopiscium	Internalising	-34800.000	171	-0.353	-0.090	0.997
Megasphaera	Internalising	-15830.000	171	-0.355	-0.081	0.995
Streptococcus	Externalising	2.088	171	-0.414	-0.200	0.639
Flavonifractor	Externalising	-434.000	171	-0.398	-0.131	0.374
Escherichia/Shigella	Externalising	-2.781	171	-0.423	-0.131	0.108
Intestinibacter	Externalising	123.700	171	-0.399	-0.205	0.187
Pseudocitrobacter	Externalising	-165.500	171	-0.409	-0.194	0.647
Enterococcus	Externalising	-34.090	171	-0.411	-0.193	0.554
Clostridium_sensu_stricto_1	Externalising	-17.760	171	-0.412	-0.180	0.564
Staphylococcus	Externalising	21.410	171	-0.428	-0.192	0.580
Rothia	Externalising	1.048	171	-0.427	-0.195	0.973
Veillonella	Externalising	3.593	171	-0.416	-0.155	0.043

[a] OR – Odds Ratio



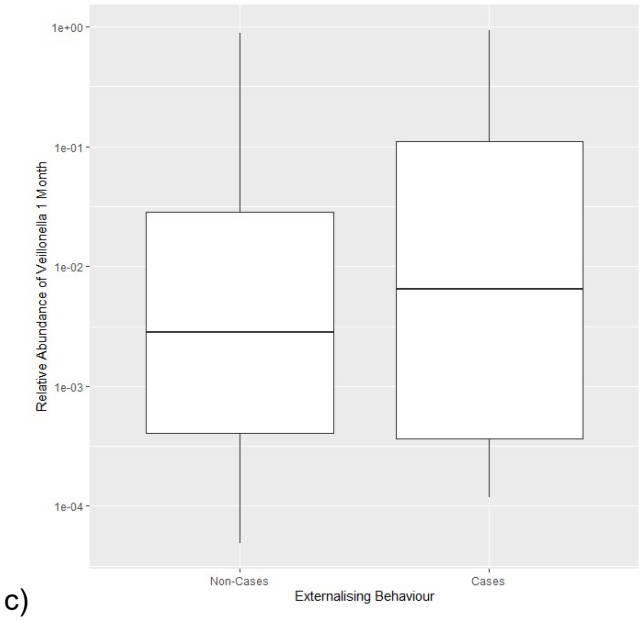


Figure 5.1. Graphs to show the significant difference in relative abundance of extracted bacteria between cases and non-cases at 4-years old, in a) *Pseudocitrobacter,* b) *Collinsella,* and c) *Veillonella* at 1-month.

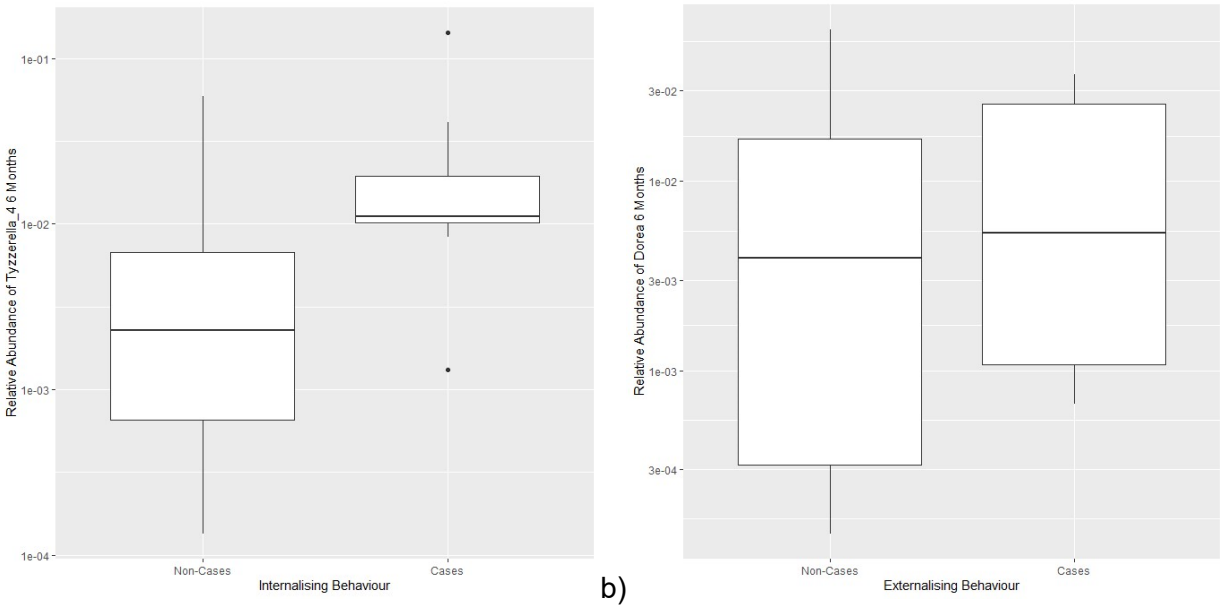
Table 5.3. Logistic regression between relative abundance of bacteria measured at 6-months, and clinically significant SDQ outcomes measured at 4-years.

Microbiota Genus	SDQ	OR[a]	DF	Q1	Q2	P value
Sutterella	Total Problem	-95180.000	248	-0.336	-0.336	0.994
Romboutsia	Total Problem	24.210	196	-0.279	0.000	0.215
Prevotella	Total Problem	-108000.000	196	-0.227	0.000	0.996
Tyzzerella_4	Internalising	177.100	196	-0.276	-0.081	0.004
Sutterella	Internalising	-95720.000	248	-0.350	-0.350	0.994
Eggerthella	Internalising	-578.900	196	-0.383	-0.108	0.412
Blautia	Internalising	0.169	196	-0.398	-0.122	0.966
Lachnoclostridium_5	Internalising	-181.000	196	-0.398	-0.120	0.775
Ruminococcaceae_UCG-013	Internalising	-96280.000	196	-0.387	-0.100	0.994
Subdoligranulum	Internalising	-285.600	196	-0.403	-0.116	0.763
Alistipes	Internalising	-24.250	196	-0.381	-0.116	0.645
Flavonifractor	Internalising	10.410	196	-0.388	-0.121	0.729
Raoultella	Internalising	2.269	196	-0.398	-0.123	0.986
Sellimonas	Internalising	-366.300	196	-0.395	-0.123	0.611
Akkermansia	Internalising	2.008	196	-0.380	-0.115	0.221
Fusobacterium	Internalising	-10600.000	196	-0.365	-0.059	0.994
Prevotella_9	Internalising	-134000.000	248	-0.337	-0.337	0.988
Ruminococcaceae_UCG-014	Internalising	-813.000	196	-0.394	-0.120	0.772
Desulfovibrio	Internalising	910.700	196	-0.302	-0.039	0.732
Roseburia	Internalising	16.910	196	-0.395	-0.122	0.671
Fusicatenibacter	Internalising	-24.730	196	-0.399	-0.124	0.812
Lachnospira	Internalising	-136.200	196	-0.396	-0.113	0.580
Ruminiclostridium_5	Internalising	-110.400	196	-0.393	-0.121	0.650
Streptococcus	Internalising	-98.180	196	-0.331	-0.013	0.097
Lactococcus	Internalising	-185.100	196	-0.404	-0.114	0.514
Lachnoclostridium	Internalising	1.208	196	-0.396	-0.122	0.819
Blautia	Internalising	0.169	196	-0.398	-0.122	0.966
Actinomyces	Internalising	-108.200	196	-0.399	-0.119	0.575
Peptoclostridium	Internalising	12.370	196	-0.397	-0.122	0.875
Enterobacter	Internalising	-3.572	248	-0.332	-0.328	0.727
Gemella	Externalising	-973.700	196	-0.388	-0.182	0.319
Romboutsia	Externalising	3.065	196	-0.393	-0.197	0.881
Holdemanella	Externalising	-4605.000	196	-0.385	-0.190	0.994
Dorea	Externalising	136.200	196	-0.363	-0.176	0.010
Actinomyces	Externalising	24.816	196	0.382	-0.196	0.744
Granulicatella	Externalising	138.800	196	-0.392	-0.193	0.742

Table 5.3 Continued

Microbiota Genus	SDQ	OR[a]	DF	Q1	Q2	P value
Hungatella	Externalising	-3.381	196	-0.388	-0.192	0.873
Prevotella	Externalising	-1082.000	196	-0.406	-0.180	0.418
Subdoligranulum	Externalising	-95510.000	248	-0.412	-0.412	0.991

[a] OR – Odds Ratio



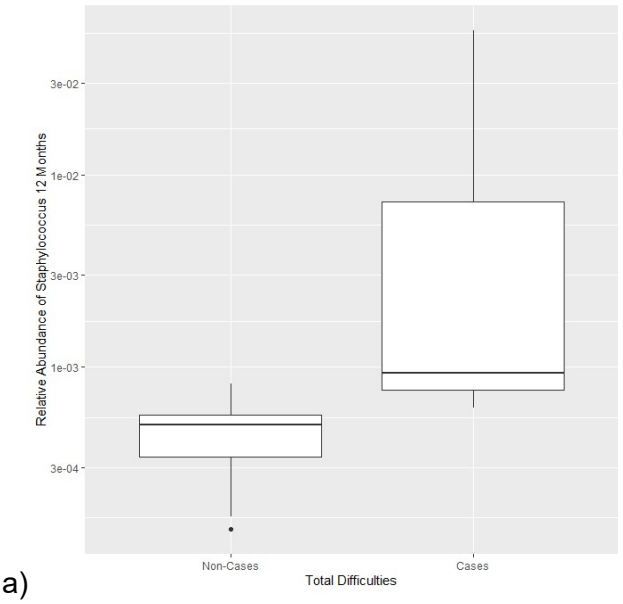
a) b)

Figure 5.2. Graphs to show the difference in relative abundance of extracted bacteria between cases and non-cases at 4-years old, in a) *Tyzzerella_4, and* b) *Dorea* at 6-months.

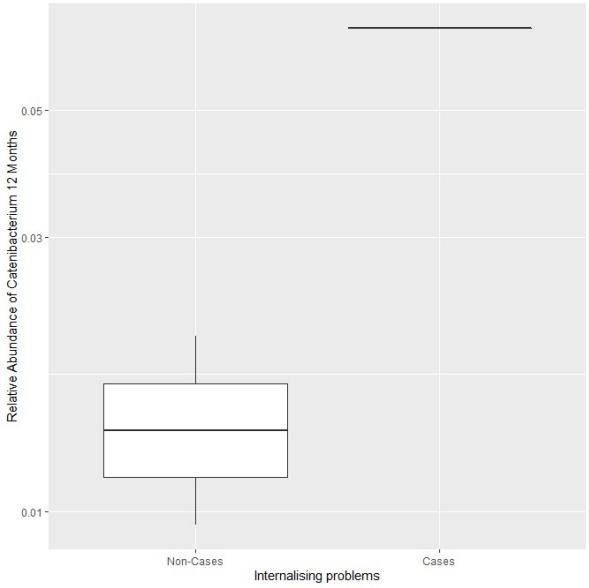
Table 5.4. Logistic regression between relative abundance of bacteria measured at 12-months, and clinically significant SDQ outcomes measured at 4-years.

Microbiota Genus	SDQ	OR[a]	DF	Q1	Q2	P value
Eubacterium	Total Problem	291.300	194	-0.314	-0.123	0.264
Staphylococcus	Total Problem	3803.000	194	-0.246	-0.085	0.019
Odoribacter	Total Problem	-61190.000	194	-0.324	0.000	0.995
Odoribacter	Internalising	-55610.000	194	-0.374	0.000	0.995
Desulfovibrio	Internalising	724.500	194	-0.370	-0.069	0.070
Prevotella_9	Internalising	-82790.000	214	-0.376	-0.376	0.981
Eubacterium	Internalising	-94430.000	194	-0.364	0.000	0.995
Alistipes	Internalising	-49.080	194	-0.364	-0.086	0.590
Lachnospira	Internalising	-62.310	194	-0.350	-0.081	0.441
Ruminococcaceae_UCG-004	Internalising	73.400	194	-0.367	-0.085	0.732
Subdoligranulum	Internalising	-3.301	194	-0.386	-0.087	0.579
Dialister	Internalising	-1.533	194	-0.368	-0.087	0.833
Tyzzerella_3	Internalising	-180600.000	194	-0.361	0.000	0.994
Lachnospiraceae_FCS020_group	Internalising	-129.800	194	-0.375	-0.085	0.708
Catenibacterium	Internalising	151.400	194	-0.284	-0.034	0.004
Flavonifractor	Internalising	17.910	194	-0.373	-0.089	0.777
Anaerofilum	Internalising	-52.510	194	-0.364	-0.083	0.559
Ruminococcus_1	Internalising	28.930	194	-0.368	-0.087	0.729
Megamonas	Internalising	-89100.000	214	-0.370	-0.370	0.983
Turicibacter	Internalising	126.700	194	-0.334	-0.080	0.016
Tyzzerella_4	Internalising	16.480	194	-0.383	-0.082	0.456
Erysipelotrichaceae_UCG-003	Externalising	-927.000	194	-0.361	0.000	0.497
Anaerostipes	Externalising	-7.863	194	-0.349	-0.167	0.492
Eubacterium	Externalising	350.000	194	-0.336	-0.156	0.089
Coprococcus_2	Externalising	-4.181	194	-0.347	-0.164	0.958
Parasutterella	Externalising	-51.760	194	-0.349	-0.163	0.661
Staphylococcus	Externalising	2696.000	194	-0.327	-0.155	0.130
Lachnospiraceae_UCG-004	Externalising	-1983.000	194	-0.343	-0.113	0.238
Ruminococcaceae_UCG-013	Externalising	1.106	194	-0.348	-0.162	0.973

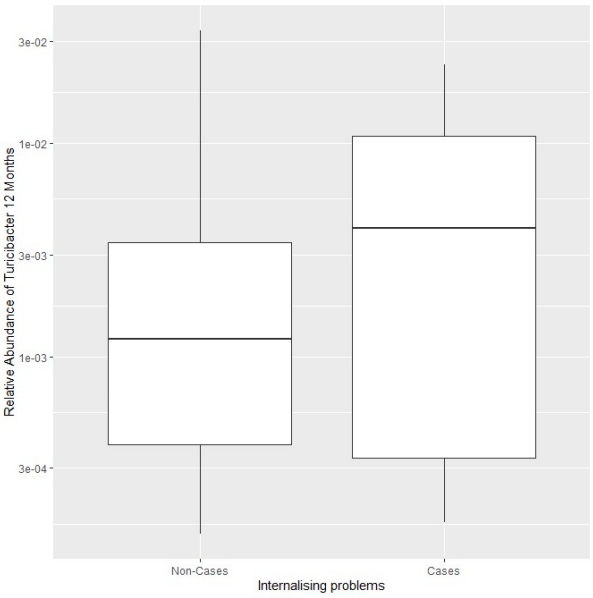
[a] OR – Odds Ratio



a)



b)



c)

Figure 5.3. Graphs to show the difference in relative abundance of extracted bacteria between cases and non-cases at 4-years of age, in a) *Staphylococcus,* b) *Catenibacterium,* and c) *Turicibacter* at 12-months,

Table 5.5. Logistic regression between SELBAL balance parameters measured at 1-, 6-, and 12-months, and clinically significant SDQ outcomes measured at 4-years.

Month	SDQ Scale	OR[a]	DF	Q1	Q2	P value
1	Internalising Problems	0.351	171	-0.331	-0.089	0.302
1	Externalising Problems	-0.169	171	-0.426	-0.172	0.150
1	Total Difficulties	-0.086	171	-0.328	-0.127	0.735
6	Internalising Problems	0.595	196	-0.364	-0.106	0.024
6	Externalising Problems	0.030	196	-0.397	-0.141	0.907
6	Total Difficulties	0.608	196	-0.270	-0.005	0.205
12	Internalising Problems	0.527	183	-0.374	-0.069	0.148
12	Externalising Problems	0.407	183	-0.363	-0.133	0.054
12	Total Difficulties	1.181	183	-0.281	-0.056	0.012

[a] OR – Odds Ratio

Previously, in chapter 4, it was established that *Collinsella*, a bacterium that is associated with exclusively breastfed infants whose mothers consume a low-fat diet, measured at one month of age, was significantly associated with emotional problems subscale, which is associated with internalising problems rather than externalising. This difference in result is indicative that the patterns that can be seen when investigating the continuous SDQ outcome scales are not necessarily the patterns that will emerge when investigating clinical caseness. This may suggest that although higher relative abundance of *Collinsella* at 1-month may be disruptive it may not be in a homogenous way. Further investigation is warranted, which would benefit from larger participant group with greater number of cases. The result of increased relative abundance of *Veillonella* predicting increased risk of externalising problems*,* aligns with previous findings from chapter 4 that showed and association with the Hyperactivity/inattention subscale. The bacterium *Veillonella* is part of the commensal flora of both the mouth and gastrointestinal (GI) tract and is known for the fermentation of lactate. A recent study investigating the GM in preschool children found that relative abundance of *Veilonella* was significantly correlated with internalizing problems (van de Wouw et al., 2022), which somewhat contradicts the finding of this study. However, the microbiota samples were taken between 3 and 5 years of age, compared to 1-month of age in this study. These results lend more weight to the notion that timing of colonisation is vitally important, and that over colonisation by bacteria, that would otherwise be part of the normal flora of the gut, at sensitive periods can lead to increased risk of clinically significant behavioural problem. Furthermore, this emphasises the need for longitudinal analyses of the relationship between microbiota and behaviour.

At 6-months of age a strong pattern of association that was previously established in chapter 4 was found. Increased relative abundance of *Tyzzerella_4*, measured at 6-months of age, has been found to increase risk of clinically significant internalising problems at 4-years of age. In chapter 4 it was

established that increased relative abundance of *Tyzzerella_4* was positively associated with continuous scores of both emotional problems and peer problems subscales of the SDQ. These two subscales are added together to form the internalising problems scales. This result was also again significant in the unadjusted model, which suggests that *Tyzzerella_4* is a significant bacterium of interest in understanding the gut-brain-axis (GBA) mechanisms of influencing behaviour in early childhood. As *Tyzzerella_4* is specifically associated with markers of gut inflammation (Jaimes et al., 2021), it may be possible to identify these inflammatory markers and develop interventions that can reduce inflammation in a timely manner and lessen the effect of this bacteria. Additionally, at 6-months of age increased relative abundance of *Dorea* was associated with elevated risk of externalising problems at 4-years of age. In chapter 4 it was established that the relative abundance of *Dorea* was positively associated with the continuous scores on the conduct problems subscales. Again *Dorea* is a genera of bacteria that is associated with inflammation, specifically found in higher relative abundance of those who suffer with irritable bowel syndrome (IBS) (Maharshak et al., 2018), and individuals who are overweight/obese (Gallè et al., 2020). From these two results it can be seen that both of these bacteria are consistent in findings for both continuous subscales and risk of clinical caseness. Furthermore, both bacteria are associated with inflammation of the gut are associated with elevated risk of clinical caseness of both the internalising and externalising problems. It is therefore plausible that bacteria that contribute to inflammation overall at 6-months of age may be associated with mechanisms that alter behaviour, however in order to explore this fully future research would need to further investigate the functional composition of the gut microbiome. This would allow identification of metabolic pathways of bacteria that are significant, such as metabolism of HMOs, and also the metabolic biproducts of the bacteria that significantly contribute to behavioural development.

At 12-months of age increased relative abundance of *Staphylococcus* was associated with increased likelihood of cases measured as total difficulties, and elevated risk of internalising problems at 4-years was also predicted by increased relative abundance of both *Catenibacterium* and *Turicibacter. Staphylococcus* is a common pathogen that is associated with a number of clinical diseases; however, this is also a normal part of the flora of the body and has not previously been associated with either internalising or externalising behavioural problems. *Catenibacterium* is a Gram-negative bacterium that metabolises glucose, produces several metabolites including acetic acid, lactic acid, and butyric acid, which are associated with butyrate production. This bacterium has been previously associated with cognitive ability in three year olds, specifically this bacteria was found in greater relative abundance in those whose language was non-impaired compared to language impaired children (Vaher, Bogaert, Richardson, & Boardman, 2022). However, there is no further existing literature that has previously linked this bacterium with internalising behaviour problems. Again, the novelty of these findings warrants further investigation to confirm whether these findings are repeatable. The genus *Turicibacter* is an interesting one that has been associated with host fat metabolism, including

regulation of cholesterol, and adipose tissue mass by modifying host bile acids and lipid metabolism in mice. It is therefore plausible that this bacterium influences behaviour through metabolism of dietary fat intake.

Finally, when investigating the compositional balance of the bacteria that were extracted using the SELBAL method there was a significant relationship between this balance at 6-months of age and elevated risk of clinically significant internalising problem scores, indicating that when there are increased relative abundances of the bacteria identified there is an increased risk of the problems occurring. Similarly, there was a significantly increased risk of clinically significant total difficulties scores in relation to increased compositional balance scores at 12-months.

This study presents a novel method of investigating the likelihood of clinical caseness. Through use of logistic regression methods, it has been possible to thoroughly investigate the predictive relationship of several important genera of interest relating to the GM, at three important time points of development in the first year of life. Furthermore, it has been possible to expand upon previous associations established in Chapter 4 and confirm important relationships between bacteria of interest, specifically *Tyzzerella_4* measured at 6-months of age and internalising problems, and *Dorea,* also measured at 6-months of age, and externalising problems. There are, however, limitations to this present study. Firstly, it should be mentioned that the relationships that have been established should be viewed somewhat tentatively as the number of clinically significant cases that has been established for each microbiota time point is relatively low. These small case numbers call into question the reliability of the analyses performed, especially in infant samples where the relative abundance of microbes may be very small. Further exploration of the underlying mechanisms by which the identified bacteria act would benefit from investigation using whole genome sequencing to establish the functional composition of the gut microbiome as well as the compositional. Finally, as this was a population derived birth cohort, and not a cohort that was specifically targeting aberrant psychosocial development, it is to be expected that the number of clinically significant cases would be relatively low. The strengths of this particular cohort lie in explaining variance in the dimensional psychosocial measures as explored in chapters 3 and 4. Further investigation could focus on targeted recruitment of clinical cases, although this would be somewhat difficult to achieve with concurrent measures of microbiota from birth. Given that it is expected that 10% of the population would score in the abnormal range of scores on the SDQ, it is possible to estimate the minimum required cases necessary to carry out these investigations and to recruit the numbers necessary, although a rather large cohort would be necessary. In order to firmly establish these relationships, repetition is necessary, and increased cohort size would be beneficial.

The results of this study present findings that show significant direct associations between bacteria identified at 1-, 6-, and 12-months of age and clinical cases established at 4-years using the SDQ.

One interesting finding is the confirmation of the associations of the generas *Tyzzerella_4* and *Dorea*, which were previously associated with the continuous scales of emotional and peer problems and conduct problems respectively (See Chapter 4). A novel finding for the genera *Pseudocitrobacter* and the association with increased risk of clinically significant internalising problems also warrants further investigation. In order to achieve this future research should employ whole genome sequencing methods to both confirm these taxonomical composition results as well as extend to include functional compositional findings, in order to elucidate greater understanding of the underlying mechanisms. Additionally, future research should include larger numbers of participants in order to increase the number of cases. This would allow for the investigation of dietary influence upon clinical caseness, which was not possible in the present study.

CHAPTER 6
EARLY CHILDHOOD DEVELOPMENT: THE RELATIONSHIP WITH THE GUT MICROBIOME AND DIET – A PROJECT PROPOSAL AND PATIENT, PARTICIPANT, INVOLVEMENT (PPI)

6.1 Introduction

As set out in the chapter 1, the literature review, there is considerable evidence supporting the influence of early maternal feeding practices upon GM development across the first year of life. Furthermore, there is evidence that early feeding practices influence behavioural outcomes via GM (Carlsson et al*.,* 2018). Current literature supports evidence that reduced diversity of gut microbiome and colonisation by aberrant species of bacteria is associated with poorer physical and mental health outcomes (Rook et al., 2018). Evidence of differing developmental trajectories between breast/formula fed infants provides some insight into potential causal pathways. However, there has to date been a paucity of investigation into the influence that early maternal feeding practices, including solid food introduction, has upon the relationship between GM development and behavioural outcomes. By investigating these relationships in healthy infants, it is possible to develop reference material that can further current understanding of the relationship between diet and GM development, and how this may lead to poorer neurodevelopmental trajectories. This chapter presents a proposal for a subsequent study that was not possible to undertake during the COVID-19 pandemic. PPI that investigates the feasibility and likelihood of participant uptake of the study proposed is also reported.

Chapters 2 to 5 of this thesis have demonstrated several significant associations between the composition and diversity of the GM, and both temperament and behavioural outcomes. In a systematic review of the relationship between the gut microbiome and temperament it was established that in addition to the significant associations between both alpha and beta diversity, increased relative abundance of *Bifidobacteria,* and increased relative abundance of butyrate producing bacteria and positive scores of the dimensions of temperament it was also found there are several improvements that can be made to study designs in this field. Firstly, in order to begin to understand causal links between GM diversity and composition, and behavioural outcomes, it is necessary to conduct studies of longitudinal design with concurrent GM and behaviour measures (Alving-Jessep et al., 2022). In this proposed study, the direct relationship between gut microbiota and behavioural outcome will be explored during the period of GM maturation, between birth and up to 41-months of life. Furthermore, the influence of dietary intake shall be explored through moderation analyses. It was established in chapter 3, that moderation analyses, rather than mediation, are the most appropriate type of analyses for these interactions. It was shown that it is possible to establish dietary patterns during solid food introduction that have significant influences upon the relationship between specific genera of bacteria and later behaviour. In particular it was established that there are several influences of diet upon the relationship between composition of GM, the relative abundance of

Collinsella for instance, and emotional problem scores. For this reason, the current project proposes to use similar method to explore these relationships further.

Furthermore, there is need for a tool for power calculations to determine sample size in microbiome research. However, as this has yet to be established, it is currently only possible to determine the sample size needed for this research by evaluating previous research of this type and the research performed in chapters 3-5. From chapter 3 it can be seen that when drawing conclusions from the dietary data, some of the diet clusters were too small in size to give strength to arguments. This indicates that a greater sample size is necessary when conducting this type of research. For this reason, it is it is suggested in this chapter that samples of upwards of 1000 participants are necessary to be able to evaluate fully the relationships put forward in this chapter.

In addition to improvements of study design it has also been determined that there is a need for more detailed investigation of the introduction of solid food to the infant diet. In chapters 3 and 4 it was possible to use diet clusters established during the period of introduction of solid food to evaluate the influence upon the relationship between GM composition measured at 1-, 6- and 12-months of age, and SDQ outcomes measured at 4-years. However, from these analyses additional questions arose. Firstly, it was not possible to parse the dietary information into elements including type of dietary fibre being consumed, soluble vs insoluble, and also type of meat consumed, red vs white. This information is important as it is known that variation in meat consumption influences the composition of the GM (Shen et al., 2010), and additionally specific bacteria types, such as *Collinsella* are able to colonise the GM based on the consumption of fibre in the diet (Gomez-Arango et al., 2018; Gomez Arango, Barrett, Callaway, & Nitert, 2015; Zhao et al., 2018). For this reason, the project presented in this chapter proposes the use of a diet diary, modelled on the APrON infant feeding diary developed by Jarman et al., (2018).

In addition to the in-depth exploration of the diet through use of diet diaries, it is also proposed that the milk-based diet is explored in greater depth. This shall be extended from a measure of breast milk exclusivity to also include breastmilk microbiome. It was established in chapter 4 that there is a relationship between increased relative abundance of *Collinsella*, measured at 1-month of age, and increased emotional problems measured at 4-years of age, but only in those receiving breastmilk. This is an indication that there may be transference of the maternal GM through breastmilk to the child. For this reason, investigation of the association between breastmilk microbiome and child microbiome should be explored. Furthermore, a measure of maternal gut microbiome will be taken, in the form of faecal samples, and measured as a covariate of this relationship. Further to the increased measure of breastmilk quality, it will also be established which type of formula is being fed to the infant. This will allow for analysis of variation in composition and allow for the assessment of HMO replacement by either fructo- or oligosaccharides in the formula.

Finally, the influence of parental feeding practices will be measured through use of observation of mealtime interactions. This will allow for observation of infant temperament, and eating behaviour such as cues of satiety, avoidance/acceptance, and rejection of food. Furthermore, this will allow for the assessment of parental sensitivity in responding to a child, and also responses to the eating behaviours. Several studies have indicated that infant temperament determines not only the timing of introduction to solid food (Rogers and Blissett (2019), but also may influence the responses to infant cues during feeding. Through investigation of these items the proposed study will expand upon the work carried out throughout this thesis and add to our current understanding of the complex relationship between GM composition and diversity, diet, and behavioural outcomes from birth throughout very early childhood.

6.1.1 Aims

The aim of this chapter is to present a proposal for a subsequent study and determine whether the proposed project will be seen as feasible and acceptable by target participants. In order to achieve this the proposed research design and methodology presented in section 6.2 was presented to a target audience through patient participant involvement (PPI) and feedback was gathered regarding the project suitability.

6.1.2 Proposed research questions

The study proposed in this chapter will investigate the following research questions:

1. Is the relationship between GM and temperament measured during the first two years of life, moderated by diet?
2. Does dietary fibre, during solid food introduction, moderate the relationship between GM, and behavioural and neurodevelopment?
3. Does type of protein consumed during solid food introduction, red meat vs white meat vs fish, moderate the relationship between GM, and behavioural and neurodevelopment?
4. Does Breastmilk microbiota influence the relationship between infant gut microbiota and behaviour?
5. Does maternal diet moderate the relationship between infant breastmilk microbiota, GM and behaviour?
6. Does maternal feeding responsiveness, measured through mealtime observation, influence the relationship between infant diet and GM?
7. Does infant eating behaviour influence the relationship between infant diet, GM, and behavioural and neurodevelopment?

6.2 Proposed research design and methods

6.2.1 Participants

Participants will be recruited through parent forums, pregnancy forums, word of mouth, advertisement with parent groups and opportunity samples recruited through poster placement. Participants will be recruited whilst the mother is pregnant following a healthy 20-week scan. In order for sufficient power of analysis 1000 participants to the study will be recruited. This will allow for the necessary complexity within analysis of the data and take into account the amount of attrition that is to be expected over a 2-year period. The study will be submitted to the Health and Life Sciences Ethics Committee of Aston University. Informed consent will be obtained from parents of the infants participating in the study. Additionally, all necessary Human Tissue Act (HTA) considerations for consent, storage and disposal of samples will be put in place to ensure proper handling of faecal samples. Participants will receive remuneration for their participation in this study, which will consist of £20 gift vouchers at each of the five participation sessions, totalling £100 by completion of the study.

6.2.2 Dietary measures

Dietary intake will be measured in the form of a three-day diet diary collected on three non-consecutive days, for a 24-hour period, during the week prior to each faecal sample collection. This will be modified from the APrON infant feeding diary developed by Jarman et al., (2018). This diet diary was developed to be a valid and reliable measure of total consumption within a 24-hour period and focused on the collection of milk-based diet. However, it is also possible to adapt this diary to include collection of both milk-based diet and the introduction of very first food types. As the diary includes measurement of all nutritional intake within the 24-hour period it is possible to understand milk-based practices more accurately, as many combinations of breast/formula feeding are possible. The 1-month collection point will be a milk-based diet diary. This will include breast (either direct or indirect), formula and supplementary feeding types. In the 6-, 12- and 24-month diet diaries, in addition to the milk-based questions, complementary feeding intake will be measured, and will include measures of type of feeding, frequency, and quantity. The type of feeding for milk-based diet will include details of how often the infants are breastfed and/or formula fed, formula composition (oligosaccharide replacement and pro/pre-biotic content). The method of breastfeeding (direct vs. indirect i.e., through expressed milk) will also be measured as this is known to alter the microbiome of the breastmilk and further impact the IGM. Solid food that is introduced will be measured for frequency and quantity and analysed for red vs. white meat content, fruit and vegetable content (analysed further for soluble and insoluble fibre content) and dietary fat content, which are all elements known to influence gut microbiome composition.

6.2.3 Behavioural measures

The Bayley Scales of Infant and Toddler development; Fourth Edition (Bayley & Aylward, In Press), will be used to measure infant development, at 12-, and 24-months of age, across the domains of Cognitive, Language, Motor, Social-Emotional and Adaptive Behaviour. This is a robust measure (taken directly by researchers trained to be reliable in conducting the tests) and is considered the gold standard for measuring infant development. The tests are suitable for use in infants between the ages of 1 and 42 months of age. In order to use the scale appropriately, components of the scale will be adjusted according to the specific instructions to reflect the age of the infant at collection. Additionally, the Ages and Stages questionnaire 3rd edition (ASQ3) will be completed by parents at 12- and 24-months of age. This questionnaire is a parent report, suitable from 1-monthto 5.5 years of age. The questionnaire screens in four developmental areas at 12-months of age including, communication, gross motor control, fine motor control, and problem solving. Schonhaut et al., (2013) established validity of the ASQ3, showing that it has good psychometric properties, and agreement with the Bayley III scale.

Temperament will be measured using a 3-day diary in 1-month-olds to record the frequency and duration of crying and sleeping in a 24-hour period. In 6-, 12- and 24-month-olds a parent report using the Very Short Infant Behaviour Questionnaire (IBQ) – Revised (Putnam et al., 2006), and the Children's Behaviour Questionnaire – very short form (CBQ) (Putnam et al., 2014) will be collected. Putnam et al., (2014), found that the use of the short and very short versions of the IBQ as a caregiver measure of temperament in infants, has high reliability and validity. The average age range for which this has been validated is 3- to 12-months of age. Therefore, this is a good measure of temperament for this study.

The baby eating behaviour questionnaire (BEBQ) (Llewellyn et al., 2011) and the amended child eating behaviour questionnaire (CEBQ) (Wardle *et al.* 2012) will be used to assess eating behaviour at the three collection points. The CEBQ is a psychometric assessment that has been found to be a valid and reliable scale for the assessment of the eating styles and eating behaviour of young children (Carnell & Wardle, 2007). The questionnaire is a parent-report, using a five-point Likert scale to assess behaviours including food responsiveness, enjoyment of food, and satiety responsiveness. The BEBQ was validated by Llewellyn et al., (2011) and was derived from the validated CEBQ. It is a standardised measure of infant appetite and is related to milk-feeding behaviours.

6.2.4 Observational measures

Finally, three interactions will be video recorded, observing mother and child during mealtime. This will take place only if the child has already been introduced to some food prior to the home visit. Videos will be analysed and scored using the Ainsworth maternal sensitivity scales, with both the parent and

child components (Ainsworth, 1969). Additionally, measures of frequency and strength of satiety cues, approach and avoidance behaviours, acceptance /rejection of food offered and infant emotional response to food will be taken. Maternal feeding responsiveness will also be measured from the videos. In order to maintain consistency, the interaction will be performed with mothers only as it has been found previously that mothers are more often perceived to be the primary caregiver and more often responsible for feeding their children (Patrick et al., 2005).

6.2.5 Biological measures

Two types of biological measures will be requested from the participants at several time points throughout this study. Firstly, faecal samples will be requested from the mother at 36-38 weeks of pregnancy, this will be used to identify the relationship between maternal GM composition and the first GM composition patterns that are present in infancy. Secondly, faecal samples will be collected from the child at 1-, 6-, 12-, and 24-months of age. These will be used to analyse the relationship between dietary intake, GM composition and diversity, and behavioural outcomes.

A second type of sample that will be investigated is expressed breastmilk. This will be collected at the same time points as the child faecal samples for those mothers that are still breastfeeding. This will be used to assess the microbiome composition of the breastmilk to assess the influence of breastmilk microbiome composition upon the composition of the GM. For children that are receiving any form of formula milk, either exclusively or mixed fed infants, the type and quantity of formula given will be recorded in the diet diary.

All samples will be immediately extracted for DNA upon reaching the laboratory and then stored in a HTA approved freezer at -80^{o}C, before transportation to the University of Birmingham for analysis of the 16S rRNA gene using illumine MiSeq techniques. To address previously discussed issues of low *Bifidobacteria* relative abundance in samples analysed using only the V4 hypervariable region (See Chapter 2), primers targeting both the V4 and also the V1-3 regions will be used. Amplicon Sequence Variants (ASV) will be identified for each of the samples.

6.2.6 Covariates.

A number of demographic, health and environmental covariates will be collected inclusive of maternal pre and postnatal diet quality, gestational health measures, maternal pre- and post-natal weight, birth method and gestational age, number of siblings, geographic location, attendance to nursery, and use antibiotics. This will be conducted via questionnaires. Using directed acyclic graphing (DAG) it will be determined which covariates are necessary for inclusion relating to each hypothesis.

6.2.7 Procedure.

Following recruitment and the completion of consent, the participants will be fully enrolled in the study. For the data collection at 36-38 weeks, the questionnaires, diet diaries, and biological samples will be completed at home. All necessary materials will be sent one week prior to the completion date, this will include biological labels, personal protective equipment (PPE), and sample pots necessary to safely collect the samples. Once the faecal sample is collected, the participant will contact the researcher via a study phone and arrange for a collection from their home. Following a healthy birth, the participant will notify the researcher via email of the date of birth and the subsequent visits can be arranged. The 1-, and 6- month sessions are conducted entirely at home. Participants will arrange sample collection similarly to the prenatal visit. Questionnaire material will be collected with the biological samples. At 12-, and 24, months a lab visit will be conducted. The questionnaire material will be completed prior to the visit. Faecal samples can be brought into the session or collected during.

6.3 Patient and participant involvement methods.

6.3.1 Study design and participants

Participants were recruited as an opportunity sample from the Birmingham area through use of a poster advertisement and word of mouth (See Appendix 6). Participants were eligible to take part if they were either expecting parents or parents of children up to the age of 10-years, who were fluent in English. All participants were shown a presentation of the proposed study and then given the opportunity to clarify any points that they were unsure of with the researcher. They were then requested to fill in a short questionnaire (See Appendix 7). Following completion of the questionnaire, participants received a £10 voucher as a thank you for taking part in the study. Participants took part in the study through use of an online conferencing call using Zoom software.

6.3.2 Video presentation

A video was created using the Microsoft Power point and Apple iMovie applications to put together a presentation which explained the proposed study. The contents of the power point are contained within Appendix 8 and were kept to a simple visual representation of the information given. Audio was also recoded using the iMovie application, the contents of which describe the proposed research mentioned in section 6.2. In total the runtime for the presentation was 5 minutes and 56 seconds. The video was presented in this manner as each PPI participant was recruited separately and providing the information in this way controls for possible variations that would arise with a live presentation. Participants were asked to imagine that they were being recruited for the project presented to them in the video, and to consider whether they would agree to take part in the project, what they thought

could be improved, and what they would suggest that could increase the likelihood of others taking part in the future.

6.3.3 Questionnaire

A very short questionnaire was designed to ascertain the opinions of PPI participants regarding the design, of the proposed project. Participants were asked whether they understood the information that they were being presented with in the video and if not, which part they were uncertain of. Further questions requested feedback about the study design about their happiness to complete the tasks requested by the study, and the willingness to provide biological samples. Furthermore, the participants were asked whether they would consider the remuneration to be sufficient for the amount of involvement that they were being asked to provide. The full questionnaire can be found in Appendix 7. The questions were a mixed design comprising of open-ended questions, 5-point Likert scale questions, and dichotomous yes/no questions. The questionnaire was developed and presented on the Qualtrics software platform; the participants were provided with a link sent through the Zoom chat software.

6.4 Results

6.4.1 Study Sample

A total of 10 participants were recruited into the PPI study, all of which completed both the video and questionnaire in full. Eight females and two males participated in the study ranging in age from 22 to 45 years of age. All participants had children aged between birth and 10-years of age, none were expectant parents.

6.4.2 Questionnaire results

Two of the PPI Questionnaire items addressed clarity of the project and the presentation, enquiring whether the participants were understanding the material presented to them. Question 1 addressed whether the participants understood the information provided to them in the video questionnaire was clear and understandable. This was answered favourably, with 90% of participants finding it completely clear and one participant finding it somewhat clear (See table 6.2 for full quantitative results). In addition, 90% of the participants stated that it was clear why the study was being proposed. The participant that did answered ‘No’ to question 2 completed question 3 the open-ended question and stated, as quoted below, that the reason for doing the study

‘Could have been clearer on the why and so what factor. Process to be followed was very clear’.

Question 4, “Would you be willing to be contacted for just over 2-years for the duration of the project?”, was again answered quite positively with 40% answering ‘Yes’, 50% answering ‘Maybe’, and only one participant stating that they would ‘Probably not’ be willing to participate. Similarly, with regard to completion of the study questionnaires the overall indication is that most participants would be willing to complete, with 80% stating ‘Yes’ and 20% ‘Maybe’.

Results of the questionnaire indicate that there is a more mixed feeling towards biological samples in the form of maternal and infant faecal samples and also breastmilk collection. Of the 10 participants 50% were either ‘Happy’ or ‘Somewhat happy’ to donate maternal faecal samples, 20% were either ‘Somewhat unhappy’ or ‘Not happy at all’, and 30% were neutral. They were slightly happier overall with donating samples of infant faecal matter with 80% either ‘Happy’ or ‘Somewhat happy’, only 1 participant was ‘Somewhat unhappy’, and one was neutral. Similar to infant samples most were either ‘Happy’ or ‘Somewhat happy’, with 90% of participants falling in these categories. Only 1 participant again was ‘Somewhat unhappy’. The participant that was ‘Somewhat unhappy’ was the same participant throughout all of the biological sample items.

The final question addressing whether the level of remuneration was considered substantial enough for the amount of time and effort that was being requested of the participants was again answered favourably. All of the participants answered that they think that the reward is enough for taking part.

Further to the items above with multi choice answers, there were also three open-ended questions, that provided participants with an opportunity to give more detailed feedback and express opinion. The first of these questions addressed if there was anything that would put the participant off taking part in the project. The majority of participants expressed that there was indeed something that would put them off, with only 2 participants not being put off at all. Two of the answers addressed biological sample collection aversion stating, *‘Collecting my own poo sample’* and *‘The inconvenience of handling poo’*, as the reasons. The remaining answers addressed concerns about timing of the sessions at either during the pregnancy due to concerns about premature birth, and also collection at 1-monthof age and commitment/flexibility of the appointment, being a new parent and the amount of time required for a new parent to complete the tasks.

Item 12 addressed improvements that can be made to the project. Again, these revolved around reducing the session times and questionnaires. Two of the participants requested more information about the link between gut and brain and the findings following the research conclusion. Whilst it is possible to give information during the debrief and feedback session it is not possible to give great detail about the reason for the experiment as this would lead to participant bias and could potentially influence the way in which the child is being parented or fed. Whilst it is necessary to acknowledge that the act of taking part in research such as this naturally influences participants it is also necessary to keep the influence to a minimum. Finally, one participant also suggested an increased incentive for

the final session of to increase likelihood of retaining the participants throughout the 5 sessions. The current incentive is the potential prize of £100, that can be won through entering a prize draw at the end of the 5 sessions. Although this is not given to every participant it should provide a significant incentive to stay in the project.

Finally, item 13 provided the opportunity to make any other comments or suggestions about the proposed project. One participant suggested that study sessions are carried out at the same time as the health visitor services offered to a mother following birth. A second participant, suggested that the terminology used during the presentation was incorrect, referring to the use of the word ‘health’, which was perceived as referring to only physical health and not inclusive of mental health. This shall be addressed and made clearer going forward. Secondly, use of the terminology ‘lab-visit’ was suggested as being a term that should be softened somewhat.

Table 6.1 Quantitative results of Patient Participant Questionnaire.

Question	Scale	Results
Question 1. Was the information provided to you in the video clear and easy to understand?	• Completely • Somewhat • Not very • Not at all	9 (90%) 1 (10%) 0 0
Question 2. Is the reason for doing the study clear?	• Yes • No	9 (90%) 1 (10%)
Question 4. Would you be willing to be contacted for just over 2-years for the duration of the project?	• Yes, definitely • Maybe • I don't know • Probably not • Definitely not	4 (40%) 5 (50%) 0 1 (10%) 0
Question 5. Would you be willing to complete the questionnaires for the project?	• Yes, definitely • Maybe • I don't know • Probably not • Definitely not	8 (80%) 2 (20%) 0 0 0
Question 6. Would you be happy to donate a sample of poo from the mother?	• Happy • Somewhat happy • Neither happy not unhappy • Somewhat unhappy • Not happy at all	3 (30%) 2 (20%) 3 (30%) 1 (10%) 1 (10%)
Question 7. Would you be happy to donate samples of poo from the child?	• Happy • Somewhat happy • Neither happy not unhappy • Somewhat unhappy • Not happy at all	5 (50%) 3 (30%) 1 (10%) 1 (10%) 0

Table 6.1 Continued

Question	Scale	Results
Question 8. Would you be happy to donate samples of breastmilk?	• Happy • Somewhat happy • Neither happy not unhappy • Somewhat unhappy • Not happy at all	8 (80%) 1 (10%) 0 0 1 (10%)
Question 9. Would you be happy to record yourself and child during mealtime interactions?	• Happy • Somewhat happy • Neither happy not unhappy • Somewhat unhappy • Not happy at all	6 (60%) 3 (30%) 0 1 (10%) 0
Question 10. Do you think that the reward for taking part is enough?	• Yes • No	10 (100%) 0

Table 6.2. Full responses to open-ended questions

Question	Response
Question 3. If you answer to question 2 was No, can you explain what other information you would need?	• Could have been clearer on the why and do so what factor. Process to be followed was very clear
Question 11. Is there anything that would put you off taking part in the project?	• The questionnaire maybe a bit too long. • Duration of research and quite a lot of different media to provide or complete • The length of time to answer the questionnaires: as a new parent I feel it may be hard to find 1-1.5 hours to fill in a form. • Possibly the time commitment when having a 4-week-old baby, as long as there was flexibility around time of appointments etc, it could work. • My children were prem so i would worry about the 36–38-week contact • The nutrition journal takes a lot of dedication. Also recording mealtime could be awkward. • Collecting my own poo sample • The inconvenience of handling poo.
Question 12. Is there anything that you think can be improved in the project?	• Maybe shorten the questionnaire • Shorter duration and reduced inputs. Incentive to do more would relate to me learning something about me or my child specifically • I think everything is fairly concise and well organised. I wonder if participants would be interested in taking part for 2-years. But that being said I believe I would do so, and there a big gap between each session/ child's development so I believe people would be happy that it isn't extremely "full on" • Personally, before agreeing I would want to understand more about the type of questions you were going to ask and have a real understanding of the time needed. I'd be really interested in reading your findings too. • For me it would be more interesting if there would be a more detailed explanation of what links between gut and brain are being researched • No, it’s very good • I think that more money for the final session would be a good idea to keep people involved.

Table 6.2 Continued

Question	Response
Question 13. Please share with us any other comments or suggestions you have for us here.	• Would it be possible to try and coincide these sessions roughly with the scheduled health visitor appointments to try and minimise how much a mother has to do during the month, such as having to fill out questionnaires for Health Visitor anyway, so one more for this study at the same kind of time, wouldn't be too much trouble for example. • I picked up on the wording 'health' for physical health and mental health. Implying health was just physical - that didn't feel right. I also felt the word 'lab visit' could be softened somewhat. • I would love to learn the results of the study! • Very well executed

6.5 Discussion

The primary aim of this chapter was to put forward a new proposed study through evaluation of the evidence and results determined in chapters 2 to 5. Secondly, to determine whether the proposed project will be achievable and favourable to potential participants. Through conducting a study involving patient and participant feedback, it has been determined that the participants were interested in the study proposed, however in its current state the amount of involvement was perceived as quite burdensome, and there is some significant work that needs to be undertaken in order for participant attrition to be reduced and participants to stay until completion.

Results of the PPI investigation indicate that the information that was presented to the participants was well received and understood, however there was some indication that the participants would have liked to receive greater depth of information regarding why the study was being proposed and carried out. This is a somewhat complex situation, as whilst the information could have been provided including an in-depth explanation of the relationship between diet, GM and behaviour, this could potentially have biased the results by changing participant behaviour. It would not be possible to provide this information to the potential participants as the longitudinal nature of the design could potentially lead to altered behaviours, dietary intake, and changes in parenting. This would be undesirable for the purposes of the research and for this reason information regarding the links

between gut microbiota and behaviour shall be provided but not an in-depth discussion about how this can be influenced, shaped or changed by our diet.

One interesting finding of this PPI investigation was the influence of type of biological samples. There is much more opposition to collecting maternal faecal samples as opposed to either infant faecal samples or breastmilk. This would appear to be due to the handling aspect of the faecal sample, which has been indicated in the open-ended questions. It would appear that handling of infant samples is much more acceptable, potentially due to the necessity of the parent to handle the sample when cleaning the nappy regardless of whether they take part or not. Maternal faecal samples would pose greater opposition as this is not something that is typically handled. However, there was only a 20% opposition to completing this part of the project and therefore it has not been determined that this element of the project should be revised. In order to reduce the amount that this aspect of the project puts off potential participants, the future study will emphasise that this sample type is not mandatory, and it is an optional contribution to the project.

One of the major concerns overall is the amount of time that is necessary to complete the questionnaires at each time point. In particular the complexity of the diet diary, and the amount of time that would be required during the very early visit at month 1. For this reason, the design shall be re-evaluated in order to streamline the amount of time that is required of the parent at month 1. It should be noted however that during the brief presentation that a rough guide of 1-1.5 hours was given for each timepoint. This was provided as an estimate for over all time of each session, and a greater breakdown of the schedule has not been provided. The reason behind this was to not over burden the PPI participant with information. In actuality, the earlier visits will be a much simpler process, and will be shorter in duration than the later visits. Therefore, the suggestions that have been given in the feedback are for the main part already addressed. However, a further review of the duration of each visit shall be conducted to ensure that the participants are not over-burdened during a period of significant change and potential stress.

Finally, evaluating the suggestions and comments that were made regarding the project as a whole, there was one suggestion that the study could potentially work with health visitor services to minimise the number of visits that are being carried out at the very beginning of the post-partum period. This is an interesting suggestion and one that is worth exploring, potentially even using midwifery services to streamline both the recruitment and initial visits and reduce burden upon the mother. This will require further exploration of feasibility to explore this route fully to take into consideration additional NHS ethical approval should midwifery assistance be sought for recruitment, and additional costs should the study use the services of a research nurse to support the recruitment.

In conclusion this project has received overall a favourable review when presented to the general public, indicating that following the aforementioned evaluations and potential revisions it should move forward to the next stage of planning.

CHAPTER 7
GENERAL DISCUSSION

7.1 Aims and hypotheses of the thesis

The overall aim of this thesis was to investigate the changes in diversity and composition of the gut microbiome during the period of GM maturation from birth, how this is related to behavioural outcomes and temperament throughout early childhood, and how this relationship can be influenced by diet. In order to achieve this, the first aim of the thesis was to synthesise the findings of previous research, which was achieved through completion of a systematic review, entitled “The development of the gut microbiome and temperament during infancy and early childhood: A systematic review”, presented in chapter 2, which also investigated current microbiota sampling techniques and analysis techniques, and investigation of key confounding variables.

The second aim of this thesis was to investigate the maturation of the GM throughout the first year of life, and the influence of dietary intake upon behaviour outcomes measured in 4-year-olds through a longitudinal approach. This was achieved through several investigations. Firstly, following the assessment of previous literature in chapter 2, two bacteria types were identified as bacteria of interest. These included the genus *Bifidobacteria*, and butyrate producing bacteria. The objectives of Chapter 3 were to investigate whether early infant and childhood gut microbiota composition characteristics are related to behaviour at 4-years of age, and through use of a multi-mediation approach to investigate whether GM diversity and composition mediates the relationship between early childhood diet and behaviour at 4-years. Chapter 3 hypothesised that.

1. Diversity and composition of gut microbiota will be associated with SDQ scores.
2. Gut microbiota diversity and composition will mediate the relationships between early childhood diet and SQ scores.

From the follow up analyses conducted in chapter 3, it was established that, rather than the GM mediating the relationship between diet and behaviour, that there was an indication that diet moderates the relationship between GM composition and diversity and SDQ outcomes. In other words, the GM did not explain any direct relationship between diet and behaviour, but rather the relationships between GM and behaviour vary dependent on diet. Therefore chapter 4 aimed to widen this investigation through examination of the composition of the whole microbiota. Further investigation of both the direct relationship between GM composition measured at 1-, 6-, and 12-months and behavioural scores measured using the SDQ at 4-years of age, and the moderating effect

of dietary intake, was conducted. This was achieved through multiple linear regressions carried out in Chapter 4, in which it was hypothesised that.

1. Bacteria of interest identified at each time point would be significantly associated with SDQ subscales and total problems.
2. Diet will significantly moderate the relationship between GM composition and behaviour measured using the SDQ at 4-years.

The third aim of this thesis was to investigate whether it is possible to predict clinical caseness of SDQ outcomes, internalising and externalising problems, and total difficulties measured at 4-years, from GM composition during the first year of life. This was investigated through further use of the SELBAL technique to identify bacteria of interest from clinically significant internalising, externalising, and total difficulties scores. Furthermore, likelihood of GM bacteria predicting clinical caseness was established through use of logistic regression techniques. This was investigated in Chapter 5, in which it was hypothesised that.

1. The composition of the GM will significantly predict clinical cases measured on the SDQ scales of internalising problems, externalising problems and total difficulties measured at 4-years.

7.2 Summary of the results

Significant associations between the gut microbiota and temperament during early childhood were established in chapter 2. Figures 2.2 and 2.3, synthesise the findings of the systematic review performed to meet the requirements of the first aim. Chapters 3, 4, and 5, found several significant associations between GM, composition and diversity, and behaviour outcomes, GM and dietary intake, and the moderating effect of diet upon the relationship between GM and behaviour. Figures 7.1 and 7.2 present simplified diagram of the significant associations found in chapters 3 and 4. Figure 7.3 presents a synthesis of the results for empirical chapters 3-5 of this thesis, highlighting key findings at each microbiota time point. These findings are discussed in detail in the sections below.

7.2.1 GM relationship with Temperament & Behaviour

Chapter 2 demonstrated two patterns of relationship between GM diversity, and Temperament, which are distinguishable by age: from birth to 12-months, and then 12-months and over. Synthesis of results indicate that before 12-months of age there is no association between either alpha or beta diversity of the GM temperament measures (Aatsinki et al., 2019; Alving-Jessep, Botchway, Wood, Hilton, & Blissett, 2022; Kelsey et al., 2021). Findings in children from the age of 12-months up to the age of 6-years 11-months, show that higher alpha diversity is related to Surgency/Extraversion in both males and females, High-Intensity Pleasure in males only and lower levels of effortful control in females only (Christian et al., 2015).

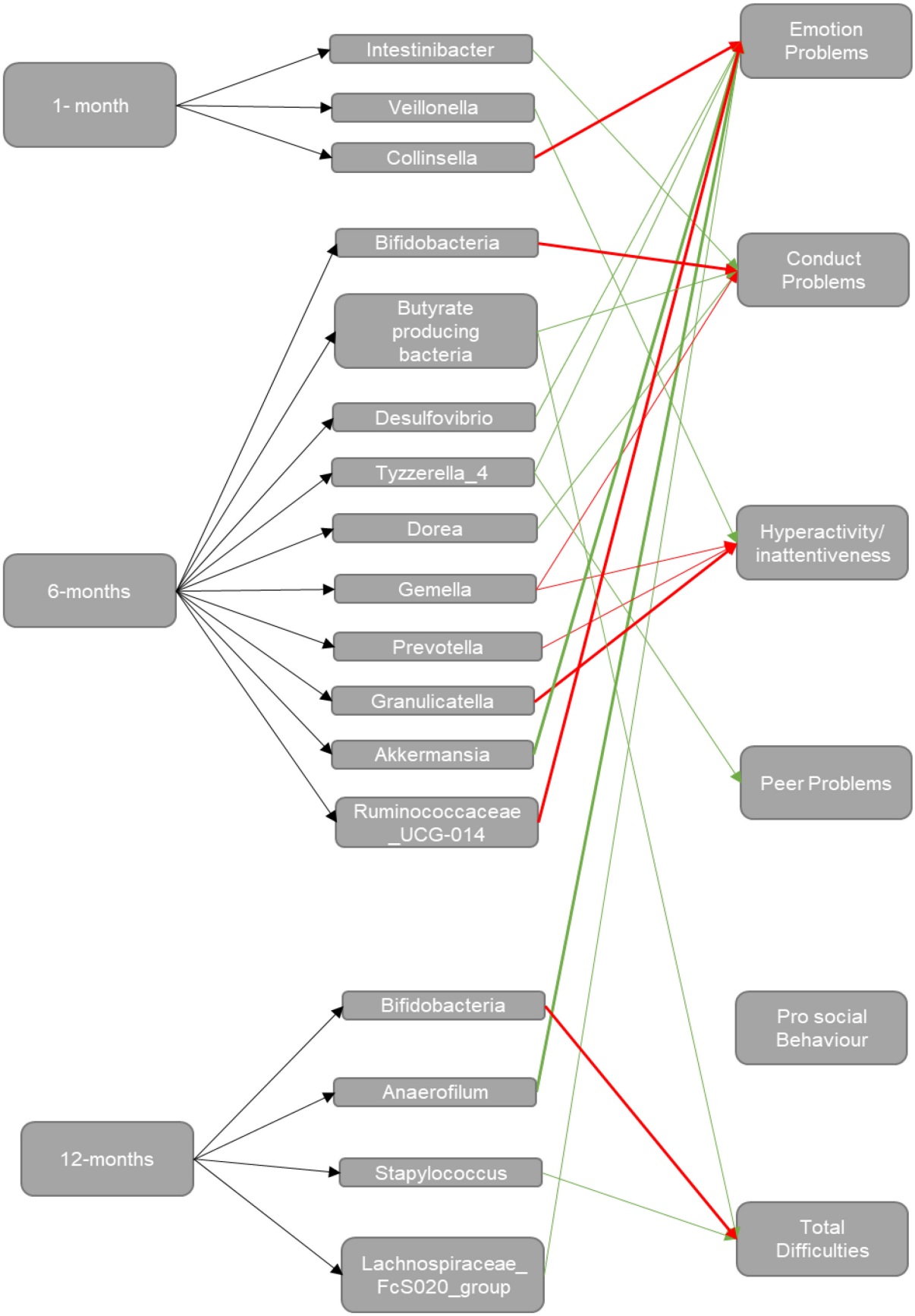


Figure 7.1. Synthesis figure illustrating significant direct and indirect relationships between GM, and SDQ outcomes at 4-years.
Note. Green arrows denote positive relationships and red arrows negative. Thicker arrows indicate relationships significant only when moderated by diet.

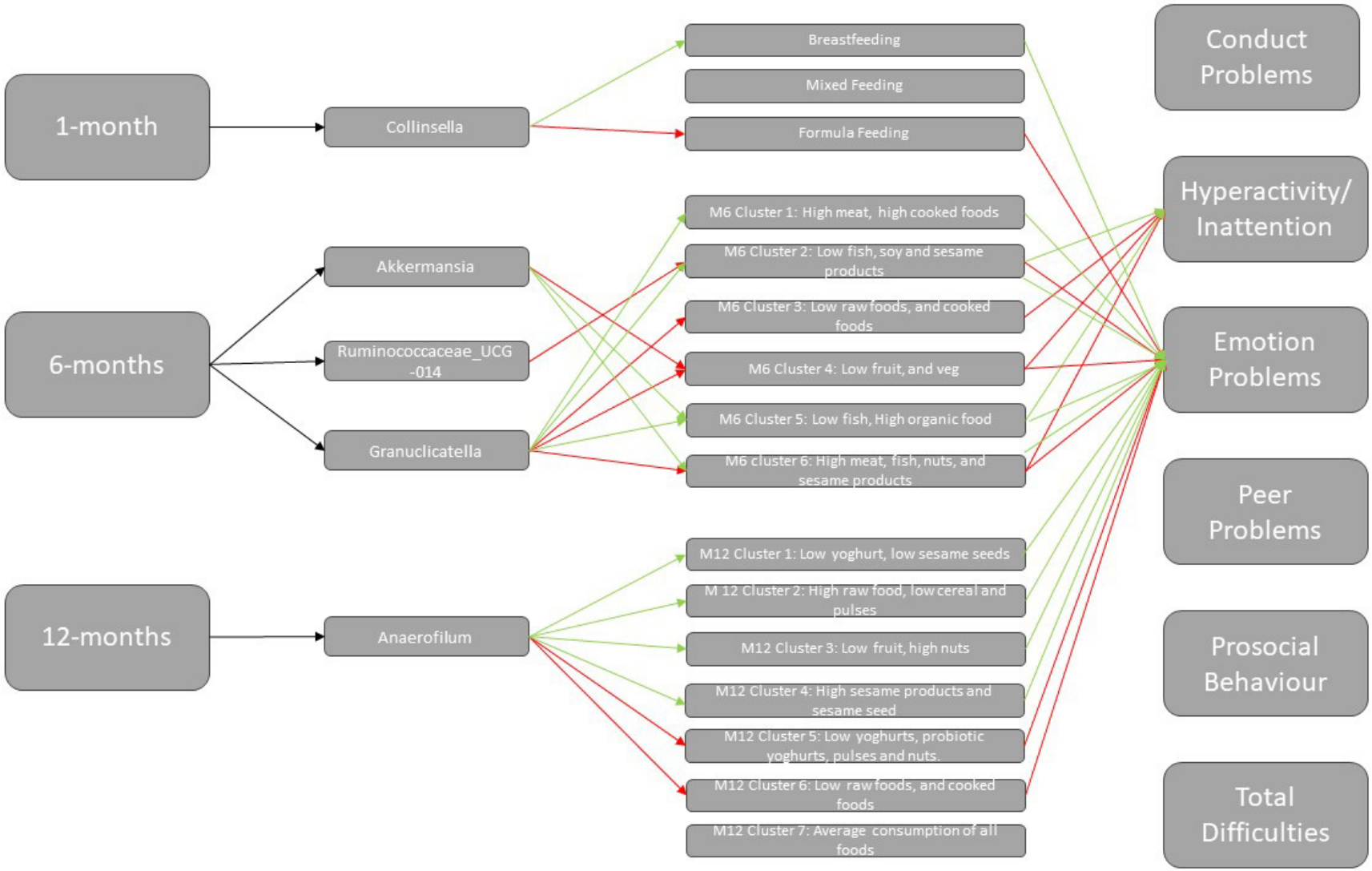


Figure 7.2. Synthesis figure illustrating significant moderation relationships between GM, and SDQ outcomes at 4-years. These results present the original findings and not without reduced cluster sizes as there were no significantly large changes as a result of the additional analyses.
Note. Green arrows denote positive relationships and red arrows negative.

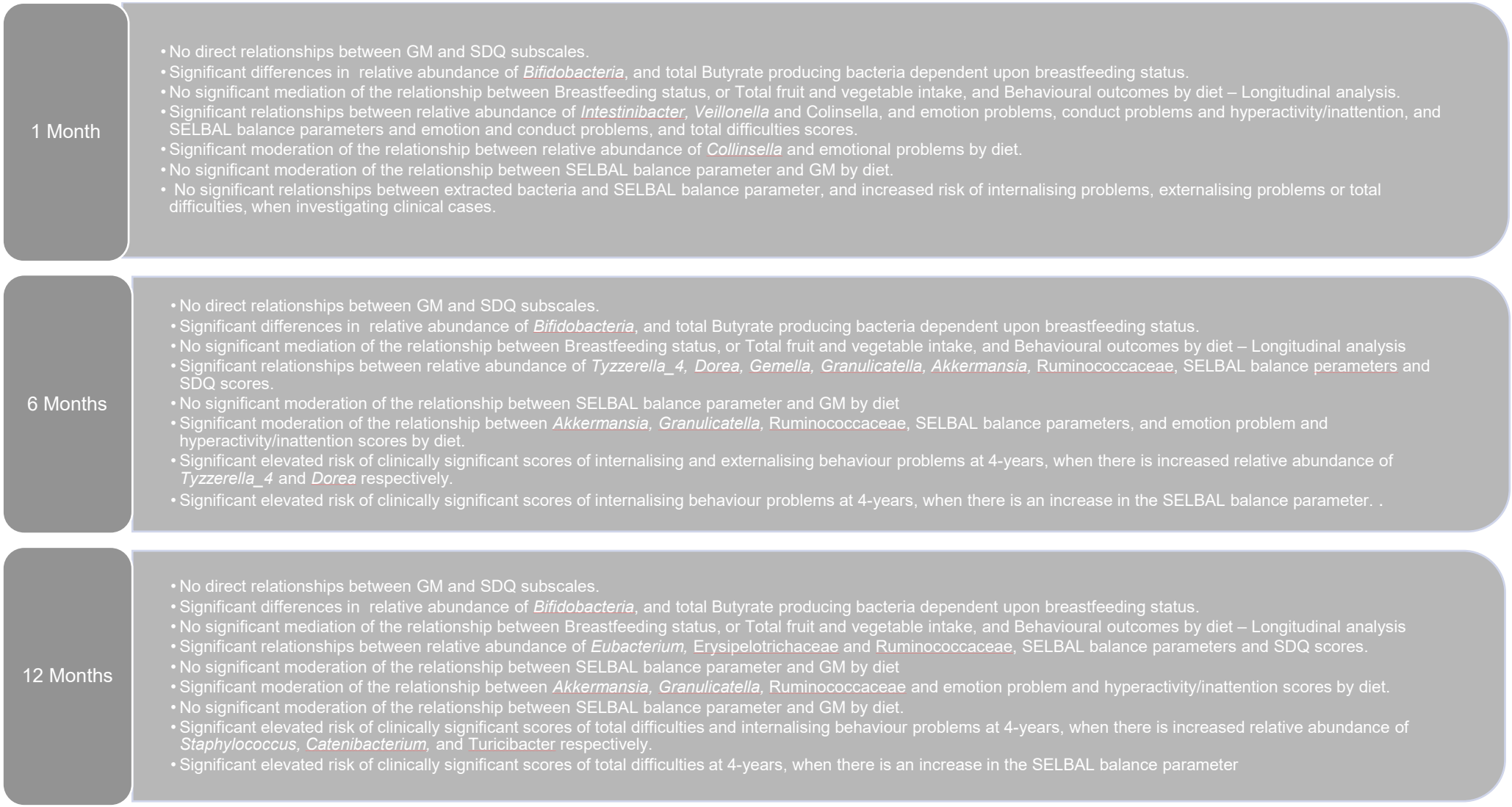


Figure 7.3. Synthesis figure illustrating overall findings of empirical chapters 3-5, the relationship between Diet, GM during the first 12 months, and SDQ outcomes at 4-years.

Furthermore, there were unique community structures associated with Surgency/Extraversion in males and Fear in females, measured as beta diversity (Christian et al., 2015). These findings somewhat support the literature previously examined in chapter 1, literature review, of this thesis, which found that dysregulation of the GM is a potential contributory factor to alterations in temperament through modification of the HPA axis (Luczynski et al., 2016). Although the terminology 'dysregulation' is somewhat outdated, at the time of publication, the term referred to altered alpha diversity of the GM, characterised by elevated alpha diversity in children under 12-months of age or by lowered alpha diversity in all individuals older than 12-months, compared to those of healthy individuals, and a composition that was either depleted of a number of commensal bacteria or significantly elevated in pathogenic species. Since completion of chapter 2, and subsequent submission for publication, a further paper examining the relationship between GM composition and diversity, and Temperament in 12-month-olds, has been published by Fox et al. (2021). Using similar Illumina MiSeq techniques, and the Infant Behaviour Questionnaire-Revised (IBQ-R; Putnam et al., 2014), their results further support that in 12-months olds beta diversity is associated with Surgency/Extraversion, however they do not differentiate by sex in this sample. This finding further supports the conclusions drawn from synthesis of literature in chapter 2.

Synthesis of findings, presented in chapter 2, regarding the relationship between composition of the GM and temperament found several interesting patterns. Firstly, increased relative abundance of *Bacteroides* were found to be associated with lower scores of High-Intensity Pleasure, cuddliness, and duration of orienting, in 6-month-olds , and reduced levels of sadness and impulsivity, and increased levels of inhibitory control, in 5–7-year-olds : temperament traits typically associated with better developmental outcomes, for example lower levels of impulsivity and depression measured at 4-6-years of age have previously been found to have similarly low levels on the same dimensions in later childhood and into adulthood (Lahey et al., 2016). These results are interesting because investigation into adult models has found significant associations between increased relative abundance of *Bacteroides* with less positive outcomes such as major depressive disorder and increased anxiety (Mason et al., 2020; Zhang et al., 2022). This further supported the necessity of this thesis to investigate the development of the gut microbiota during the sensitive period of maturation, as work differentiating early development should not expect to find the same patterns of results in both adult and child studies and vice versa.

A second significant pattern of association relates to increased relative abundance of *Bifidobacteria*. This group of bacteria has been significantly discussed throughout Chapter 1, but specifically in sections 1.1 and 1.2. There is a wealth of evidence that has been established that *Bifidobacteria* is associated with breastfeeding exclusivity and furthermore health, both physical and psychological (Hidalgo-Cantabrana et al., 2018; Wong, Iwabuchi, & Xiao, 2019). The findings of chapter 2 of this

thesis show that there is a potential link between increased relative abundance of *Bifidobacteria* and increased ability of the temperamental trait emotional regulation. As temperament, specifically emotional regulation, has been shown to influence the way in which behavioural strategies develop on an individual level (Rothbart et al., 2011), and there is a wealth of literature establishing that early childhood temperament problems is associated with social behavioural development and behavioural problems (Jones & Sloan, 2018; Sanson et al., 2004; Yoleri, 2014), *Bifidobacteria* was established as a bacterium of interest for the investigation carried out in chapter 3.

Finally, a third pattern of association between GM composition and temperament was established in chapter 2. The pattern showed that positive temperament traits were associated with GM communities biased towards short-chain fatty acid production from a metabolism based on dietary fibres and complex carbohydrates. These bacteria specifically belong to a group known as the butyrate producing bacteria. This finding presented evidence to support butyrate producing bacteria as a bacteria group of interest pertinent to the investigation in chapter 3, and further established the necessity for deeper investigation into the influence of diet upon the relationship between GM and behavioural development.

Results of chapters 3 and 4 demonstrated the relationship between both diversity and composition of the GM measured at 1-, 6-, and 12-months of life and behavioural outcomes measured as the continuous subscales and total difficulties scores of the SDQ at 4-years. A somewhat surprising result in chapter 3, was that there were no significant relationships between either relative abundance of *Bifidobacteria*, or total butyrate producing bacteria, with SDQ subscales or total score, which, given the synthesised results of chapter 2, is somewhat contradictory. This may suggest that although measures of temperament and behaviour measures (Jones & Sloan, 2018; Sanson et al., 2004; Yoleri, 2014), such as the SDQ are related, they are not the same and therefore patterns of GM relationships with each of these measures are different. When investigating the composition of the whole microbiota, as opposed to pre-determined bacteria groups, several bacteria at each timepoint were related to SDQ scores at 4-years. One genera of bacteria *Tyzzerella_4* has previously been associated with gut inflammation (Jaimes et al., 2021). This bacterium was associated with peer problems at the age of 4 years. Although the relative abundance of this microbe, within the microbiota of children aged under 1 year, is relatively small, the results that were found persisted after both correction for false discovery rates and removal of outliers that would potentially influence and drive the significant result. This evidence again provides further motivation to explore the influence that dietary intake has upon composition of the GM, and further how this may influence the relationship between GM and behavioural outcomes.

7.2.2 The relationship between diet and GM

Chapter 3 of this thesis investigated the impact of dietary intake upon both the diversity and composition of GM, measured as relative abundance of *Bifidobacteria*, and total butyrate producing bacteria. At 1-month of age, dietary intake was measured as milk-based diet, qualified by level of breastfeeding exclusivity. It was determined that the microbiota sub-cohort of the Barwon Infant Study, described in Section 3.5.1, showed significant differences in alpha diversity, and total butyrate producing bacteria, between those receiving exclusive breastfeeding, mixed feeding, and exclusive formula feeding. An interesting result is that contrary to previous literature as described in section 1.1, there was no significant effect of breastfeeding exclusivity upon relative abundance of *Bifidobacteria* at either 1-, or 6- months of age. Furthermore, there was no indication that this result changes once solid foods are introduced at 6- months of age, which has previously been linked with breastfeeding cessation and an increase in alpha diversity of the microbiota. One interesting and unexpected finding in this group was that at 12-months of age mixed-fed infants showed significantly higher relative abundance of *Bifidobacteria* compared to both breastfed and formula fed infants. Looking at previous findings it would be expected that breastfed infants would have the highest relative abundance of *Bifidobacteria*, which would be slightly lower in mixed fed infants and significantly lower in formula fed infants (Stewart et al., 2018b). It is surprising that the breastfed infants in this groups showed relative abundance of *Bifidobacteria* that is comparatively so low. Additionally, when looking at the mixed fed infants' group, we do not have a measure of how much breastmilk is being received relative to the mount of formula that is being given. It can be seen in figure 3.3 d, that there is a far greater range of *Bifidobacteria* relative abundance in the mixed fed infant group, which would indicate that the amount of formula being given varies widely. However, this still does not explain why the exclusively breastfed infants are so low, replication of this result in future studies would allow us to determine whether this is an unexpected result of this particular cohort. Finally, at each time point, formula fed infants showed significantly elevated abundance of butyrate producing bacteria compared to those receiving either mixed feeding, or exclusive breastfeeding. As this finding is present at all time points, it indicates that this pattern persists after solid food introduction. This result is consistent with previous literature that highlights the earlier increased diversity and adult like microbiota that can be found in the guts of infants receiving exclusive formula feeding compared to those receiving either mixed or exclusive breastfeeding (Stewart et al., 2018). It is furthermore evidence that cessation of breastfeeding at any time point is associated with the maturation of the gut microbiota (Stewart et al., 2018b; Valles et al., 2014), and indicates that the maturation process may begin as early as 1-month of age, although further dietary evidence is necessary to support this postulation. Interestingly butyrate producing bacteria are perceived to have beneficial influence upon mental health in adults (Du Toit, 2019).

Chapter 3 further investigated the influence of solid foods introduced into the diet, and measured in this investigation at 6-, and 12-months of age. It was found that the total intake of fruits, vegetables,

and dietary fibres at 12-months of age was not associated with total number of butyrate producing bacteria. However, it was not possible from the dietary measure used in this investigation to accurately identify good sources of dietary fibre, compared to poor sources, or the types of fibres, either water soluble or insoluble, from the types of fruits and vegetables consumed. This methodological limitation is further discussed in section 7.3 of this chapter. As a result of this finding, dietary intake was then investigated through use of PCA analysis and k-means clustering to identify dietary patterns at 6- and 12-months of age, which will be further discussed in the next section.

7.2.3 The influence of dietary intake upon the relationship between GM & Behaviour

During the longitudinal analyses performed in chapter 3, it was determined that there was no mediating effect of microbiota composition or diversity, upon the relationship between diet and SDQ outcomes. As this was the first study to investigate in depth relationships between diet, GM, and behaviour, and it focused on a specific set of GM measures, it is not wholly surprising that an absence of significant results was found. It is well established in the existing literature that there is a direct relationship between GM diversity and composition and dietary intake, and between GM and behaviour. Furthermore, it is well established in microbiota research that dynamic interactions exist between environmental factors, microbiota and host (Xia, Sun, & Chen, 2018), as shown in the model shown in figure 7.2. Thus, the lack of significant mediating relationships was not sufficient evidence to rule out other potential interactive relationships between these three factors. Therefore, follow up analyses were performed to determine the influence of diet upon the relationship between GM and SDQ outcomes. It was established during follow up analyses that rather than a mediating effect of GM composition, dietary intake moderated the relationship between GM and SDQ outcomes.

Chapter 3 demonstrated that the relationship between relative abundance of *Bifidobacteria* measured at 6-months and conduct problems measured at 4-years of age was significantly moderated by diets composed of higher-than-average meat consumption and lower consumption of all other food types. However, as mentioned in section 3.7, this cluster consisted of only two children and therefore conclusions drawn from this result should be viewed with the utmost caution. Further discussion of limitations of analyses due to number of participants will be further discussed in section 7.3. Chapter 3 also showed that diet significantly moderated the relationship between relative abundance of *Bifidobacteria* measured at 12-months, and the sub-scale measuring hyperactivity/inattention at 4-years. Specifically, the moderation effect was driven by two clusters, one with much lower-than-average consumption of yoghurts and lower than average consumption of sesame products, and the second cluster with lower-than-average raw food and cooked food, and higher-than-average pre-processed or packaged foods. These novel findings indicate that diet should be considered as a potential moderator of the relationship between GM and behaviour. In particular when investigating the effect of the complex interactions that lead to behavioural outcomes, it is not simply that diet or

GM leads to behaviour. Rather the GM requires specific dietary substrates in order to produce the metabolites that either leads to vulnerability or protective influence upon behaviour, which can be seen during in depth investigation of dietary patterns. This line of investigation was continued in chapter 4, when investigating the composition of the whole microbiota.

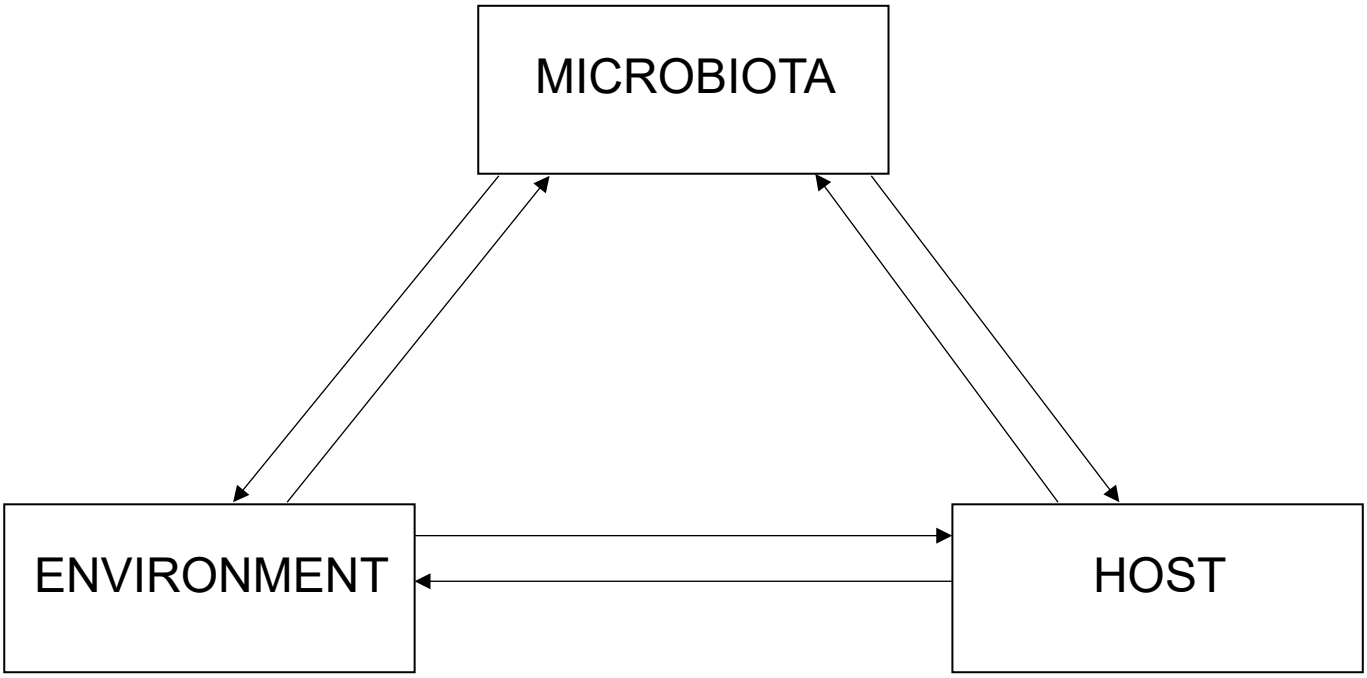


Figure 7.4 The dynamic relationship between host, microbiota, and environment.

There was an interesting pattern of results that showed that diet significantly moderated the relationship between GM composition and the emotional problems subscale of the SDQ. Furthermore, it was found that breastfeeding status at 1-month of age moderated the relationship between relative abundance of *Collinsella* and emotional problems at 4-years. This result highlights the possibility that diet as early as 1-month of age may create conditions within the gut which moderate bacterial capacity to influence later behavioural outcomes. Furthermore, *Collinsella* is associated with low levels of dietary fibre intake in the maternal diet (Gomez-Arango et al., 2018) and can be transferred from mother to child through breastfeeding (Jost et al., 2014). At 6-months of age relative abundance of both *Akkermansia* and an unclassified genus of the family Ruminococcaceae*,* and emotional problems measured at 4-years of age were also significantly moderated by diet. Specifically, *Akkermansia* was moderated by diet with higher-than-average nut consumption, and higher than average cooked foods, although it was not possible to ascertain the dietary qualities of the cooked food, such as olive oil usage, which contributed to strong positive relationship between relative abundance of *Akkermansia* and emotional problem scores. And, finally at 12-months of age relative abundance of *Anaerofilium* was associated with emotional problems at 4-years, and significantly moderated by diet. Of course, a number of other factors have been shown to influence the development of emotional problems in

children, these include parental availability, parental illness, social economic status (Easterbrooks, Bureau, & Lyons-Ruth, 2012; Sigal, Perry, Robbins, Gagné, & Nassif, 2003), and maternal socialisation, (Yagmurlu & Altan, 2010), so there are multiple interacting mechanisms by which emotional problems may develop. These results at month 6 persisted following correction for small number of individuals in clusters and the removal of cluster 1 from the analyses. Following similar corrections of the small cluster sizes at 12-months, there was no longer a significant moderation of the relationship between relative abundance of *Anaerofilium* and emotional problems. There were however two new significant moderation relationships between conduct problems and relative abundance of *Eubacterium* and an unclassified group of Ruminococcaceae. With regard to diet, the metabolic pathways associated with emotional regulation and microbiota involve a number of factors including epigenetics, insulin pathways, and the HPA axis. A review carried out by Michels (2019) investigated the relationship between emotional regulation, emotional over-eating, stress and appetite, and how these can interact with the GBA to lead to emotional problems and obesity. They highlight the inflammatory system as potential pathway of interest to the GM-diet-behaviour pathway; however, it is also concluded that the complex interplay between diet, GM and emotional problem behaviour has yet to be explored in depth with regard to the metabolome. Metabolomics can be used to explore the metabolic profiles of organisms within the GM and the interplay between the metabolites, (the small molecules such as SCFAs produced through the bacterial metabolic processes and the digestion of substrates ingested in the host diet), and their interaction with the GBA. This again highlights the necessity of whole-genome sequencing, which allows for detailed exploration of bacterial metabolic processes, through use of genome-scale metabolic network models (GEMs), in order to explore these relationships further (Fang, Lloyd, & Palsson, 2020).

Chapter 5 investigated whether composition of gut microbiota measured at 1-, 6-, and 12-months of age predicts identified cases with clinically significant behavioural outcomes. Although the relationship between GM composition was investigated with regard to behaviour measured using the SDQ at 4-years, in chapter 4, this was established through use of continuous scores. Chapter 5 investigated clinically significant cases using established cut points for the same scale, which was hypothesised to produce different patterns of results compared to those found using continuous scores. At 1-month of age (n=289) there were 13 cases identified for internalising scale of the SDQ, 19 for the externalising scale and 12 for total difficulties. At 6-months of age (n=298), the cases numbers were the same, and at 12-months (n=251) there were slightly fewer datasets, with 12 internalising cases, 14 externalising cases and 8 total difficulties cases. For a full description of the participants see section 5.3.1. Through use of the SELBAL method of identifying candidate bacteria of interest, and a logistic regression approach to establish likelihood of clinical caseness based on relative abundance of these identified bacteria, several relationships were established. For each of the time point of microbiota sampling, bacteria were found that significantly predicted clinical cases indicating externalising problems or

internalising problems. However, there were no bacteria that indicated clinical cases classified on the basis of 'total difficulties'.

At 1-month of age, increased relative abundance of *Pseudocitrobacter,* was associated with increased risk of clinically significant internalising problems, and both *Collinsella* and *Veillonella* were associated with increased likelihood of externalising problems at 4-years of age. As established in chapter 4 *Collinsella* can transfer from mother to child through breastfeeding, and is associated with maternal obesity during pregnancy, and low dietary fibre intake. This result indicates that maternal diet may play a significant role in the relationship between infant diet, GM, and behaviour, and therefore should be considered as a variable of interest and more than a covariate in future studies, see chapter 6 section 6.1.2. However, the association established in chapter 4 was found in association with emotional problems, which is part of the internalising problem scale rather than the externalising scale. One plausible explanation is that the relationship between *Collinsella* and behaviour through interaction of the GBA is highly complex and probably not linear. It may be possible to determine that elevated levels at this time point are disruptive to a healthy composition of the microbiota and behavioural development, however it is not possible to determine any further mechanistic information which illuminates pathways to the development of more internalising problems or externalising disorders from the data within this study. With regard to relative abundance of *Veillonella,* in chapter 4 increased relative abundance was found to be associated with elevated Hyperactivity/Inattention, likewise it was also associated with clinically significant externalising problems in chapter 5. Therefore, unlike *Collinsella,* the continuous scales show potentially congruent results to the results predicting caseness. These results are again echoed in the 6-month measurements with significant prediction of caseness on the internalising problems scale at 4-years of age, and also significant associations with both the continuous scores of Hyperactivity/Inattention and conduct problems. From this it is possible to determine that the patterns of association determined using continuous subscales of the SDQ are not necessarily indicative of associations that predict clinical caseness.

These results indicate several potential directions for future research, one of which would include concurrent measurements of maternal dietary intake for as long as the child is breastfed. This would give us an insight into the components of maternal diet that can be linked through breastfeeding to the variation in gut microbiota composition. Furthermore, investigations into the microbiota composition of breastmilk would be beneficial to determine which bacteria are transferred to the child. The results presented in this thesis further established the importance of considering diet as more than a covariate in investigations into the relationship between GM and behaviour. These future directions shall be discussed in greater depth in section 7.3.

7.2.4 SELBAL balance parameters relationship with Behaviour.

As mentioned in chapters 4 and 5, the SELBAL balance parameter is a measure of compositional balance relating to the outcome of interest. In chapter 4 the outcomes of interest were each of the 5 subscales of the SDQ, and total difficulties scores. In chapter 5, this was similarly measured with regard to both internalising and externalising problems and again total difficulties. Several direct relationships were established that showed strongly positive relationships between the compositional balance and the SDQ subscales at each of the time points 1-month, 6-months and 12-months. This balance parameter is a relatively new method of investigating the influence of whole microbiota and is flexible enough to also allow for continuous outcome variables, which is a strength of this method.

Further to the direct relationships that were established, there were no significant moderations of the relationships between the balance parameters and SDQ scores by diet. This is interesting considering the specific extracted bacteria are highly related to the balance parameters, and therefore this further highlights the need to explore complex systems such as the gut microbiota. It gives further evidence to the caution that should be taken when both reviewing results presented and the methods chosen to explore complex systems such as the gut microbiota. Finally, with regard to chapter 5, logistic regressions were performed to determine whether balance parameters would be predictive of clinically significant scores of internalising and externalising problems scores as well as total difficulties. It was discovered that at 6 months, increased balance parameters were significantly associated with elevated risk of internalising problems, and at 12-months increased balance parameters were significantly associated with elevated risk of clinically significant total difficulties scores. However, these results should again be viewed cautiously as the number of clinical cases for each scale and at each time point was small. Nonetheless, given the nature of the data being drawn from a population derived birth cohort, it can be expected that approximately 10% of any population would score in the abnormal range and therefore small numbers are to be expected.

7.2.5 Project proposal and patient, participant involvement

Chapter 6 synthesised the findings from the previous 5 chapters and brought them together to propose a new study. The project is a longitudinal design and aims to bring together the findings of the current thesis and address some of the limitations. Many of the limitations addressed in section 7.3, and the future directions suggested have been applied when designing this study. Furthermore, chapter 6 conducted a study utilising patient participant involvement (PPI), in order to understand whether the proposed study would appeal to the target participant group, what components of the study would deter participants from taking part, whether there was anything that would possibly encourage participants further to take part.

From the PPI study it was determined that there is some significant work that needs to be carried out in order to streamline and fine tune the proposed study. This ultimately would lead to increased participant uptake and reduced participant attrition. Overall, the information that was presented to the participants was well received and understood. Although, the PPI participants did feel that there would be increased interest if there was more information provided about why the study was being carried out. This is complex as there is a fine line between informing participants and providing information that may bias the results. It may be possible to provide more information about the gut bacteria and behaviour, however it is unlikely that the study will be redesigned to include more information about how this relationship would be shaped by diet for this reason.

There were several concerns with regard to the amount of time that would be required to complete the diet diary, video observations, and questionnaires. This shall be addressed and further streamlining of the amount of information requested will be carried out. An interesting suggestion that was made was the potential use of maternity services, including midwives, and/or health visitors to conduct the first visit at 1-month. Further information is required to ascertain whether this is possible, whether there would be any complications with passing this through NHS ethics, and an evaluation of the monetary impact of hiring a research nurse will be carried out. Although time was a concern, the amount of remuneration suggested in the proposal presentation, was deemed as satisfactory for the amount of time required.

Finally, the collection of biological samples, particularly faecal samples, was found to be off putting for some PPI participants. In particular the collection of maternal faecal samples, this was likely due to dislike of handling faeces, and the perceived unclean aspect. Infant samples were less opposed, which is most likely due to the necessity for parents to change infant nappies, therefore to some degree they are required to handle infant faeces. This dislike will be taken into account and further efforts will be made to emphasise that biological samples are voluntary and not mandatory for study completion. Overall, the proposed study received a favourable review, however, there are still revisions necessary in the next stage of planning.

7.3 Methodological critique and future directions

There are several methodological strengths of this thesis. Firstly, the longitudinal design presented in chapters 3-5. The concurrent measures of dietary intake and microbiota sampling at multiple time points allows for more thorough investigation of dietary influence upon the GM composition and diversity. Secondly the rigorous identification of covariates of interest using DAG methodology presented in Appendix 2, ensured that environmental factors that may contribute to the models investigated are taken into account. In doing so it is possible to establish with more confidence the direct relationships between GM and behaviour. Furthermore, a significant strength of this thesis is the use of novel statistical approaches to investigate the relationship between GM, behaviour, and dietary

intake, establishing diet as factor that should in future investigations be considered as more than a covariate. As mentioned in section 7.2.3, variation in GM composition and diversity alone cannot explain the complex interactions that lead to behavioural outcomes. The metabolic by-products that interact with the GBA, and result in either protective or detrimental influence upon behavioural outcomes, are for some genera of bacteria dependent upon the dietary substrates that are introduced into the environment. Therefore, diet should no longer be measured as a covariate when investigating GM, but a thorough investigation of each dietary factor that is commonly introduced during solid food introduction should be undertaken.

There were several methodological limitations identified during the process of this thesis, some of which were identified as limitations of studies included in the chapter 2 systematic review. Taking a systematic approach, it is possible to identify methodological limitations from design of the study, through microbiota pipelines, and materials used.

In chapter 2, the systematic review identified that investigations of the relationship between GM and temperament often lack the ability to determine causality due to study design. It was suggested that concurrent measures of both gut microbiota and measurement in a longitudinal approach would be the most beneficial design and allow the possibility to determine causal routes of influence. This critique is one that is also relevant to chapters 3-5, as although there were concurrent measures of microbiota and diet, there was not a measure of microbiota taken at 4-years of age, at the time of the outcome (SDQ) measures. However, this was due in part to the global pandemic COVID-19, which resulted in delayed analysis of the microbiota samples taken in the 4-year-old wave of the Barwon Infant Study cohort. Again, there was also no measure of diet taken during the 4-year-old wave, which would have given a more complete picture of the development as 41-months is the typical point of GM maturation to adult enterotypes, and a measure at 4-years or 48 months would have allowed critical analysis of the predictors of the relationship between behaviour, GM, and diet once an adult like pattern has been established.

An additional methodological point of interest is related to a statistical approach available to establish the number of participants necessary for studies of this type. Currently in the field of microbiota research, there is little to no use of power calculations to establish sample sizes. Studies of this nature can range from participant numbers in the low hundreds to studies of upwards of 5000 participants . Using previous literature to justify sample sizes has become a standard approach in this field, which has led to vast differences in sample sizes. This is also a significant challenge due to the number of environmental covariates that are often established as being necessary for incorporation into statistical analyses of microbiota data. For all these reasons a robust statistical approach to determine sample sizes is necessary, and warrants development. The issue of sample size was encountered during the interpretation of results in chapters 3 and 4, when investigating dietary clusters, as there

were clusters that consisted of as few as 2 participants. Furthermore, due to the small number of clinically significant cases, established through use of the standard scoring procedure the SDQ, within this overall sample it was not possible to extend the investigation of caseness to include the influence of diet upon the ability of GM composition to predict clinically significant cases determined by the SDQ at 4-years. Although many analyses were statistically significant it is necessary to use a large amount of caution when interpreting these results. For example, results of the moderation analysis performed in chapter 3 generated a dietary cluster with only 2 participants, and whilst the results are statistically significant, the ability replicate this finding is called into question, reducing the reliability of this result. Additionally, this result also brings into question whether a dietary cluster of only 2 is a meaningful pattern, thus limiting interpretation regarding its clinical or practical significance. The microbiota subset of the BIS study included a total of 324 participants, which would have significantly benefited from increased numbers to include at least 1000 participants as suggested in the project designed in chapter 6 of this thesis. As mentioned in chapter 2, there are no set statistical methods that are currently used when justifying sample sizes in microbiome studies. In general, the participant sizes vary widely, and can range from a few hundred to the thousands. The ideal direction for this branch of research would be to develop a statistical method in order to have a microbiome-based power analysis. However, until this is developed, it is possible to look to more traditional methods of power analysis.

Furthermore, it is important when designing these studies to take into account the microbiota pipeline and quality assessment, which can reduce the number of final samples, as well as attrition from longitudinal studies. There was an approximately 25% drop out from attrition between the 6-month and 12-month time points, and the quality assessments performed during the BIS pipeline reduced sample sizes by approximately 10% further. Power calculations will therefore also need to adjust for potential drop out encountered in longitudinal designs, the figures presented here could potentially be used as a guide to estimate the scale of the drop out over time.

Regarding the microbiota pipeline, there are two elements that warrant in depth critique. Firstly, although there are established protocols for collection of faecal samples in adult humans, as laid out by the Human Microbiome Project (Integrative, 2014), there are few established protocols for infant samples. Faecal samples of children who are receiving a milk-based diet or being introduced to the first solid foods are substantially different from adult samples. Increased sugar concentration in samples from the milk-based diet and increased water content may impact the way in which samples should be stored. It can be clearly seen in Chapter 2 that the studies synthesised in the systematic review use a number of collection techniques, that ranged from donation as fresh, to freezing at home, and storage at ambient room temperatures at home. Variation in the storage technique has been known to influence the samples through repeated freeze-thaw cycles, which alters the ratio of Firmicutes to Bacteroidetes due to alteration in the structure of the cells of Gram-positive bacteria

when they are frozen (Choo, Leong, & Rogers, 2015). Chapters 3-5 also had a variation in storage technique prior to arrival at the laboratory, however, this was taken into account through controlling for fresh vs. frozen during the statistical analysis. For these reasons, future research should look to establishing a clear protocol for the collection, storage and processing of infant faecal samples in young children.

A second aspect of the microbiota pipeline that warrants further consideration is the selection of both hypervariable regions targeted when using 16S rRNA techniques to identify bacteria. The most prominently targeted hypervariable region in use is the V4 hypervariable region, which was also investigated in Chapters 3-5, which was also used in an analysis of the same cohort and published by Loughman et al. (2020). Synthesis of the papers included in the systematic review presented in Chapter 2 of this thesis, showed a total of three combinations of hypervariable regions in analyses, including V1-3, V4 only, and V3-4. These hypervariable regions are the genetic markers of bacteria that show a considerable level of diversity amongst species and can be used as markers for identification. Variation in hypervariable selection contributes to bias in species detection within microbiota analyses, and for this reason it is often not possible to conduct meta-analyses of papers that select for widely different genetic markers. This lack of congruence in the field also makes the generalisability of results particularly difficult, and furthermore restricts the ability of review to conduct meta-analyses. With regard to the particular age group of interest, it can be seen from previous literature established in Chapter 1, that in individuals typically denoted as healthy there is a dominance by the species *Bifidobacteria*. However, there is evidence that coding for the V4 hypervariable region only is significantly biased away from *Bifidobacteria* species and biased towards increased levels of *Firmicutes* (Alcon-Giner et al., 2017; Biol-Aquino et al., 2019). It could be plausible that this is a contributory factor for the absence of results in chapter 3, which took an *a priori* approach to investigate relative abundance of *Bifidobacteria.* By selecting for hypervariable regions that account for community structures dominated by *Bifidobacteria,* widening the hypervariable regions to include the v1-3 regions it is possible to account for the bias and allow for future results to be more generalisable. For these reasons is it also important that when selecting pipelines for investigations of microbiota, from birth throughout the maturation process of the microbiota, it is necessary to carefully select the hypervariable regions that do not bias the findings. Furthermore, this should be established in an agreed protocol for scientists investigating the infant microbiota in order to make findings more reliable.

A further methodological limitation of this thesis identified in chapters 3-5, arose when analysing results of dietary intake pertaining to the consumption of first solid foods. With measures of dietary intake, it is common to use food frequency questionnaires (FFQ). The dietary measured completed at 6- and 12-months was in this style and covered many of the typically common food types presented to children when introducing the first solid foods. However, the detail of this questionnaire was not in

depth enough to parse out necessary dietary information relevant to microbiota composition. The original design of the longitudinal BIS study was to investigate direct relationships between GM composition and diversity, health and behaviour. The microbiome subset further aimed at investigating the influence of diet upon GM and allergy profiles. For this reason, the FFQ used was fit for purpose as it separately measured the common dietary allergens (Vuillermin et al., 2015). As described in chapter 1, section 1.2, dietary components such as protein type, white meat and fish, as opposed to red meat, influence GM composition (Shen et al., 2010). Furthermore, increased dietary fat intake is associated with increased inflammation and dietary fibre intake has been associated with specific bacteria such as *Collinsella*, and butyrate producing bacteria (Gomez-Arango et al., 2018; Zhao et al., 2018). In order to be able to distinguish these important dietary attributes, future research should consider using a diet diary, to establish types and quantities of foods consumed. It may also be possible to develop a specific dietary questionnaire that is designed to measure diet in a way that is pertinent to microbiota composition and diversity. In doing so it would also create a standardised tool for integrating diet and microbiota/microbiome research, creating greater consistency within the field.

A final limitation is the use of parent completed questionnaires within research. With regard to the measured of outcome variables, the overall method for measuring both temperament and behavioural outcomes has utilised parent completed questionnaires. The BIS does utilise a highly validated and reliable measure in the Strengths and Difficulties questionnaire (Goodman & Goodman, 2009), which has been found to be a good predictor of child mental health and behavioural problems. However, it is often questioned whether parents are the best reporters of their child's behaviours, due to bias in either a positive or negative direction, influence of the parent's own temperament, and other characteristics (Bayly & Gartstein, 2013). Furthermore, collection of dietary data is complex and has been found to be often inaccurate. Simplicity is often key with the use of Food Frequency Questionnaire (FFQ), and FFQs are inexpensive, and easy to administer in both paper format and online (Ayoubi et al., 2021). However, there is not currently a questionnaire that is designed to address the dietary factors necessary to fully explore the microbiota. This could be a potential future direction that would allow for the production of a tool that can be used to integrate dietary measures with microbiome/metabolome factors. Diet diaries have also been found to be a suitable alternative, however in longitudinal studies this may prove to be overburdening for the participant as indicated in chapter 6.

7.4 Conclusion.

Despite the previously mentioned limitations, this thesis has significantly contributed to the field of microbiota research concerned with investigating GM maturation and behaviour during early childhood. It has been established that it is possible to predict behavioural outcomes measured using the SDQ at 4-years, from GM composition and diversity measures taken throughout the first year of

life. In doing this, these studies confirm previous literature that highlights the first year of life as being a significant period of influence upon development of behaviour by the GM (Cowan et al., 2020). However, it is not possible to determine the exact mechanisms involved in this development without further investigation, which would benefit from using whole genome sequencing to investigate functional and taxonomic composition, as well as modelling the metabolic potential of the bacteria present. It is then possible to map further how these metabolic processes are influence by dietary intake and further influence the complex interaction with the GBA. Additionally, by using thorough statistical methodological techniques, the analyses in this thesis ruled out potential confounding variables that could otherwise potentially be attributed to contributing to the variation in SDQ scores. Future research should ensure that limitations established in section 7.3 of this chapter are addressed to establish a statistically validated sample size calculation process, and additionally establish a pipeline from faecal sample collection to microbiota analysis that is optimally suited to infant participant groups. Additionally, future research should include microbiota measures that are taken concurrently with behavioural measures, in order to confirm that the predictive results that can be seen from earlier measurements are accurate and cannot be better explained by current microbiota composition. This thesis would also benefit from a comparison of microbiota samples taken at all stages of maturation. By including measures at 4-years, it would be possible to establish the relationship with a GM that has reached maturation and resembles the adult enterotypes. In doing so it is possible to thoroughly explore whether variations in GM composition during sensitive periods of GM maturation have greater influence over behavioural variation or whether it can be explained by current GM composition.

In addition to the direct relationship between GM and behaviour, this thesis has provided novel insight into the role of dietary intake and potential mechanisms by which the GM composition and diversity influence behavioural outcomes measured at 4-years of age. It has been established that there is no significant mediation effect of GM composition or diversity, upon the relationship between dietary intake and behaviour. However, it was established that diet characterised by exclusivity of breastfeeding at 1-month of age, and by patterns of diet characterised by solid foods introduced at 6- and 12-months, was a significant moderator of the relationship between GM and behaviour measured using the SDQ at 4-years. Future research should continue to include dietary intake as a variable of interest to further enhance our understanding of the complex relationship between diet, GM, and behaviour as presented in Figure 7.2. In doing so it may be possible in the future to establish methods of influencing diet in order to establish best practice for solid food introduction or interventions that can potentially promote healthy behavioural development.

Aatsinki, A.-K., Lamichhane, S., Isokaanta, H., Sen, P., Krakstrom, M., Amaral Aves, M., . . . Perasto, L. E. (2023). Dynamics of Gut Metabolome and Microbiome Maturation during Early Life. *medRxiv*, 2023.2005. 2029.23290441.

Aatsinki, A. K., Lahti, L., Uusitupa, H. M., Munukka, E., Keskitalo, A., Nolvi, S., . . . Karlsson, L. (2019). Gut microbiota composition is associated with temperament traits in infants. *Brain Behavior and Immunity, 80*, 849-858. doi:10.1016/j.bbi.2019.05.035

Abildgaard, A., Elfving, B., Hokland, M., Wegener, G., & Lund, S. (2017). Probiotic treatment reduces depressive-like behaviour in rats independently of diet. *Psychoneuroendocrinology, 79*, 40-48.

Adesman, A., Soled, D., & Rosen, L. (2017). Formula Feeding as a Risk Factor for Attention-Deficit/Hyperactivity Disorder: Is Bisphenol A Exposure a Smoking Gun? *Journal of Developmental and Behavioral Pediatrics, 38*(7), 545-551. doi:10.1097/dbp.0000000000000468

Adlerberth, I., & Wold, A. E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatrica, 98*(2), 229-238. doi:10.1111/j.1651-2227.2008.01060.x

Agostoni, C., Decsi, T., Fewtrell, M., Goulet, O., Kolacek, S., Koletzko, B., . . . Nutr, E. C. (2008). Complementary feeding: A commentary by the ESPGHAN committee on nutrition. *Journal of Pediatric Gastroenterology and Nutrition, 46*(1), 99-110. doi:DOI 10.1097/01.mpg.0000304464.60788.bd

Aizawa, E., Tsuji, H., Asahara, T., Takahashi, T., Teraishi, T., Yoshida, S., . . . Kunugi, H. (2016). Possible association of Bifidobacterium and Lactobacillus in the gut microbiota of patients with major depressive disorder. *Journal of Affective Disorders, 202*, 254-257. doi:10.1016/j.jad.2016.05.038

Alcon-Giner, C., Caim, S., Mitra, S., Ketskemety, J., Wegmann, U., Wain, J., . . . Hall, L. J. (2017). Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *Bmc Genomics, 18*(1), 841. doi:10.1186/s12864-017-4229-x

Alimujiang, A., Colditz, G. A., Gardner, J. D., Park, Y., Berkey, C. S., & Sutcliffe, S. (2018). Childhood diet and growth in boys in relation to timing of puberty and adult height: the Longitudinal Studies of Child Health and Development. *Cancer Causes & Control, 29*(10), 915-926.

Allport, G. W. (1937). Personality: A psychological interpretation.

Alving-Jessep, E., Botchway, E., Wood, A. G., Hilton, A. C., & Blissett, J. M. (2022). The development of the gut microbiome and temperament during infancy and early childhood: A systematic review. *Developmental Psychobiology, 64*(7), e22306.

Anatolitou, F. (2012). Human milk benefits and breastfeeding. *Journal of Pediatric and Neonatal Individualized Medicine (JPNIM), 1*(1), 11-18.

Angelsen, N. K., Vik, T., Jacobsen, G., & Bakketeig, L. S. (2001). Breast feeding and cognitive development at age 1 and 5 years. *Archives of Disease in Childhood, 85*(3), 183-188. doi:DOI 10.1136/adc.85.3.183

Aparicio, E., Canals, J., Arija, V., De Henauw, S., & Michels, N. (2016). The role of emotion regulation in childhood obesity: implications for prevention and treatment. *Nutrition Research Reviews, 29*(1), 17-29.

Arden, M. A. (2010). Conflicting influences on UK mothers' decisions to introduce solid foods to their infants. *6*(2), 159-173.

Arenz, S., Rückerl, R., Koletzko, B., & von Kries, R. (2004). Breast-feeding and childhood obesity—a systematic review. *International journal of obesity, 28*(10), 1247-1256.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., . . . Consortium, M. (2011). Enterotypes of the human gut microbiome (vol 473, pg 174, 2011). *Nature, 474*(7353). doi:10.1038/nature10187

Asnicar, F., Berry, S. E., Valdes, A. M., Nguyen, L. H., Piccinno, G., Drew, D. A., . . . Khatib, H. A. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nature Medicine, 27*(2), 321-332.

Attar, N. (2015). Early life dysbiosis increases asthma risk. *Nature Reviews Microbiology, 13*(11), 660-660.

Avenevoli, S., Baio, J., Bitsko, R. H., Blumberg, S. J., Brody, D. J., Crosby, A., . . . Hedden, S. L. (2013). Mental health surveillance among children--United States, 2005-2011.

Ayoubi, S. S., Yaghoubi, Z., Pahlavani, N., Philippou, E., MalekAhmadi, M., Esmaily, H., . . . Norouzy, A. (2021). Developed and validated food frequency questionnaires in Iran: A systematic literature review. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences, 26*.

Azad, M. B., Becker, A. B., Guttman, D. S., Sears, M. R., Scott, J. A., Kozyrskyj, A. L., & Canadian Healthy Infant Longitudinal Development Study, I. (2013). Gut microbiota diversity and atopic disease: does breast-feeding play a role? *J Allergy Clin Immunol, 131*(1), 247-248. doi:10.1016/j.jaci.2012.10.044

Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., . . . Investigators, C. S. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ, 185*(5), 385-394. doi:10.1503/cmaj.121189

Azadbakht, L., & Esmaillzadeh, A. (2012). Dietary patterns and attention deficit hyperactivity disorder among Iranian children. *Nutrition, 28*(3), 242-249.

Bach Knudsen, K. E., Lærke, H. N., Hedemann, M. S., Nielsen, T. S., Ingerslev, A. K., Gundelund Nielsen, D. S., . . . Schioldan, A. G. (2018). Impact of diet-modulated butyrate production on intestinal barrier function and inflammation. *Nutrients, 10*(10), 1499.

Backhed, F., Fraser, C. M., Ringel, Y., Sanders, M. E., Sartor, R. B., Sherman, P. M., . . . Finlay, B. B. (2012). Defining a Healthy Human Gut Microbiome: Current Concepts, Future Directions, and Clinical Applications. *Cell Host & Microbe, 12*(5), 611-622. doi:10.1016/j.chom.2012.10.012

Baker, C. (2020). Mental health statistics for England: prevalence, services and funding.

Barbara, G., Stanghellini, V., Brandi, G., Cremon, C., Di Nardo, G., De Giorgio, R., & Corinaldesi, R. (2005). Interactions between commensal bacteria and gut sensorimotor function in health and disease. *Official journal of the American College of Gastroenterology| ACG, 100*(11), 2560-2568.

Bardanzellu, F., Fanos, V., & Reali, A. (2018). Human Breast Milk-Acquired Cytomegalovirus Infection: Certainties, Doubts And Perspectives. *Curr Pediatr Rev*. doi:10.2174/1573396315666181126105812

Barros-Santos, T., Silva, K. S. O., Libarino-Santos, M., Cata-Preta, E. G., Reis, H. S., Tamura, E. K., . . . Marinho, E. A. V. (2020). Effects of chronic treatment with new strains of Lactobacillus plantarum on cognitive, anxiety-and depressive-like behaviors in male mice. *Plos One, 15*(6), e0234037.

Bauer, G., Ewald, L., Hoffman, J., & Dubanoski, R. (1991). Breastfeeding and cognitive development of three-year-old children. *Psychological reports, 68*(3_suppl), 1218-1218.

Bayly, B., & Gartstein, M. (2013). Mother's and father's reports on their child's temperament: Does gender matter? *Infant behavior & development, 36*(1), 171.

Benoit, B., Laugerette, F., Plaisancie, P., Geloen, A., Bodennec, J., Estienne, M., . . . Michalski, M. C. (2015). Increasing fat content from 20 to 45 wt% in a complex diet induces lower endotoxemia in parallel with an increased number of intestinal goblet cells in mice. *Nutrition Research, 35*(4), 346-356. doi:10.1016/j.nutres.2015.01.005

Bergstrom, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., . . . Licht, T. R. (2014). Establishment of Intestinal Microbiota during Early Life: a Longitudinal, Explorative Study of a Large Cohort of Danish Infants. *Applied and Environmental Microbiology, 80*(9), 2889-2900. doi:10.1128/Aem.00342-14

Biol-Aquino, M. A., Perdiz, C. J., Borlagdan, M., Alcantara, J. D., & Mallillin, A. (2019). Differences in the bacterial profiles of infant gut by birth process, milk diet, and choice of 16S rRNA gene target region. *Human Microbiome Journal, 13*, 100062.

Bode, L. (2015). The functional biology of human milk oligosaccharides. *Early Hum Dev, 91*(11), 619-622. doi:10.1016/j.earlhumdev.2015.09.001

Boehm, G., & Stahl, B. (2007). Oligosaccharides from milk. *Journal of Nutrition, 137*(3), 847s-849s. Retrieved from <Go to ISI>://WOS:000244538300017

Bolling, K., Grant, C., Hamlyn, B., & Thornton, A. (2007). Infant feeding survey 2005. In: London: The Information Centre for Health and Social Care.

Borkent, J., Ioannou, M., Laman, J. D., Haarman, B. C., & Sommer, I. E. (2022). Role of the gut microbiome in three major psychiatric disorders. *Psychological Medicine, 52*(7), 1222-1242.

Borre, Y. E., O'Keeffe, G. W., Clarke, G., Stanton, C., Dinan, T. G., & Cryan, J. F. (2014). Microbiota and neurodevelopmental windows: implications for brain disorders. *Trends in Molecular Medicine, 20*(9), 509-518. doi:10.1016/j.molmed.2014.05.002

Bosheva, M., Tokodi, I., Krasnow, A., Pedersen, H. K., Lukjancenko, O., Eklund, A. C., . . . Cercamondi, C. I. (2022). Infant formula with a specific blend of five human milk oligosaccharides drives the gut microbiota development and improves gut maturation markers: A randomized controlled trial. *Frontiers in Nutrition, 9*, 920362.

Brinkworth, G. D., Noakes, M., Clifton, P. M., & Bird, A. R. (2009). Comparative effects of very low-carbohydrate, high-fat and high-carbohydrate, low-fat weight-loss diets on bowel habit and faecal short-chain fatty acids and bacterial populations. *British Journal of Nutrition, 101*(10), 1493-1502. doi:10.1017/S0007114508094658

Brown, A., & Rowan, H. (2016). Maternal and infant factors associated with reasons for introducing solid foods. *Maternal and Child Nutrition, 12*(3), 500-515. doi:10.1111/mcn.12166

Burgess, J. A., Dharmage, S. C., Allen, K., Koplin, J., Garcia-Larsen, V., Boyle, R., . . . Lodge, C. J. (2019). Age at introduction to complementary solid food and food allergy and sensitization: A systematic review and meta-analysis. *Clinical & Experimental Allergy, 49*(6), 754-769.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal, 11*(12), 2639-2643.

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., & Holmes, S. P. (2016). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research, 5*.

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology, 28*(2), 203-209. Retrieved from <Go to ISI>://WOS:000359016600007

Carlson, A. L., Xia, K., Azcarate-Peril, M. A., Goldman, B. D., Ahn, M., Styner, M. A., . . . Knickmeyer, R. C. (2018). Infant Gut Microbiome Associated With Cognitive Development. *Biol Psychiatry, 83*(2), 148-159. doi:10.1016/j.biopsych.2017.06.021

Chaturvedi, P., Warren, C. D., Altaye, M., Morrow, A. L., Ruiz-Palacios, G., Pickering, L. K., & Newburg, D. S. (2001). Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. *Glycobiology, 11*(5), 365-372. doi:10.1093/glycob/11.5.365

Chen, L.-W., Xu, J., Soh, S. E., Aris, I. M., Tint, M.-T., Gluckman, P. D., & Tan, K. H. (2020). Implication of gut microbiota in the association between infant antibiotic exposure and childhood obesity and adiposity accumulation. *International journal of obesity, 44*(7), 1508-1521.

Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports, 5*(1), 1-10.

Christian, L. M., Galley, J. D., Hade, E. M., Schoppe-Sullivan, S., Dush, C. K., & Bailey, M. T. (2015). Gut microbiome composition is associated with temperament during early childhood. *Brain Behavior and Immunity, 45*, 118-127. doi:10.1016/j.bbi.2014.10.018

Clarke, G., Grenham, S., Scully, P., Fitzgerald, P., Moloney, R., Shanahan, F., . . . Cryan, J. (2013). The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Molecular Psychiatry, 18*(6), 666-673.

Clarke, G., O'Mahony, S. M., Dinan, T. G., & Cryan, J. F. (2014). Priming for health: gut microbiota acquired in early life regulates physiology, brain and behaviour. *Acta Paediatrica, 103*(8), 812-819. doi:10.1111/apa.12674

Cloherty, M., Alexander, J., & Holloway, I. (2004). Supplementing breast-fed babies in the UK to protect their mothers from tiredness or distress. *Midwifery, 20*(2), 194-204. doi:10.1016/j.midw.2003.09.002

Cong, X., Xu, W., Janton, S., Henderson, W. A., Matson, A., McGrath, J. M., . . . Graf, J. (2016). Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender. *Plos One, 11*(4), e0152751. doi:10.1371/journal.pone.0152751

Costantini, I., Paul, E., Caldwell, D. M., López-López, J. A., & Pearson, R. M. (2020). Protocol for a systematic review and network meta-analysis of randomised controlled trials examining the effectiveness of early parenting interventions in preventing internalising problems in children and adolescents. *Systematic Reviews, 9*(1), 1-17.

Cowan, C. S. M., Dinan, T. G., & Cryan, J. F. (2020). Annual research review: Critical windows – the microbiota–gut–brain axis in neurocognitive development. *Journal of Child Psychology and Psychiatry, 61*(3), 353-371. doi:10.1111/jcpp.13156

Cresci, G. A., & Bawden, E. (2015). Gut Microbiome: What We Do and Don't Know. *Nutr Clin Pract, 30*(6), 734-746. doi:10.1177/0884533615609899

Crozier, S. R., Robinson, S. M., Borland, S. E., & Inskip, H. M. (2006). Dietary patterns in the Southampton Women's Survey. *European journal of clinical nutrition, 60*(12), 1391-1399.

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience, 13*(10), 701-712. doi:10.1038/nrn3346

Daly, M., Sutin, A. R., & Robinson, E. (2022). Longitudinal changes in mental health and the COVID-19 pandemic: Evidence from the UK Household Longitudinal Study. *Psychological Medicine, 52*(13), 2549-2558.

Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C., & Gilad, Y. (2014). Seasonal variation in human gut microbiome composition. *Plos One, 9*(3), e90731.

Davis, E. C., Dinsmoor, A. M., Wang, M., & Donovan, S. M. (2020). Microbiome Composition in Pediatric Populations from Birth to Adolescence: Impact of Diet and Prebiotic and Probiotic Interventions. *Dig Dis Sci, 65*(3), 706-722. doi:10.1007/s10620-020-06092-x

Davis, E. C., Wang, M., & Donovan, S. M. (2017). The role of early life nutrition in the establishment of gastrointestinal microbial composition and function. *Gut Microbes, 8*(2), 143-171. doi:10.1080/19490976.2016.1278104

Davis, E. P., Glynn, L. M., Schetter, C. D., Hobel, C., Chicz-Demet, A., & Sandman, C. A. (2007). Prenatal exposure to maternal depression and cortisol influences infant temperament. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(6), 737-746.

De Crescenzo, F., Lennox, A., Gibson, J., Cordey, J., Stockton, S., Cowen, P., & Quested, D. (2017). Melatonin as a treatment for mood disorders: a systematic review. *Acta Psychiatrica Scandinavica, 136*(6), 549-558.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., . . . Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A, 107*(33), 14691-14696. doi:10.1073/pnas.1005963107

de Vos, W. M., Tilg, H., Van Hul, M., & Cani, P. D. (2022). Gut microbiome and health: mechanistic insights. *Gut, 71*(5), 1020-1032.

de Weerth, C. (2017). Do bacteria shape our development? Crosstalk between intestinal microbiota and HPA axis. *Neuroscience and Biobehavioral Reviews, 83*, 458-471. doi:10.1016/j.neubiorev.2017.09.016

Del-Ponte, B., Quinte, G. C., Cruz, S., Grellert, M., & Santos, I. S. (2019). Dietary patterns and attention deficit/hyperactivity disorder (ADHD): a systematic review and meta-analysis. *Journal of Affective Disorders, 252*, 160-173.

Dinan, T. G., Stilling, R. M., Stanton, C., & Cryan, J. F. (2015). Collective unconscious: How gut microbes shape human behavior. *Journal of Psychiatric Research, 63*, 1-9. doi:10.1016/j.jpsychires.2015.02.021

Dogra, S., Sakwinska, O., Soh, S.-E., Ngom-Bru, C., Brück, W. M., Berger, B., . . . Chong, Y.-S. (2015). Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *Mbio, 6*(1), e02419-02414.

Donovan, S. M., & Comstock, S. S. (2016). Human Milk Oligosaccharides Influence Neonatal Mucosal and Systemic Immunity. *Ann Nutr Metab, 69 Suppl 2*(Suppl 2), 42-51. doi:10.1159/000452818

Du Toit, A. (2019). The gut microbiome and mental health. *Nature Reviews Microbiology, 17*(4), 196-196.

Easterbrooks, M. A., Bureau, J.-F., & Lyons-Ruth, K. (2012). Developmental correlates and predictors of emotional availability in mother–child interaction: A longitudinal study from infancy to middle childhood. *Development and Psychopathology, 24*(1), 65-78.

Fallani, M., Amarri, S., Uusijarvi, A., Adam, R., Khanna, S., Aguilera, M., . . . Young, D. (2011a). Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology, 157*(5), 1385-1392.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., . . . Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science, 352*(6285), 560-564. doi:10.1126/science.aad3503

Fan, W., Huo, G., Li, X., Yang, L., Duan, C., Wang, T., & Chen, J. (2013). Diversity of the intestinal microbiota in different patterns of feeding infants by Illumina high-throughput sequencing. *World J Microbiol Biotechnol, 29*(12), 2365-2372. doi:10.1007/s11274-013-1404-3

Fang, X., Lloyd, C. J., & Palsson, B. O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology, 18*(12), 731-743.

Fanos, V., Reali, A., Marcialis, M. A., & Bardanzellu, F. (2018). What you have to know about Human Milk Oligosaccharides. *Journal of Pediatric and Neonatal Individualized Medicine, 7*(1). doi:10.7363/070137

Farrow, C., & Blissett, J. (2006). Maternal cognitions, psychopathologic symptoms, and infant temperament as predictors of early infant feeding problems: A longitudinal study. *International Journal of Eating Disorders, 39*(2), 128-134. doi:10.1002/eat.20220

Flannery, J. E., Stagaman, K., Burns, A. R., Hickey, R. J., Roos, L. E., Giuliano, R. J., . . . Sharpton, T. J. (2020). Gut Feelings Begin in Childhood: the Gut Metagenome Correlates with Early Environment, Caregiving, and Behavior. *Mbio, 11*(1). doi:10.1128/mBio.02780-19

Foster, J. A., & McVey Neufeld, K. A. (2013). Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci, 36*(5), 305-312. doi:10.1016/j.tins.2013.01.005

Fox, M., Lee, S. M., Wiley, K. S., Lagishetty, V., Sandman, C., Jacobs, J. P., & Glynn, L. M. (2021). Development of the infant gut microbiome and temperament across the first year of life. *American Journal of Human Biology, 33*. Retrieved from <Go to ISI>://WOS:000636277500038

Frankiensztajn, L. M., Elliott, E., & Koren, O. (2020). The microbiota and the hypothalamus-pituitary-adrenocortical (HPA) axis, implications for anxiety and stress disorders. *Current Opinion in Neurobiology, 62*, 76-82. doi:10.1016/j.conb.2019.12.003

Frese, S. A., Hutton, A. A., Contreras, L. N., Shaw, C. A., Palumbo, M. C., Casaburi, G., . . . Henrick, B. M. (2017). Persistence of supplemented Bifidobacterium longum subsp. infantis EVC001 in breastfed infants. *Msphere, 2*(6), e00501-00517.

Gabert, L., Vors, C., Louche-Pélissier, C., Sauvinet, V., Lambert-Porcheron, S., Drai, J., . . . Michalski, M. C. (2011). 13C tracer recovery in human stools after digestion of a fat-rich meal labelled with [1, 1, 1-13C3] tripalmitin and [1, 1, 1-13C3] triolein. *Rapid Communications in Mass Spectrometry, 25*(19), 2697-2703.

Gallè, F., Valeriani, F., Cattaruzza, M. S., Gianfranceschi, G., Liguori, R., Antinozzi, M., . . . Romano Spica, V. (2020). Mediterranean diet, physical activity and gut microbiome composition: a cross-sectional study among healthy young Italian adults. *Nutrients, 12*(7), 2164.

Gao, W., Salzwedel, A. P., Carlson, A. L., Xia, K., Azcarate-Peril, M. A., Styner, M. A., . . . Knickmeyer, R. C. (2019). Gut microbiome and brain functional connectivity in infants-a preliminary study focusing on the amygdala. *Psychopharmacology, 236*(5), 1641-1651. doi:10.1007/s00213-018-5161-8

Gartstein, M. A., & Skinner, M. K. (2018). Prenatal influences on temperament development: The role of environmental epigenetics. *Development and Psychopathology, 30*(4), 1269-1303.

Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., & Blumberg, S. J. (2019). Prevalence and treatment of depression, anxiety, and conduct problems in US children. *The Journal of pediatrics, 206*, 256-267. e253.

Golding, J., Emmett, P. M., & Rogers, I. S. (1997). Breast feeding and infant mortality. *Early Hum Dev, 49 Suppl*, S143-155. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9363423

Gomez-Arango, L. F., Barrett, H. L., Wilkinson, S. A., Callaway, L. K., McIntyre, H. D., Morrison, M., & Dekker Nitert, M. (2018). Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. *Gut Microbes, 9*(3), 189-201.

Gomez-Arango, L. F., Barrett, H. L., Callaway, L. K., & Nitert, M. D. (2015). Probiotics and pregnancy. *Curr Diab Rep, 15*(1), 567. doi:10.1007/s11892-014-0567-0

Gomez-Llorente, C., Plaza-Diaz, J., Aguilera, M., Munoz-Quezada, S., Bermudez-Brito, M., Peso-Echarri, P., . . . Gil, A. (2013). Three main factors define changes in fecal microbiota associated with feeding modality in infants. *J Pediatr Gastroenterol Nutr, 57*(4), 461-466. doi:10.1097/MPG.0b013e31829d519a

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 400-403.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of abnormal child psychology, 38*(8), 1179-1191.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586.

Grummer-Strawn, L. M., Scanlon, K. S., & Fein, S. B. (2008). Infant feeding and feeding transitions during the first year of life. *Pediatrics, 122*, S36-S42. doi:10.1542/peds.2008-1315d

Hair, A. B., Peluso, A. M., Hawthorne, K. M., Perez, J., Smith, D. P., Khan, J. Y., . . . Abrams, S. A. (2016). Beyond necrotizing enterocolitis prevention: improving outcomes with an exclusive human milk–based diet. *Breastfeeding Medicine, 11*(2), 70-74.

Hamlyn, B. (2002). *Infant feeding 2000: A survey conducted on behalf of the Department of Health, the Scottish Executive, the National Assembly for Wales and the Department of Health, Social Services and Public Safety in Northern Ireland*: TSO.

Harder, T., Bergmann, R., Kallischnigg, G., & Plagemann, A. (2005). Duration of breastfeeding and risk of overweight: a meta-analysis. *American journal of epidemiology, 162*(5), 397-403.

Harrison, E., Drake, T., Ots, R., & Harrison, M. E. (2020). Package “finalfit.”. *Retrieved February, 29*, 2020.

Haszard, J. J., Russell, C. G., Byrne, R. A., Taylor, R. W., & Campbell, K. J. (2019). Early maternal feeding practices: Associations with overweight later in childhood. *Appetite, 132*, 91-96. doi:10.1016/j.appet.2018.10.008

Haycraft, E., Farrow, C., Meyer, C., Powell, F., & Blissett, J. (2011). Relationships between temperament and eating behaviours in young children. *Appetite, 56*(3), 689-692.

Health, N. I. o. (2018). Quality assessment tool for observational cohort and cross-sectional studies. National heart, lung, and blood Institute, 2014. In.

Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., . . . Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences, 108*(7), 3047-3052.

Hellsing, C., Örtqvist, A. K., Hagel, E., Mesas-Burgos, C., Gustafsson, U. O., & Granström, A. L. (2022). Delivery mode and risk of gastrointestinal disease in the offspring. *Acta Obstetricia et Gynecologica Scandinavica, 101*(10), 1146-1152.

Henrick, B. M., Hutton, A. A., Palumbo, M. C., Casaburi, G., Mitchell, R. D., Underwood, M. A., . . . Frese, S. A. (2018). Elevated Fecal pH Indicates a Profound Change in the Breastfed Infant Gut Microbiome Due to Reduction of Bifidobacterium over the Past Century. *Msphere, 3*(2). doi:ARTN e00041-18

Hidalgo-Cantabrana, C., Delgado, S., Ruiz, L., Ruas-Madiedo, P., Sánchez, B., & Margolles, A. (2018). Bifidobacteria and their health-promoting effects. *Bugs as Drugs: Therapeutic Microbes for the Prevention and Treatment of Disease*, 73-98.

Homann, C.-M., Rossel, C. A., Dizzell, S., Bervoets, L., Simioni, J., Li, J., . . . Mommers, M. (2021). Infants’ first solid foods: impact on gut microbiota development in two intercontinental cohorts. *Nutrients, 13*(8), 2639.

Horwood, L. J., & Fergusson, D. M. (1998). Breastfeeding and later cognitive and academic outcomes. *Pediatrics, 101*(1). DOI 10.1542/peds.101.1.e9

Hoskinson, C., Dai, D. L., Del Bel, K. L., Becker, A. B., Moraes, T. J., Mandhane, P. J., . . . Azad, M. B. (2023). Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease. *Nature Communications, 14*(1), 4785.

Huh, S. Y., Rifas-Shiman, S. L., Taveras, E. M., Oken, E., & Gillman, M. W. (2011). Timing of Solid Food Introduction and Risk of Obesity in Preschool-Aged Children. *Pediatrics, 127*(3), E544-E551. doi:10.1542/peds.2010-0740

Hupe, M. (2019). EndNote X9. *Journal of Electronic Resources in Medical Libraries, 16*(3-4), 117-119.

Integrative, H. (2014). The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host & Microbe, 16*(3), 276-289.

Jaimes, J. D., Slavíčková, A., Hurych, J., Cinek, O., Nichols, B., Vodolánová, L., . . . Havlík, J. (2021). Stool metabolome-microbiota evaluation among children and adolescents with obesity, overweight, and normal-weight using 1H NMR and 16S rRNA gene profiling. *Plos One, 16*(3), e0247378.

Jarman, M., Edwards, K., & Blissett, J. (2022). Influences on the dietary intakes of preschool children: A systematic scoping review. *International Journal of Behavioral Nutrition and Physical Activity, 19*(1), 1-15.

Jenkins, T. A., Nguyen, J. C., Polglaze, K. E., & Bertrand, P. P. (2016). Influence of tryptophan and serotonin on mood and cognition with a possible role of the gut-brain axis. *Nutrients, 8*(1), 56.

Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., . . . Shi, J. (2015). Altered fecal microbiota composition in patients with major depressive disorder. *Brain, behavior, and immunity, 48*, 186-194.

Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., . . . Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in microbiology, 159*(3), 187-193.

Jones, N. A., & Sloan, A. (2018). Neurohormones and temperament interact during infant development. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1744), 20170159.

Jost, T., Lacroix, C., Braegger, C. P., & Chassard, C. (2012). New insights in gut microbiota establishment in healthy breast fed neonates. *Plos One, 7*(8), e44595. doi:10.1371/journal.pone.0044595

Jost, T., Lacroix, C., Braegger, C. P., Rochat, F., & Chassard, C. (2014). Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environmental Microbiology, 16*(9), 2891-2904. doi:10.1111/1462-2920.12238

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., . . . Madsen, K. L. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology, 7*, 459.

Kämpfer, P., Glaeser, S. P., Raza, M. W., Abbasi, S. A., & Perry, J. D. (2014). Pseudocitrobacter gen. nov., a novel genus of the Enterobacteriaceae with two new species Pseudocitrobacter faecalis sp. nov., and Pseudocitrobacter anthropi sp. nov, isolated from fecal samples from hospitalized patients in Pakistan. *Systematic and applied microbiology, 37*(1), 17-22.

Kaplan, B. J., Giesbrecht, G. F., Leung, B. M., Field, C. J., Dewey, D., Bell, R. C., . . . Pop, V. J. (2014). The Alberta Pregnancy Outcomes and Nutrition (APrON) cohort study: rationale and methods. *Maternal & child nutrition, 10*(1), 44-60.

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H.-M., Korja, R., Ekholm, E., . . . Paunio, T. (2018). Cohort profile: the FinnBrain birth cohort study (FinnBrain). *International Journal of Epidemiology, 47*(1), 15-16j.

Kashtanova, D. A., Popenko, A. S., Tkacheva, O. N., Tyakht, A. B., Alexeev, D. G., & Boytsov, S. A. (2016). Association between the gut microbiota and diet: Fetal life, early childhood, and further life. *Nutrition, 32*(6), 620-627. doi:10.1016/j.nut.2015.12.037

Kelly, J. R., Borre, Y., Brien, C. O., Patterson, E., El Aidy, S., Deane, J., . . . Dinan, T. G. (2016). Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *Journal of Psychiatric Research, 82*, 109-118. doi:10.1016/j.jpsychires.2016.07.019

Kelsey, C. M., Prescott, S., McCulloch, J. A., Trinchieri, G., Valladares, T. L., Dreisbach, C., . . . Grossmann, T. (2021). Gut microbiota composition is associated with newborn functional brain connectivity and behavioral temperament. *Brain Behavior and Immunity, 91*, 472-486. doi:10.1016/j.bbi.2020.11.003

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development, 40*(1), 64-75.

Kinlein, S. A., Phillips, D. J., Keller, C. R., & Karatsoreos, I. N. (2019). Role of corticosterone in altered neurobehavioral responses to acute stress in a model of compromised hypothalamic-pituitary-adrenal axis function. *Psychoneuroendocrinology, 102*, 248-255.

Kitaoka, M. (2012). Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides. *Adv Nutr, 3*(3), 422s-429s. doi:10.3945/an.111.001420

Knickmeyer, R. C., Gouttard, S., Kang, C. Y., Evans, D., Wilber, K., Smith, J. K., . . . Gilmore, J. H. (2008). A Structural MRI Study of Human Brain Development from Birth to 2 Years. *Journal of Neuroscience, 28*(47), 12176-12182. doi:10.1523/Jneurosci.3479-08.2008

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., . . . Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 4578-4585. doi:10.1073/pnas.1000081107

Kong, Q., Tian, P., Zhao, J., Zhang, H., Wang, G., & Chen, W. (2021). The autistic-like behaviors development during weaning and sexual maturation in VPA-induced autistic-like rats is accompanied by gut microbiota dysbiosis. *Peerj, 9*, e11103.

Korpela, K. (2021). Impact of delivery mode on infant gut microbiota. *Annals of Nutrition and Metabolism, 77*(3), 11-19.

Kramer, M. S., & Kakuma, R. (2012). Optimal duration of exclusive breastfeeding. *Cochrane Database Syst Rev*(8), CD003517. doi:10.1002/14651858.CD003517.pub2

Kretschmer, T., Hickman, M., Doerner, R., Emond, A., Lewis, G., Macleod, J., . . . Heron, J. (2014). Outcomes of childhood conduct problem trajectories in early adulthood: findings from the ALSPAC study. *European Child & Adolescent Psychiatry, 23*(7), 539-549.

La Rosa, P. S., Brooks, J. P., Deych, E., Boone, E. L., Edwards, D. J., Wang, Q., . . . Shannon, W. D. (2012). Hypothesis testing and power calculations for taxonomic-based human microbiome data. *Plos One, 7*(12), e52078.

Lahey, B. B., Lee, S. S., Sibley, M. H., Applegate, B., Molina, B. S., & Pelham, W. E. (2016). Predictors of adolescent outcomes among 4–6-year-old children with attention-deficit/hyperactivity disorder. *Journal of abnormal psychology, 125*(2), 168.

Lanfumey, L., Mongeau, R., & Hamon, M. (2013). Biological rhythms and melatonin in mood disorders and their treatments. *Pharmacology & therapeutics, 138*(2), 176-184.

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature, 444*(7122), 1022-1023. doi:10.1038/4441022a

Li, Y., Ren, L., Wang, Y., Li, J., Zhou, Q., Peng, C., . . . Shen, X. (2022). The Effect of Breast Milk Microbiota on the Composition of Infant Gut Microbiota: A Cohort Study. *Nutrients, 14*(24), 5397.

Ling, J., & Zahry, N. R. (2021). Relationships among perceived stress, emotional eating, and dietary intake in college students: Eating self-regulation as a mediator. *Appetite, 163*, 105215.

Lipsanen, J., Elovainio, M., Hakulinen, C., Tremblay, M. S., Rovio, S., Lagström, H., . . . Viikari, J. (2020). Temperament profiles are associated with dietary behavior from childhood to adulthood. *Appetite, 151*, 104681.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association, 83*(404), 1198-1202.

Liu, R. T., Rowan-Nash, A. D., Sheehan, A. E., Walsh, R. F., Sanzari, C. M., Korry, B. J., & Belenky, P. (2020). Reductions in anti-inflammatory gut bacteria are associated with depression in a sample of young adults. *Brain, behavior, and immunity, 88*, 308-324.

Loman, M. M., & Gunnar, M. R. (2010). Early experience and the development of stress reactivity and regulation in children. *Neuroscience & biobehavioral reviews, 34*(6), 867-876.

Loughman, A., Ponsonby, A.-L., O'Hely, M., Symeonides, C., Collier, F., Tang, M. L., . . . Pezic, A. (2020). Gut microbiota composition during infancy and subsequent behavioural outcomes. *EBioMedicine, 52*, 102640.

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology, 19*(1), 29-41.

Luczynski, P., Whelan, S. O., O'Sullivan, C., Clarke, G., Shanahan, F., Dinan, T. G., & Cryan, J. F. (2016). Adult microbiota-deficient mice have distinct dendritic morphological changes: differential effects in the amygdala and hippocampus. *European Journal of Neuroscience, 44*(9), 2654-2666. doi:10.1111/ejn.13291

Luque, V., Closa-Monasterolo, R., Escribano, J., & Ferre, N. (2015). Early Programming by Protein Intake: The Effect of Protein on Adiposity Development and the Growth and Functionality of Vital Organs. *Nutrition and Metabolic Insights, 8*, 49-56. doi:10.4137/Nmi.S29525

Ma, B., Liang, J., Dai, M., Wang, J., Luo, J., Zhang, Z., & Jing, J. (2019). Altered Gut Microbiota in Chinese Children With Autism Spectrum Disorders. *Front Cell Infect Microbiol, 9*, 40. doi:10.3389/fcimb.2019.00040

Ma, X., Ding, J., Ren, H., Xin, Q., Li, Z., Han, L., . . . Ren, Z. (2023). Distinguishable influence of the delivery mode, feeding pattern, and infant sex on dynamic alterations in the intestinal microbiota in the first year of life. *Microbial Ecology*, 1-15.

Mackie, R. I., Sghir, A., & Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *American Journal of Clinical Nutrition, 69*(5), 1035S-1045S. doi:10.1093/ajcn/69.5.1035s

Madan, A., Thompson, D., Fowler, J. C., Ajami, N., Salas, R., Frueh, B., . . . Petrosino, J. (2020). The gut microbiota is associated with psychiatric symptom severity and treatment outcome among individuals with serious mental illness. *Journal of Affective Disorders, 264*, 98-106.

Madan, J. C., Hoen, A. G., Lundgren, S. N., Farzan, S. F., Cottingham, K. L., Morrison, H. G., . . . Karagas, M. R. (2016). Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants. *JAMA Pediatr, 170*(3), 212-219. doi:10.1001/jamapediatrics.2015.3732

Magnusson, K. R., Hauck, L., Jeffrey, B. M., Elias, V., Humphrey, A., Nath, R., . . . Bermudez, L. E. (2015). Relationships between Diet-Related Changes in the Gut Microbiome and Cognitive Flexibility. *Neuroscience, 300*, 128-140. doi:10.1016/j.neuroscience.2015.05.016

Maharshak, N., Ringel, Y., Katibian, D., Lundqvist, A., Sartor, R. B., Carroll, I. M., & Ringel-Kulka, T. (2018). Fecal and mucosa-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Digestive Diseases and Sciences, 63*(7), 1890-1899.

Marlicz, W., Poniewierska-Baran, A., Rzeszotek, S., Bartoszewski, R., Skonieczna-Zydecka, K., Starzynska, T., & Ratajczak, M. Z. (2018). A novel potential role of pituitary gonadotropins in the pathogenesis of human colorectal cancer. *Plos One, 13*(3). doi:ARTN e0189337

Marques, T. M., Wall, R., Ross, R. P., Fitzgerald, G. F., Ryan, C. A., & Stanton, C. (2010). Programming infant gut microbiota: influence of dietary and environmental factors. *Current Opinion in Biotechnology, 21*(2), 149-156. doi:10.1016/j.copbio.2010.03.020

Martin, M. A., & Sela, D. A. (2013). Infant gut microbiota: developmental influences and health outcomes. In *Building babies* (pp. 233-256): Springer.

Martinez, K. B., Leone, V., & Chang, E. B. (2017). Western diets, gut dysbiosis, and metabolic diseases: Are they linked? *Gut Microbes, 8*(2), 130-142.

Mason, B. L., Li, Q., Minhajuddin, A., Czysz, A. H., Coughlin, L. A., Hussain, S. K., . . . Trivedi, M. H. (2020). Reduced anti-inflammatory gut microbiota are associated with depression and anhedonia. *J Affect Disord, 266*, 394-401. doi:10.1016/j.jad.2020.01.137

Matsuyama, M., Gomez-Arango, L. F., Fukuma, N. M., Morrison, M., Davies, P. S. W., & Hill, R. J. (2018). Breastfeeding: a key modulator of gut microbiota characteristics in late infancy. *Journal of Developmental Origins of Health and Disease*, 1-8. doi:10.1017/s2040174418000624

McAndrew, F., Thompson, J., Fellows, L., Large, A., Speed, M., & Renfrew, M. J. (2012). Infant feeding survey 2010. *Leeds: health and social care information Centre, 2*(1).

McCrae, R. R., Costa Jr, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., . . . Woodfield, R. (2000). Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology, 78*(1), 173.

McGuire, M. K., Meehan, C. L., McGuire, M. A., Williams, J. E., Foster, J., Sellen, D. W., . . . Bode, L. (2017). What's normal? Oligosaccharide concentrations and profiles in milk produced by healthy women vary geographically. *American Journal of Clinical Nutrition, 105*(5), 1086-1100. doi:10.3945/ajcn.116.139980

McMeekin, S., Jansen, E., Mallan, K., Nicholson, J., Magarey, A., & Daniels, L. (2013). Associations between infant temperament and early feeding practices. A cross-sectional study of Australian mother-infant dyads from the NOURISH randomised controlled trial. *Appetite, 60*, 239-245. doi:10.1016/j.appet.2012.10.005

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *Plos One, 8*(4), e61217.

Meltzer, H., Gatward, R., Goodman, R., & Ford, T. (2003). Mental health of children and adolescents in Great Britain. *International Review of Psychiatry, 15*(1-2), 185-187.

Michels, N. (2019). Biological underpinnings from psychosocial stress towards appetite and obesity during youth: research implications towards metagenomics, epigenomics and metabolomics. *Nutrition Research Reviews, 32*(2), 282-293. doi:10.1017/s0954422419000143

Miller, E. M., Aiello, M. O., Fujita, M., Hinde, K., Milligan, L., & Quinn, E. A. (2013). Field and laboratory methods in human milk research. *American Journal of Human Biology, 25*(1), 1-11. doi:10.1002/ajhb.22334

Misiak, B., Loniewski, I., Marlicz, W., Frydecka, D., Szulc, A., Rudzki, L., & Samochowiec, J. (2020). The HPA axis dysregulation in severe mental illness: Can we shift the blame to gut microbiota? *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 102*. doi:10.1016/j.pnpbp.2020.109951

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2011). Preferred reporting items of systematic review and meta-analyses: the PRISMA statement. *Deutsche Medizinische Wochenschrift, 136*(15).

Moossavi, S., Miliku, K., Sepehri, S., Khafipour, E., & Azad, M. B. (2018). The Prebiotic and Probiotic Properties of Human Milk: Implications for Infant Immune Development and Pediatric Asthma. *Frontiers in Pediatrics, 6*, 197. doi:10.3389/fped.2018.00197

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research—Understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research, 44*(2), 465-476.

Muris, P., & Ollendick, T. H. (2005). The role of temperament in the etiology of child psychopathology. *Clinical Child and Family Psychology Review, 8*(4), 271-289. doi:10.1007/s10567-005-8809-y

Murugesan, S., Nirmalkar, K., Hoyo-Vadillo, C., García-Espitia, M., Ramírez-Sánchez, D., & García-Mena, J. (2018). Gut microbiome production of short-chain fatty acids and obesity in children. *Eur J Clin Microbiol Infect Dis, 37*(4), 621-625. doi:10.1007/s10096-017-3143-0

Ness, A., Maynard, M., Frankel, S., Smith, G. D., Frobisher, C., Leary, S., . . . Gunnell, D. (2005). Diet in childhood and adult cardiovascular and all cause mortality: the Boyd Orr cohort. *Heart, 91*(7), 894-898.

Nikolova, V. L., Hall, M. R., Hall, L. J., Cleare, A. J., Stone, J. M., & Young, A. H. (2021). Perturbations in gut microbiota composition in psychiatric disorders: a review and meta-analysis. *JAMA psychiatry, 78*(12), 1343-1354.

O'Callaghan, A., & van Sinderen, D. (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Frontiers in Microbiology, 7*. doi:ARTN 925

O'Mahony, S. M., Clarke, G., Dinan, T. G., & Cryan, J. F. (2017). Early-life adversity and brain development: Is the microbiome a missing piece of the puzzle? *Neuroscience, 342*, 37-54. doi:10.1016/j.neuroscience.2015.09.068

Ogra, P. L., Rassin, D. K., & Garofalo, R. (2006). Human milk. In *Infectious diseases of the fetus and newborn infant* (pp. 211-243): Elsevier Inc.

Olsson, L. M., Boulund, F., Nilsson, S., Khan, M. T., Gummesson, A., Fagerberg, L., . . . Bergström, G. (2022). Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host & Microbe, 30*(5), 726-739. e723.

Oozeer, R., van Limpt, K., Ludwig, T., Ben Amor, K., Martin, R., Wind, R. D., . . . Knol, J. (2013). Intestinal microbiology in early life: specific prebiotics can have similar functionalities as human-milk oligosaccharides. *American Journal of Clinical Nutrition, 98*(2), 561s-571s. doi:10.3945/ajcn.112.038893

Operto, F. F., Smirni, D., Scuoppo, C., Padovano, C., Vivenzio, V., Quatrosi, G., . . . Pastorino, G. M. G. (2021). Neuropsychological profile, emotional/behavioral problems, and parental stress in children with neurodevelopmental disorders. *Brain Sciences, 11*(5), 584.

Organization, W. H. (2017). Guidance on ending the inappropriate promotion of foods for infants and young children: implementation manual.

Ou, J. H., DeLany, J. P., Zhang, M., Sharma, S., & O'Keefe, S. J. D. (2012). Association Between Low Colonic Short-Chain Fatty Acids and High Bile Acids in High Colon Cancer Risk Populations. *Nutrition and Cancer-an International Journal, 64*(1), 34-40. doi:10.1080/01635581.2012.630164

Ou, Y., Belzer, C., Smidt, H., & de Weerth, C. (2022). Development of the gut microbiota in healthy children in the first ten years of life: associations with internalizing and externalizing behavior. *Gut Microbes, 14*(1), 2038853.

Paans, N. P., Gibson-Smith, D., Bot, M., van Strien, T., Brouwer, I. A., Visser, M., & Penninx, B. W. (2019). Depression and eating styles are independently associated with dietary intake. *Appetite, 134*, 103-110.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol, 5*(7), e177. doi:10.1371/journal.pbio.0050177

Paolicelli, R. C., Bergamini, G., & Rajendran, L. (2019). Cell-to-cell Communication by Extracellular Vesicles: Focus on Microglia. *Neuroscience, 405*, 148-157. doi:10.1016/j.neuroscience.2018.04.003

Papoutsou, S., Savva, S. C., Hunsberger, M., Jilani, H., Michels, N., Ahrens, W., . . . Consortium, I. (2018). Timing of solid food introduction and association with later childhood overweight and obesity: The IDEFICS study. *Maternal and Child Nutrition, 14*(1). doi:ARTN e12471

Parrett, A. M., & Edwards, C. A. (1997). In vitro fermentation of carbohydrate by breast fed and formula fed infants. *Archives of Disease in Childhood, 76*(3), 249-253. doi:DOI 10.1136/adc.76.3.249

Patrick, H., Nicklas, T. A., Hughes, S. O., & Morales, M. (2005). The benefits of authoritative feeding style: caregiver feeding styles and children's food consumption patterns. *Appetite, 44*(2), 243-249. doi:10.1016/j.appet.2002.07.001

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., . . . Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics, 118*(2), 511-521. doi:10.1542/peds.2005-2824

Penders, J., & van Best, N. (2022). The Development of the Gut Microbiota in Childhood and Its Distortion by Lifestyle Changes. In *Evolution, Biodiversity and a Reassessment of the Hygiene Hypothesis* (pp. 197-219): Springer.

Perez-Munoz, M. E., Arrieta, M. C., Ramer-Tait, A. E., & Walter, J. (2017). A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome, 5*(1), 48. doi:10.1186/s40168-017-0268-4

Plaza-Diaz, J., Fontana, L., & Gil, A. (2018). Human Milk Oligosaccharides and Immune System Development. *Nutrients, 10*(8). doi:10.3390/nu10081038

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56*(3), 345-365.

Ponsonby, A. L., Dwyer, T., & Couper, D. (1997). Factors related to infant apnoea and cyanosis: A population-based study. *Journal of Paediatrics and Child Health, 33*(4), 317-323.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant behavior and development, 29*(3), 386-401.

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and Assessment of Short and Very Short Forms of the Infant Behavior Questionnaire-Revised. *Journal of Personality Assessment, 96*(4), 445-458. doi:10.1080/00223891.2013.841171

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167-190.

Ramanathan, A., Gordon, S. M., & Shrestha, N. K. (2020). A case series of patients with Gemella endocarditis. *Diagnostic Microbiology and Infectious Disease, 97*(1), 115009.

Rampelli, S., Schnorr, S. L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., . . . Candela, M. (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology, 25*(13), 1682-1693. doi:10.1016/j.cub.2015.04.055

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F., & van der Ende, J. (2010). Predicting adult emotional and behavioral problems from externalizing problem trajectories in a 24-year longitudinal study. *European Child & Adolescent Psychiatry, 19*(7), 577-585.

Rigato, S., Charalambous, S., Stets, M., & Holmboe, K. (2020). Maternal depressive symptoms and infant temperament in the first year of life predict child behaviour at 3 years of age.

Rinne, M., Kalliomaki, M., Arvilommi, H., Salminen, S., & Isolauri, E. (2005). Effect of probiotics and breastfeeding on the bifidobacterium and lactobacillus/enterococcus microbiota and humoral immune responses. *J Pediatr, 147*(2), 186-191. doi:10.1016/j.jpeds.2005.03.053

Rivera-Pinto, J., Egozcue, J. J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., & Calle, M. L. (2018). Balances: a new perspective for microbiome analysis. *Msystems, 3*(4), e00053-00018.

Riviere, A., Selak, M., Lantin, D., Leroy, F., & De Vuyst, L. (2016). Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Frontiers in Microbiology, 7*. doi:ARTN 979

Rodriguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., . . . Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health and Disease, 26*. doi:ARTN 26050

Rogers, S. L., & Blissett, J. (2019). Infant temperament, maternal feeding behaviours and the timing of solid food introduction. *Maternal & child nutrition, 15*(3), e12771.

Rook, G. A., Raison, C. L., & Lowry, C. A. (2018). Childhood Microbial Experience, Immunoregulation, Inflammation, and Adult Susceptibility to Psychosocial Stressors and Depression. In *Inflammation and Immunity in Depression* (pp. 17-44): Elsevier.

Rosin, S., Xia, K., Azcarate-Peril, M. A., Carlson, A. L., Propper, C. B., Thompson, A. L., . . . Knickmeyer, R. C. (2021). A preliminary study of gut microbiome variation and HPA axis reactivity in healthy infants. *Psychoneuroendocrinology, 124*. doi:10.1016/j.psyneuen.2020.105046

Roswall, J., Olsson, L. M., Kovatcheva-Datchary, P., Nilsson, S., Tremaroli, V., Simon, M.-C., . . . Uhlén, M. (2021). Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. *Cell Host & Microbe, 29*(5), 765-776. e763.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current directions in psychological science, 16*(4), 207-212.

Rothbart, M. K. (2012). Advances in temperament. *Handbook of temperament*, 3-20.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*(5), 1394-1408.

Rothbart, M. K., & Bates, J. E. (2006). Temperament.

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion review, 3*(2), 207-213.

Russo, S., Kema, I. P., Fokkema, R. M., Boon, J. C., Willemse, P. H., de Vries, E. G., . . . Korf, J. (2003). Tryptophan as a link between psychopathology and somatic states. *Psychosomatic Medicine, 65*(4), 665-671.

Salim, S. Y., Kaplan, G. G., & Madsen, K. L. (2014). Air pollution effects on the gut microbiota: a link between exposure and inflammatory disease. *Gut Microbes, 5*(2), 215-219.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*(1), 142-170.

Saudino, K. J. (2005). Behavioral genetics and child temperament. *Journal of developmental and behavioral pediatrics: JDBP, 26*(3), 214.

Savage, J. H., Lee-Sarwar, K. A., Sordillo, J. E., Lange, N. E., Zhou, Y., O'Connor, G. T., . . . Sodergren, E. (2018). Diet during pregnancy and infancy and the infant intestinal microbiome. *The Journal of pediatrics, 203*, 47-54. e44.

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics, 27*(4), 592-593.

Schmitz, S., Saudino, K. J., Plomin, R., Fulker, D. W., & DeFries, J. (1996). Genetic and environmental influences on temperament in middle childhood: Analyses of teacher and tester ratings. *Child Development, 67*(2), 409-422.

Shen, Q., Chen, Y. A., & Tuohy, K. M. (2010). A comparative in vitro investigation into the effects of cooked meats on the human faecal microbiota. *Anaerobe, 16*(6), 572-577. doi:10.1016/j.anaerobe.2010.09.007

Sheng, J. A., Bales, N. J., Myers, S. A., Bautista, A. I., Roueinfar, M., Hale, T. M., & Handa, R. J. (2021). The hypothalamic-pituitary-adrenal axis: development, programming actions of hormones, and maternal-fetal interactions. *Frontiers in Behavioral Neuroscience, 14*, 601939.

Sigal, J. J., Perry, J. C., Robbins, J. M., Gagné, M.-A., & Nassif, E. (2003). Maternal preoccupation and parenting as predictors of emotional and behavioral problems in children of women with breast cancer. *Journal of Clinical Oncology, 21*(6), 1155-1160.

Silva, Y. P., Bernardi, A., & Frozza, R. L. (2020). The role of short-chain fatty acids from gut microbiota in gut-brain communication. *Frontiers in Endocrinology, 11*, 25.

Simons, S. S., Beijers, R., Cillessen, A. H., & de Weerth, C. (2015). Development of the cortisol circadian rhythm in the light of stress early in life. *Psychoneuroendocrinology, 62*, 292-300.

Skonieczna-Żydecka, K., Marlicz, W., Misera, A., Koulaouzidis, A., & Łoniewski, I. (2018). Microbiome—the missing link in the gut-brain axis: focus on its role in gastrointestinal and mental health. *Journal of clinical medicine, 7*(12), 521.

Sominsky, L., & Spencer, S. J. (2014). Eating behavior and stress: a pathway to obesity. *Frontiers in Psychology, 5*. doi:ARTN 434

Srinivasan, V., Smits, M., Spence, W., Lowe, A. D., Kayumov, L., Pandi-Perumal, S. R., . . . Cardinali, D. P. (2006). Melatonin in mood disorders. *The World Journal of Biological Psychiatry, 7*(3), 138-151.

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., . . . Petrosino, J. F. (2018a). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature, 562*(7728), 583-588. doi:10.1038/s41586-018-0617-x

Stilling, R. M., Dinan, T. G., & Cryan, J. F. (2014). Microbial genes, brain & behaviour–epigenetic regulation of the gut–brain axis. *Genes, Brain and Behavior, 13*(1), 69-86.

Sundekilde, U. K., Downey, E., O'Mahony, J. A., O'Shea, C.-A., Ryan, C. A., Kelly, A. L., & Bertram, H. C. (2016). The effect of gestational and lactational age on the human milk metabolome. *Nutrients, 8*(5), 304.

Team, R. C. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. In.

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology, 45*(6), 1887-1894.

Thomas, B., & Bishop, J. (2013). *Manual of dietetic practice*: John Wiley & Sons.

Thompson, A. L., Monteagudo-Mera, A., Cadenas, M. B., Lampl, M. L., & Azcarate-Peril, M. A. (2015). Milk- and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and metabolic and immune function of the infant gut microbiome. *Front Cell Infect Microbiol, 5*, 3. doi:10.3389/fcimb.2015.00003

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: the 'original sin'hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry, 51*(4), 341-367.

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., . . . Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods, 12*(10), 902-903.

Turnbaugh, P. J., & Gordon, J. I. (2009). The core gut microbiome, energy balance and obesity. *J Physiol, 587*(Pt 17), 4153-4158. doi:10.1113/jphysiol.2009.174136

Turroni, F., Foroni, E., Serafini, F., Viappiani, A., Montanini, B., Bottacini, F., . . . Ventura, M. (2011). Ability of Bifidobacterium breve to grow on different types of milk: exploring the metabolism of milk through genome analysis. *Appl Environ Microbiol, 77*(20), 7408-7417. doi:10.1128/AEM.05336-11

Uloko, A. E., Musa, B. M., Ramalan, M. A., Gezawa, I. D., Puepet, F. H., Uloko, A. T., . . . Sada, K. B. (2018). Prevalence and risk factors for diabetes mellitus in Nigeria: a systematic review and meta-analysis. *Diabetes Therapy, 9*(3), 1307-1316.

Ulrich-Lai, Y. M., Fulton, S., Wilson, M., Petrovich, G., & Rinaman, L. (2015). Stress exposure, food intake and emotional state. *Stress, 18*(4), 381-399.

Upadhyaya, S., & Banerjee, G. (2015). Type 2 diabetes and gut microbiome: at the intersection of known and unknown. *Gut Microbes, 6*(2), 85-92. doi:10.1080/19490976.2015.1024918

Vaher, K., Bogaert, D., Richardson, H., & Boardman, J. P. (2022). Microbiome-gut-brain axis in brain development, cognition and behavior during infancy and early childhood. *Developmental Review, 66*, 101038.

Valles, Y., Artacho, A., Pascual-Garcia, A., Ferrus, M. L., Gosalbes, M. J., Abellan, J. J., & Francino, M. P. (2014). Microbial Succession in the Gut: Directional Trends of Taxonomic and Functional Change in a Birth Cohort of Spanish Infants. *Plos Genetics, 10*(6). doi:ARTN e1004406

van de Wouw, M., Wang, Y., Workentine, M. L., Vaghef-Mehrabani, E., Dewey, D., Reimer, R. A., . . . Giesbrecht, G. F. (2022). Associations between the gut microbiota and internalizing behaviors in preschool children. *Psychosomatic Medicine, 84*(2), 159-169.

Verhoog, S., Taneri, P. E., Roa Díaz, Z. M., Marques-Vidal, P., Troup, J. P., Bally, L., . . . Muka, T. (2019). Dietary factors and modulation of bacteria strains of Akkermansia muciniphila and Faecalibacterium prausnitzii: a systematic review. *Nutrients, 11*(7), 1565.

Voreades, N., Kozil, A., & Weir, T. L. (2014). Diet and the development of the human intestinal microbiome. *Frontiers in Microbiology, 5*, 494. doi:10.3389/fmicb.2014.00494

Vuillermin, P., Saffery, R., Allen, K. J., Carlin, J. B., Tang, M. L., Ranganathan, S., . . . Jachno, K. (2015). Cohort profile: the Barwon infant study. *International Journal of Epidemiology, 44*(4), 1148-1160.

Wang, J., Wu, Y. J., Xiong, G. P., Chao, T. T., Jin, Q., Liu, R., . . . Yang, X. F. (2016). Introduction of complementary feeding before 4 months of age increases the risk of childhood overweight or obesity: a meta-analysis of prospective cohort studies. *Nutrition Research, 36*(8), 759-770. doi:10.1016/j.nutres.2016.03.003

Wang, M., Li, M., Wu, S., Lebrilla, C. B., Chapkin, R. S., Ivanov, I., & Donovan, S. M. (2015). Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed. *J Pediatr Gastroenterol Nutr, 60*(6), 825-833. doi:10.1097/MPG.0000000000000752

Wang, S. G., Harvey, L., Martin, R., van der Beek, E. M., Knol, J., Cryan, J. F., & Renes, I. B. (2018). Targeting the gut microbiota to influence brain development and function in early life. *Neuroscience and Biobehavioral Reviews, 95*, 191-201. doi:10.1016/j.neubiorev.2018.09.002

Wang, Y., Chen, X., Yu, Y., Liu, Y., Zhang, Q., & Bai, J. (2020). Association between Gut Microbiota and Infant's Temperament in the First Year of Life in a Chinese Birth Cohort. *Microorganisms, 8*(5). doi:10.3390/microorganisms8050753

Weaver, L. T. (2006). Rapid growth in infancy: balancing the interests of the child. *J Pediatr Gastroenterol Nutr, 43*(4), 428-432. doi:10.1097/01.mpg.0000235977.59873.e0

Wiles, N. J., Northstone, K., Emmett, P., & Lewis, G. (2009). 'Junk food'diet and childhood behavioural problems: results from the ALSPAC cohort. *European journal of clinical nutrition, 63*(4), 491-498.

Wong, C. B., Iwabuchi, N., & Xiao, J. Z. (2019). Exploring the Science behind Bifidobacterium breve M-16V in Infant Health. *Nutrients, 11*(8). doi:10.3390/nu11081724

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., . . . Lewis, J. D. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science, 334*(6052), 105-108. doi:10.1126/science.1208344

Xia, Y., Sun, J., & Chen, D.-G. (2018). *Statistical analysis of microbiome data with R*: Springer.

Yadav, M., Kapoor, A., Verma, A., & Ambatipudi, K. (2022). Functional significance of different milk constituents in modulating the gut microbiome and infant health. *Journal of Agricultural and Food Chemistry, 70*(13), 3929-3947.

Yagmurlu, B., & Altan, O. (2010). Maternal socialization and child temperament as predictors of emotion regulation in Turkish preschoolers. *Infant and Child Development: An International Journal of Research and Practice, 19*(3), 275-296.

Yan, R., Andrew, L., Marlow, E., Kunaratnam, K., Devine, A., Dunican, I. C., & Christophersen, C. T. (2021). Dietary fibre intervention for gut microbiota, sleep, and mental health in adults with irritable bowel syndrome: a scoping review. *Nutrients, 13*(7), 2159.

Yan, S., Cao, H., Gu, C., Ni, L., Tao, H., Shao, T., . . . Tao, F. (2018). Dietary patterns are associated with attention-deficit/hyperactivity disorder (ADHD) symptoms among preschoolers in mainland China. *European journal of clinical nutrition, 72*(11), 1517-1523.

Yang, J., Wei, H., Zhou, Y., Szeto, C.-H., Li, C., Lin, Y., . . . Sung, J. J. (2022). High-fat diet promotes colorectal tumorigenesis through modulating gut microbiota and metabolites. *Gastroenterology, 162*(1), 135-149. e132.

Yoleri, S. (2014). The relationship between temperament, gender, and behavioural problems in preschool children. *South African Journal of Education, 34*(2).

Zellner, D. A., Loaiza, S., Gonzalez, Z., Pita, J., Morales, J., Pecora, D., & Wolf, A. (2006). Food selection changes under stress. *Physiology & behavior, 87*(4), 789-793.

Zhang, Y., Fan, Q., Hou, Y., Zhang, X., Yin, Z., Cai, X., . . . Wang, G. (2022). Bacteroides species differentially modulate depression-like behavior via gut-brain metabolic signaling. *Brain, behavior, and immunity, 102*, 11-22.

Zhao, J., Liu, P., Wu, Y., Guo, P., Liu, L., Ma, N., . . . Zhang, J. (2018). Dietary fiber increases butyrate-producing bacteria and improves the growth performance of weaned piglets. *Journal of Agricultural and Food Chemistry, 66*(30), 7995-8004.

Zheng, H., Liang, H., Wang, Y., Miao, M., Shi, T., Yang, F., . . . Li, D. K. (2016). Altered Gut Microbiota Composition Associated with Eczema in Infants. *Plos One, 11*(11), e0166026. doi:10.1371/journal.pone.0166026

Zheng, P., Zeng, B., Zhou, C., Liu, M., Fang, Z., Xu, X., . . . Du, X. (2016). Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Molecular Psychiatry, 21*(6), 786-796.

Zhou, Y., Chen, C., Yu, H., & Yang, Z. (2020). Fecal microbiota changes in patients with postpartum depressive disorder. *Frontiers in Cellular and Infection Microbiology, 10*, 567268.

Zhu, Y., Li, Y., Zhang, Q., Song, Y., Wang, L., & Zhu, Z. (2022). Interactions between intestinal microbiota and neural mitochondria: a new perspective on communicating pathway from gut to brain. *Frontiers in Microbiology, 13*, 798917.

Zijlmans, M. A., Korpela, K., Riksen-Walraven, J. M., de Vos, W. M., & de Weerth, C. (2015). Maternal prenatal stress is associated with the infant intestinal microbiota. *Psychoneuroendocrinology, 53*, 233-245. doi:10.1016/j.psyneuen.2015.01.006

Zimmer, J., Lange, B., Frick, J.-S., Sauer, H., Zimmermann, K., Schwiertz, A., . . . Enck, P. (2012). A vegan or vegetarian diet substantially alters the human colonic faecal microbiota. *European journal of clinical nutrition, 66*(1), 53-60.

Zuo, T., Kamm, M. A., Colombel, J.-F., & Ng, S. C. (2018). Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nature Reviews Gastroenterology & Hepatology, 15*(7), 440-452.

Appendix 1

Strengths and Difficulties Questionnaire (Goodman 1997).

English- Australia edition.

Strengths and Difficulties Questionnaire

P or T 4-10

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain. Please give your answers on the basis of the child's behaviour over the last six months or this school year.

Child's name ..... Male/Female

Date of birth.....

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children, for example toys, treats, pencils	☐	☐	☐
Often loses temper	☐	☐	☐
Rather solitary, prefers to play alone	☐	☐	☐
Generally well behaved, usually does what adults request	☐	☐	☐
Many worries or often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, depressed or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often volunteers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets along better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Good attention span, sees work through to the end	☐	☐	☐

Signature ..... Date .....

Parent / Teacher / Other (Please specify):

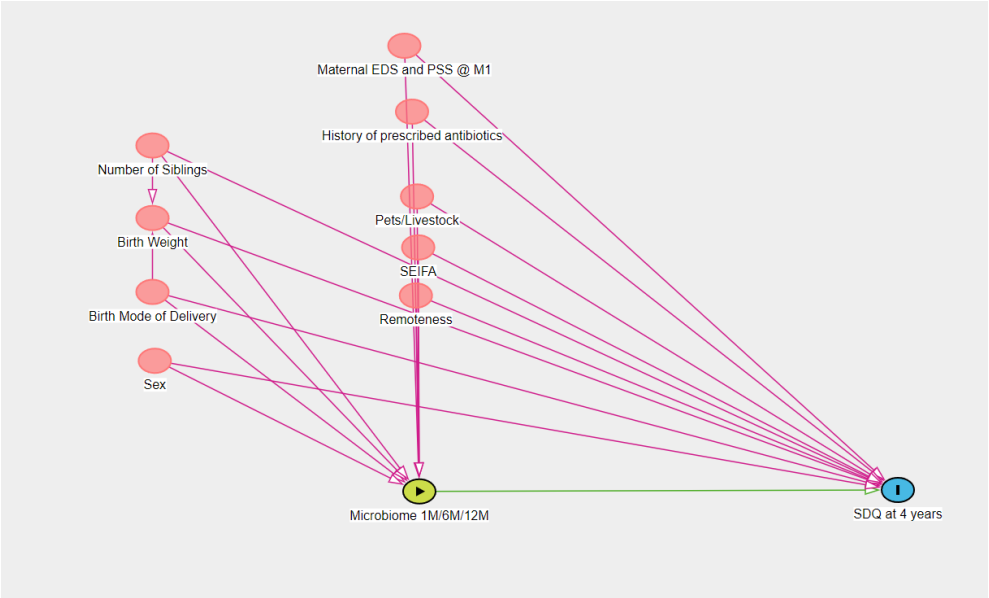
Thank you very much for your help

Appendix 2

DAGs for each hypothesis produced using DAGitty software.

1. Diversity and composition of gut microbiota will be associated with SDQ scores.

    a. Higher diversity at 4 weeks, 6-months, and 12-months of age, and lower relative abundance of genus *Bifidobacterium* and other butyrate producing bacteria, will be associated with a) higher total problem scores, b) higher subscale scores of hyperactivity, emotional symptoms, conduct problems and peer problems, and c) lower prosocial behaviour subscale scores.



dag {

bb="0,0,1,1"

"Birth Mode of Delivery" [pos="0.155,0.419"]

"Birth Weight" [pos="0.155,0.316"]

"History of prescribed antibiotics" [pos="0.381,0.168"]

"Maternal EDS and PSS @ M1" [pos="0.374,0.076"]

"Microbiome 1M/6M/12M " [exposure,pos="0.387,0.697"]

"Number of Siblings" [pos="0.155,0.215"]

"Pets/Livestock" [pos="0.385,0.286"]

"SDQ at 4-years" [outcome,pos="0.803,0.695"]

Remoteness [pos="0.384,0.424"]

SEIFA [pos="0.386,0.357"]

Sex [pos="0.157,0.515"]

"Birth Mode of Delivery" -> "Birth Weight"

"Birth Mode of Delivery" -> "Microbiome 1M/6M/12M "

"Birth Mode of Delivery" -> "SDQ at 4-years"

"Birth Weight" -> "Microbiome 1M/6M/12M "

"Birth Weight" -> "SDQ at 4-years"

"History of prescribed antibiotics" -> "Microbiome 1M/6M/12M "

"History of prescribed antibiotics" -> "SDQ at 4-years"

"Maternal EDS and PSS @ M1" -> "Microbiome 1M/6M/12M "

"Maternal EDS and PSS @ M1" -> "SDQ at 4-years"

"Microbiome 1M/6M/12M " -> "SDQ at 4-years"

"Number of Siblings" -> "Birth Weight"

"Number of Siblings" -> "Microbiome 1M/6M/12M "

"Number of Siblings" -> "SDQ at 4-years"

"Pets/Livestock" -> "Microbiome 1M/6M/12M "

"Pets/Livestock" -> "SDQ at 4-years"

Remoteness -> "Microbiome 1M/6M/12M "

Remoteness -> "SDQ at 4-years"

SEIFA -> "Microbiome 1M/6M/12M "

SEIFA -> "SDQ at 4-years"

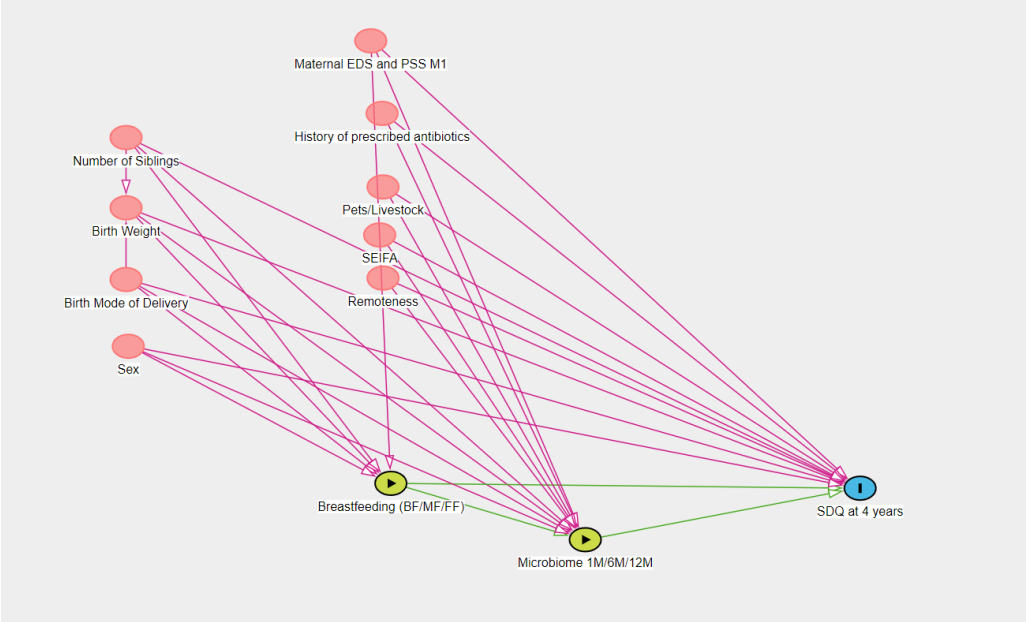
Sex -> "Microbiome 1M/6M/12M "

Sex -> "SDQ at 4-years"

}

2. Gut microbiota diversity and composition will mediate the relationships between early childhood diet and SDQ scores.

    a. It is predicted that those who are receiving breastmilk, either exclusively or via mixed feeding will have lower alpha diversity scores at 4 weeks, 6-months, and 12-months. It is further predicted that beta diversity will differ between those receiving breastmilk and those not receiving breastmilk. Furthermore, those receiving breastmilk will have higher relative abundance of genus *Bifidobacterium* (at the three timepoints), compared to those not receiving breastmilk.



dag {

bb="0,0,1,1"

"Birth Mode of Delivery" [pos="0.155,0.419"]

"Birth Weight" [pos="0.155,0.316"]

"Breastfeeding (BF/MF/FF)" [exposure,pos="0.392,0.712"]

"History of prescribed antibiotics" [pos="0.384,0.180"]

"Maternal EDS and PSS M1" [pos="0.374,0.076"]

"Microbiome 1M/6M/12M " [exposure,pos="0.566,0.793"]

"Number of Siblings" [pos="0.155,0.215"]

"Pets/Livestock" [pos="0.385,0.286"]

"SDQ at 4-years" [outcome,pos="0.812,0.719"]

Remoteness [pos="0.385,0.417"]

SEIFA [pos="0.382,0.355"]

Sex [pos="0.157,0.515"]

"Birth Mode of Delivery" -> "Birth Weight"

"Birth Mode of Delivery" -> "Breastfeeding (BF/MF/FF)"

"Birth Mode of Delivery" -> "Microbiome 1M/6M/12M "

"Birth Mode of Delivery" -> "SDQ at 4-years"

"Birth Weight" -> "Breastfeeding (BF/MF/FF)"

"Birth Weight" -> "Microbiome 1M/6M/12M "

"Birth Weight" -> "SDQ at 4-years"

"Breastfeeding (BF/MF/FF)" -> "Microbiome 1M/6M/12M "

"Breastfeeding (BF/MF/FF)" -> "SDQ at 4-years"

"History of prescribed antibiotics" -> "Microbiome 1M/6M/12M "

"History of prescribed antibiotics" -> "SDQ at 4-years"

"Maternal EDS and PSS M1" -> "Breastfeeding (BF/MF/FF)"

"Maternal EDS and PSS M1" -> "Microbiome 1M/6M/12M "

"Maternal EDS and PSS M1" -> "SDQ at 4-years"

"Microbiome 1M/6M/12M " -> "SDQ at 4-years"

"Number of Siblings" -> "Birth Weight"

"Number of Siblings" -> "Breastfeeding (BF/MF/FF)"

"Number of Siblings" -> "Microbiome 1M/6M/12M "

"Number of Siblings" -> "SDQ at 4-years"

"Pets/Livestock" -> "Microbiome 1M/6M/12M "

"Pets/Livestock" -> "SDQ at 4-years"

Remoteness -> "Microbiome 1M/6M/12M "

Remoteness -> "SDQ at 4-years"

SEIFA -> "Microbiome 1M/6M/12M "

SEIFA -> "SDQ at 4-years"

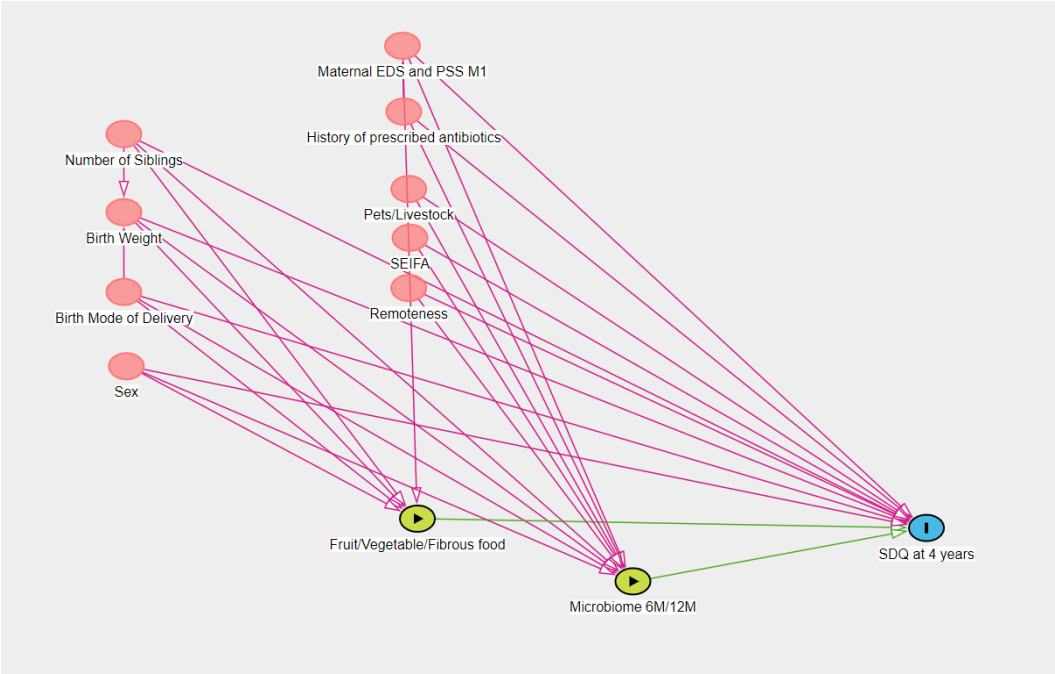
Sex -> "Breastfeeding (BF/MF/FF)"

Sex -> "Microbiome 1M/6M/12M "

Sex -> "SDQ at 4-years"

}

b. It is predicted that those children who consume fruits and vegetables, and other fibrous foods, in greater frequency, at 6-months and 12-months, will have higher relative abundance of butyrate producing bacteria, compared to children who consume those foods in lower frequency.



dag {

bb="0,0,1,1"

"Birth Mode of Delivery" [pos="0.155,0.419"]

"Birth Weight" [pos="0.155,0.316"]

"Fruit/Vegetable/Fibrous food" [exposure,pos="0.392,0.712"]

"History of prescribed antibiotics" [pos="0.381,0.186"]

"Maternal EDS and PSS M1" [pos="0.380,0.101"]

"Microbiome 6M/12M " [exposure,pos="0.566,0.793"]

"Number of Siblings" [pos="0.155,0.215"]

"Pets/Livestock" [pos="0.385,0.286"]

"SDQ at 4-years" [outcome,pos="0.803,0.724"]

Remoteness [pos="0.385,0.414"]

SEIFA [pos="0.386,0.349"]

Sex [pos="0.157,0.515"]

"Birth Mode of Delivery" -> "Birth Weight"

"Birth Mode of Delivery" -> "Fruit/Vegetable/Fibrous food"

"Birth Mode of Delivery" -> "Microbiome 6M/12M "

"Birth Mode of Delivery" -> "SDQ at 4-years"

"Birth Weight" -> "Fruit/Vegetable/Fibrous food"

"Birth Weight" -> "Microbiome 6M/12M "

"Birth Weight" -> "SDQ at 4-years"

"Fruit/Vegetable/Fibrous food" -> "SDQ at 4-years"

"History of prescribed antibiotics" -> "Maternal EDS and PSS M1"

"History of prescribed antibiotics" -> "Microbiome 6M/12M "

"History of prescribed antibiotics" -> "SDQ at 4-years"

"Maternal EDS and PSS M1" -> "Fruit/Vegetable/Fibrous food"

"Maternal EDS and PSS M1" -> "Microbiome 6M/12M "

"Maternal EDS and PSS M1" -> "SDQ at 4-years"

"Microbiome 6M/12M " -> "SDQ at 4-years"

"Number of Siblings" -> "Birth Weight"

"Number of Siblings" -> "Fruit/Vegetable/Fibrous food"

"Number of Siblings" -> "Microbiome 6M/12M "

"Number of Siblings" -> "SDQ at 4-years"

"Pets/Livestock" -> "Microbiome 6M/12M "

"Pets/Livestock" -> "SDQ at 4-years"

Remoteness -> "Microbiome 6M/12M "

Remoteness -> "SDQ at 4-years"

SEIFA -> "Microbiome 6M/12M "

SEIFA -> "SDQ at 4-years"

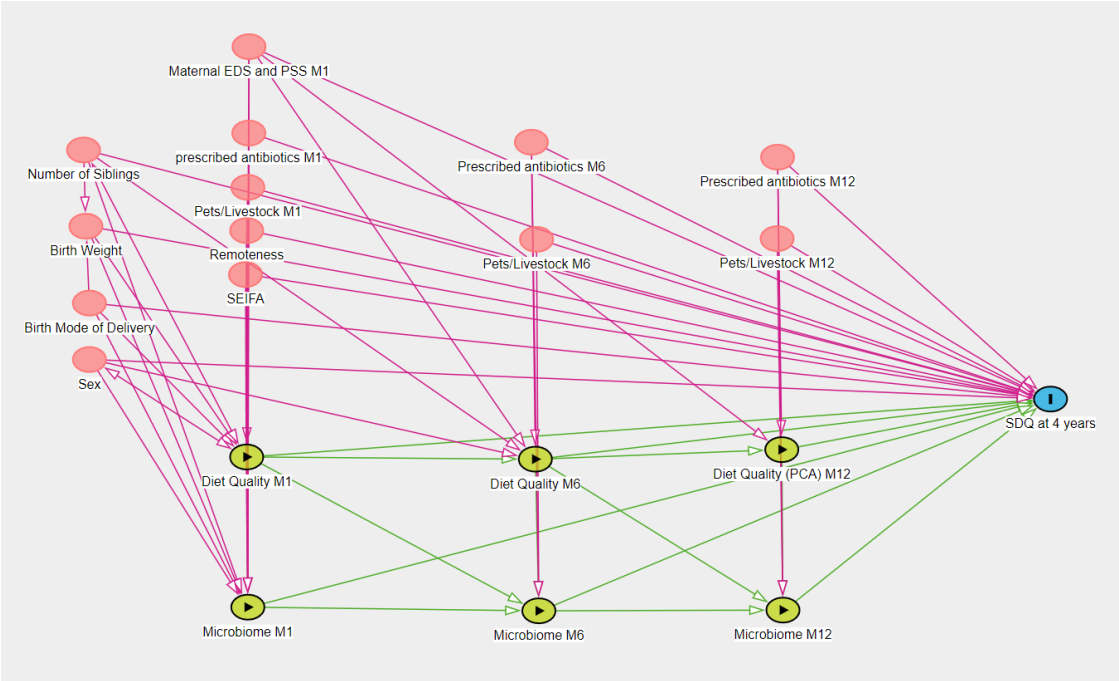
Sex -> "Fruit/Vegetable/Fibrous food"

Sex -> "Microbiome 6M/12M "

Sex -> "SDQ at 4-years"

}

c. It is predicted that early diet will be associated with SDQ scores and that the relationship between diet and SDQ scores will be mediated by microbiota composition. Specifically, it is hypothesised that dietary input, characterised by breastfeeding and higher frequency of fibrous food consumption, will predict greater abundance of taxa associated with positive developmental outcomes (as identified in previous literature, e.g. *Bifidobacteria*), which in turn will predict better SDQ scores at 4-years.



dag {

bb="0,0,1,1"

"Birth Mode of Delivery" [pos="0.093,0.435"]

"Birth Weight" [pos="0.090,0.330"]

"Diet Quality (PCA) M12" [exposure,pos="0.683,0.635"]

"Diet Quality M1" [exposure,pos="0.227,0.645"]

"Diet Quality M6" [exposure,pos="0.473,0.648"]

"Maternal EDS and PSS M1" [pos="0.229,0.085"]

"Microbiome M1" [exposure,pos="0.228,0.850"]

"Microbiome M12" [exposure,pos="0.684,0.854"]

"Microbiome M6" [exposure,pos="0.476,0.855"]

"Number of Siblings" [pos="0.088,0.226"]

"Pets/Livestock M1" [pos="0.228,0.277"]

"Pets/Livestock M12" [pos="0.679,0.347"]

"Pets/Livestock M6" [pos="0.474,0.348"]

"Prescribed antibiotics M12" [pos="0.679,0.236"]

"Prescribed antibiotics M6" [pos="0.470,0.215"]

"SDQ at 4-years" [outcome,pos="0.912,0.566"]

"prescribed antibiotics M1" [pos="0.229,0.203"]

Remoteness [pos="0.227,0.336"]

SEIFA [pos="0.226,0.396"]

Sex [pos="0.093,0.512"]

"Birth Mode of Delivery" -> "Birth Weight"

"Birth Mode of Delivery" -> "Diet Quality M1"

"Birth Mode of Delivery" -> "Microbiome M1"

"Birth Mode of Delivery" -> "SDQ at 4-years"

"Birth Weight" -> "Diet Quality M1"

"Birth Weight" -> "Microbiome M1"

"Birth Weight" -> "SDQ at 4-years"

"Diet Quality (PCA) M12" -> "Microbiome M12"

"Diet Quality (PCA) M12" -> "SDQ at 4-years"

"Diet Quality M1" -> "Diet Quality M6"

"Diet Quality M1" -> "Microbiome M6"

"Diet Quality M1" -> "SDQ at 4-years"

"Diet Quality M1" <-> "Microbiome M1"

"Diet Quality M1" <-> "Number of Siblings"

"Diet Quality M1" <-> Sex

"Diet Quality M6" -> "Diet Quality (PCA) M12"

"Diet Quality M6" -> "Microbiome M12"

"Diet Quality M6" -> "Microbiome M6"

"Diet Quality M6" -> "SDQ at 4-years"

"Maternal EDS and PSS M1" -> "Diet Quality (PCA) M12"

"Maternal EDS and PSS M1" -> "Diet Quality M1"

"Maternal EDS and PSS M1" -> "Diet Quality M6"

"Maternal EDS and PSS M1" -> "Microbiome M1"

"Maternal EDS and PSS M1" -> "SDQ at 4-years"

"Microbiome M1" -> "Microbiome M6"

"Microbiome M1" -> "SDQ at 4-years"

"Microbiome M12" -> "SDQ at 4-years"

"Microbiome M6" -> "Microbiome M12"

"Microbiome M6" -> "SDQ at 4-years"

"Number of Siblings" -> "Birth Weight"

"Number of Siblings" -> "Diet Quality M6"

"Number of Siblings" -> "Microbiome M1"

"Number of Siblings" -> "SDQ at 4-years"

"Pets/Livestock M1" -> "Microbiome M1" [pos="0.225,0.463"]

"Pets/Livestock M1" -> "SDQ at 4-years"

"Pets/Livestock M12" -> "Microbiome M12"

"Pets/Livestock M12" -> "SDQ at 4-years"

"Pets/Livestock M6" -> "Microbiome M6"

"Pets/Livestock M6" -> "SDQ at 4-years"

"Prescribed antibiotics M12" -> "Diet Quality (PCA) M12"

"Prescribed antibiotics M12" -> "Microbiome M12"

"Prescribed antibiotics M12" -> "SDQ at 4-years"

"Prescribed antibiotics M6" -> "Diet Quality M6"

"Prescribed antibiotics M6" -> "Microbiome M6"

"Prescribed antibiotics M6" -> "SDQ at 4-years"

"prescribed antibiotics M1" -> "Microbiome M1"

"prescribed antibiotics M1" -> "SDQ at 4-years"

Remoteness -> "Microbiome M1"

```
Remoteness -> "SDQ at 4-years"

SEIFA -> "Microbiome M1"

SEIFA -> "SDQ at 4-years"

Sex -> "Diet Quality M6"

Sex -> "Microbiome M1"

Sex -> "SDQ at 4-years"

}
```

Appendix 3

Longitudinal analysis of Microbiota, diet and SDQ using SEM.

full Correlation results

Table 3.4. Correlation table for all behavioural, microbiota and covariate measures to be included in SEM.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Sex of Child	-													
2. Mode of Birth	0.126	-												
3. Birthweight	0.077**	-0.166**	-											
4. Number of Siblings	0.042	-0.058	0.362**	-										
5. SEIFA	0.013	0.019	0.045	0.141*	-									
6. Remoteness	0.089	0.071	0.019	0.095	0.308**	-								
7. Number of Pets (M1)	-0.021	-0.016	-0.084	0.077	-0.063	0.156**	-							
8. Number of Pets (M12)	0.043	0.066	-0.053	0.081	0.055	0.116	0.646**	-						
9. Antibiotics duration (M1)	0.031	0.022	-0.042	-0.135	-0.050	-0.050	-0.047	0.016	-					
10. Antibiotics duration (M6)	0.068	-0.053	-0.014	0.025	0.033	-0.030	0.008	0.005	-0.005	-				
11. Antibiotics duration (M12)	0.059	-0.110*	0.026	-0.062	-0.029	-0.067	0.022	-0.074	-0.016	-0.035	-			
12. Maternal PSS score	-0.060	0.120	-0.088	0.018	-0.099	-0.045	0.040	0.115	0.051	-0.063	0.066	-		
13. Maternal EDS score	-0.062	0.104	-0.089	-0.040	-0.122	-0.079	0.068	0.050	0.013	-0.045	0.132	0.817**	-	
14. SDQ – Emotion	0.035	-0.058	0.099	-0.202**	0.037	-0.006	0.018	0.155	0.101	-0.084	0.075	0.140*	0.119	-
15. SDQ – Conduct	0.014	-0.115*	0.082	0.089	0.002	-0.030	-0.126	-0.064	0.037	-0.070	0.133	0.240**	0.201**	0.111*

Table 3.4 Continued.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
16. SDQ – Hyperactivity	0.058	-0.007	-0.037	0.069	-0.105*	0.085	0.071	0.023	0.059	-0.097	0.056	0.175*	0.178*	0.130
17. SDQ – Peer problems	0.028	0.047	-0.048	-0.040	-0.036	-0.008	0.032	0.038*	0.019	-0.074	0.023	0.065	0.017	0.259**
18. SDQ – Prosocial	-0.083	0.000	-0.111	-0.119	0.072	-0.112*	-0.012	0.137	-0.005	0.056	0.042	0.016	0.029	-0.027
19. SDQ – Total difficulties	0.057	-0.047	0.028	-0.022	-0.053	0.030	0.014	0.061*	0.087	-0.130	0.109	0.244**	0.210**	0.559**
20. SDI (M1)	-0.020	-0.052	-0.041	-0.037	0.045	-0.106	0.129	0.183*	0.100	0.136*	-0.003	-0.065	-0.080	0.094
21. SDI (M6)	0.055	-0.183	0.097	-0.053	0.022	0.021	0.074	0.069	-0.083	-0.100	0.126	0.061	0.107	0.135*
22. SDI (M12)	0.054	-0.079	0.141	0.126	0.140	0.038	0.053	0.065	0.103	-0.132	-0.121	-0.091	-0.128	0.005
23. Bifidobacteria (M1)	-0.115	-0.240**	0.144	0.043	0.070	0.011	0.007	-0.061	-0.033	-0.109	-0.020	-0.085	-0.117	-0.008
24. Bifidobacteria (M6)	0.097	-0.140**	-0.068	0.034	-0.112	-0.018	-0.055	0.025	0.075	-0.068	-0.045	0.058	0.036	-0.033
25. Bifidobacteria (M12)	0.025	-0.027	0.076	0.142	-0.130	0.061	0.079	0.033	-0.003	-0.075	-0.055	0.053	0.034	-0.028
26. Butyrate Producing (M1).	0.089	0.026	-0.047	-0.031	-0.041	0.038	-0.071	-0.080	0.153	0.011	0.018	0.100	0.058	0.134
27. Butyrate Producing (M6).	0.056	-0.089	0.046	-0.027	-0.035	-0.106	0.048	0.055	0.014	-0.064	0.178*	0.016	0.068	0.044
28. Butyrate Producing (M12).	-0.134	-0.099	0.104	-0.038	-0.084	-0.074	-0.006	0.043	0.061	-0.065	-0.054	-0.040	-0.135	0.020
29. Breastfeeding status (M1)	0.038	0.072*	-0.095	-0.065	-0.055	-0.104	0.077	0.070	0.152	-0.082	0.026	0.142	0.151*	0.025
30. Breastfeeding status (M6)	-0.039	-0.003	-0.013	-0.102*	0.028	-0.029	0.150	0.097	0.018	-0.016	0.193*	0.144*	0.216**	0.049
31. Breastfeeding status (M12)	-0.047	-0.083	0.011	-0.055	-0.083	-0.109	0.150	0.082	0.050	-0.085	0.138*	0.127	0.206**	0.055
32. Fruit, Veg, & Fibre (M6)	0.083*	-0.039	-0.049	-0.204**	-0.080	-0.091	-0.072	-0.073	0.038	-0.006	-0.065*	-0.072	-0.072	0.007
33. Fruit, Veg & Fibre (M12)	0.090	-0.045	-0.058	-0.214**	-0.085	-0.112	-0.048	-0.054	0.142	0.006	0.170**	-0.024	-0.018	0.030

Table 3.4 Continued

Variable	15	16	17	18	19	20	21	22	23	24	25	26	27	28
15. SDQ – Conduct	-													
16. SDQ – Hyperactivity	0.449**	-												
17. SDQ – Peer problems	0.141**	0.122*	-											
18. SDQ – Prosocial	-0.223**	-0.445**	-0.293**	-										
19. SDQ – Total difficulties	0.653**	0.767**	0.531**	-0.415**	-									
20. SDI (M1)	-0.021	0.008	0.035	0.006	0.044	-								
21. SDI (M6)	0.117	0.082	-0.041	0.048	0.118*	0.141	-							
22. SDI (M12)	-0.007	-0.076	-0.054	0.042	-0.058	-0.053	0.215*	-						
23. Bifidobacteria (M1)	0.066	-0.081	-0.085	0.009	-0.052	-0.278**	-0.034	0.111	-					
24. Bifidobacteria (M6)	0.075	-0.042	-0.108	0.116	-0.045	-0.121*	-0.037	0.142	0.274**	-				
25. Bifidobacteria (M12)	-0.024	-0.017	-0.037	0.026	-0.039	-0.053	-0.167*	0.016	0.111	0.229**	-			
26. Butyrate Producing (M1).	0.045	-0.048	-0.001	0.102	0.039	0.151*	0.106	0.017	-0.034	0.072	-0.027	-		
27. Butyrate Producing (M6).	0.205	0.162	0.080	-0.179	0.196	0.171	0.358	0.042	-0.099	-0.121	-0.000	-0.023**	-	
28. Butyrate Producing (M12).	0.108	0.069	-0.081	-0.052	0.053	0.080	0.138	0.181*	0.148*	0.202*	-0.077	0.020	0.070	-
29. Breastfeeding status (M1)	0.096	0.091	0.053	-0.010	0.106	0.168*	0.254**	0.038	-0.086	-0.007	-0.112	0.164**	0.355	0.128*
30. Breastfeeding status (M6)	0.092	0.198*	0.071	-0.044	0.177	0.103	0.292**	0.079	-0.077	-0.019	-0.058	0.006	0.394	0.224**
31. Breastfeeding status (M12)	0.060	0.117	0.101	0.023	0.135	0.057	0.243**	0.113	-0.043	0.055	-0.038	0.006	0.203	0.205**
32. Fruit, Veg, & Fibre (M6)	-0.009	0.066	-0.006	0.085	0.033	0.0003	0.186**	-0.038	0.052	-0.063	-0.269**	0.146*	-0.026	0.163
33. Fruit, Veg & Fibre (M12)	0.038	0.089	0.011	0.078	0.074	0.031	0.227**	-0.039	0.021	-0.058	-0.284**	0.203**	0.069	0.183*

Table 3.4 Continued

Variable	29	30	31	32	33
29. Breastfeeding status (M1)	-				
30. Breastfeeding status (M6)	0.635**	-			
31. Breastfeeding status (M12)	0.396**	0.645**	-		
32. Fruit, Veg, & Fibre (M6)	0.150	0.117	0.170	-	
33. Fruit, Veg & Fibre (M12)	0.344**	0.247*	0.257**	0.964**	-

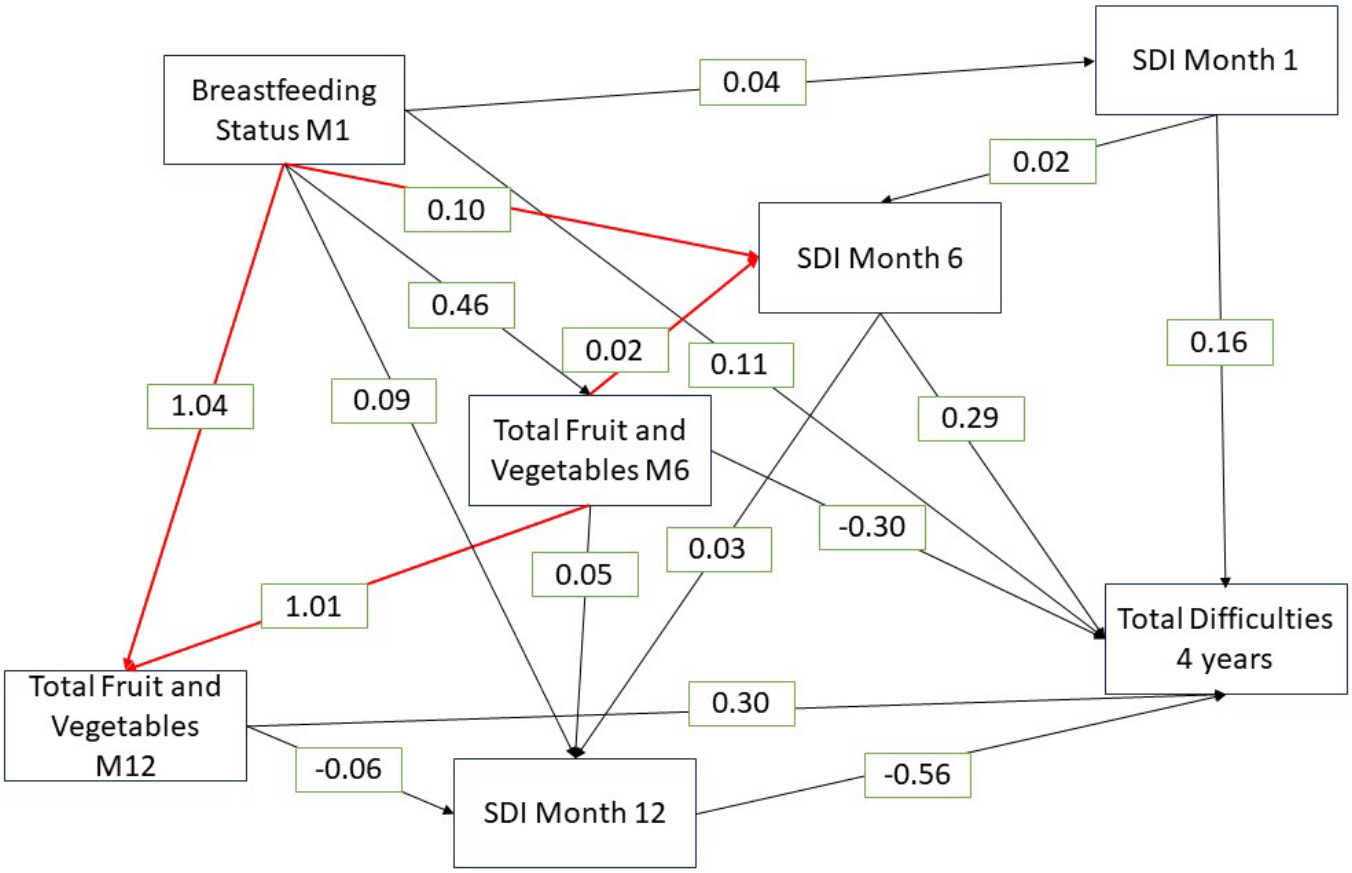
Note. * indicates significance $p < .05$, ** indicates significance $p < .01$

Appendix 4

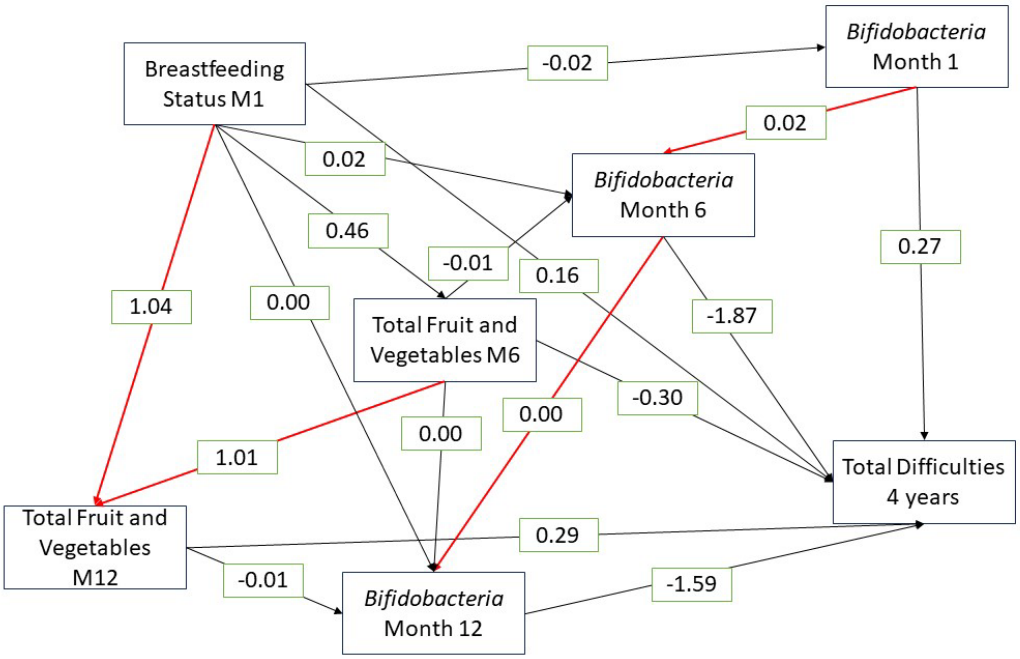
SEM diagrams for each model that reached goodness of fit parameters.

For each of the following SEM diagrams the numerical values presented next to each path represents the coefficients of the relationship. Only those models that reached the threshold to be considered a good fit have been presented in this Appendix. In each of the models the significant pathways have been highlighted in red.

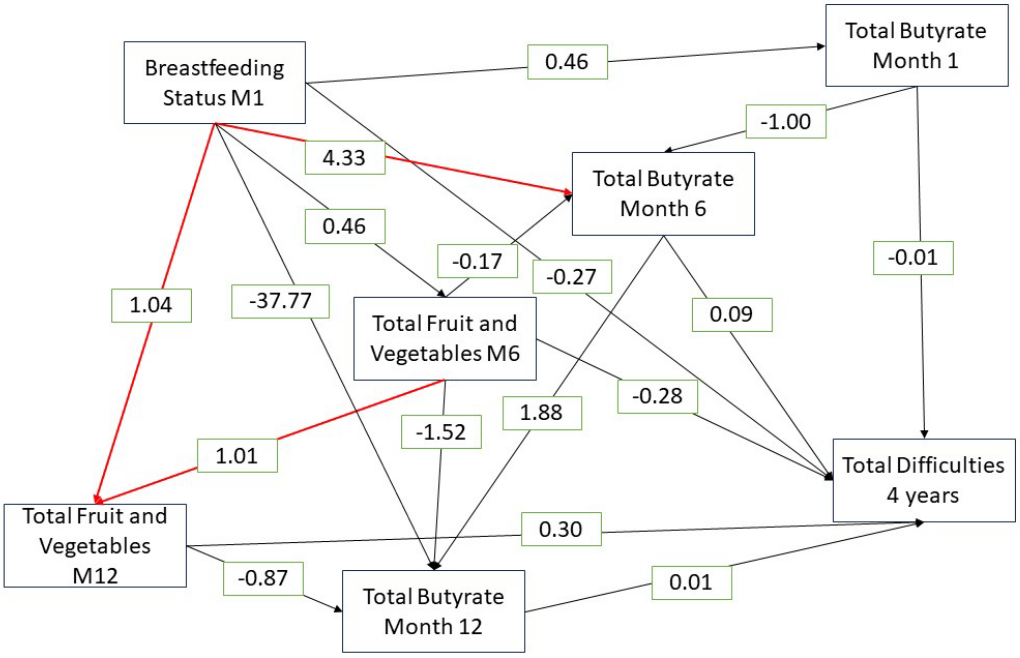
SEM 1. Total Difficulties, SDI, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p = 0.097, CFI = 0.971, and RMSEA = 0.843).



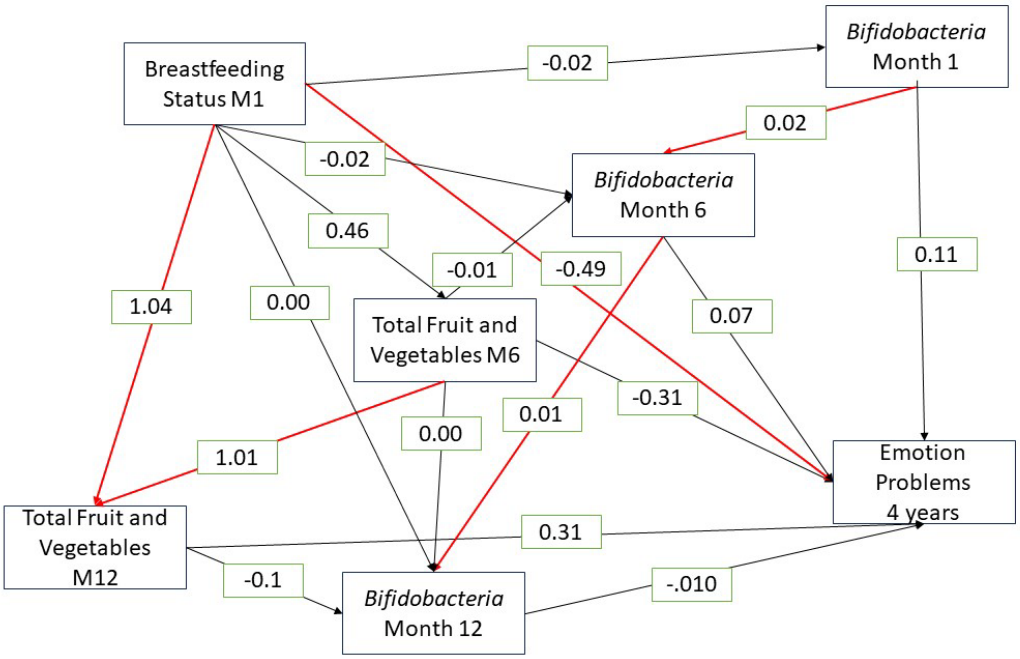
SEM 2. Total Difficulties, relative abundance of *Bifidobacteria* and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p = .370, CFI = 0.994, and RMSEA = 0.958).



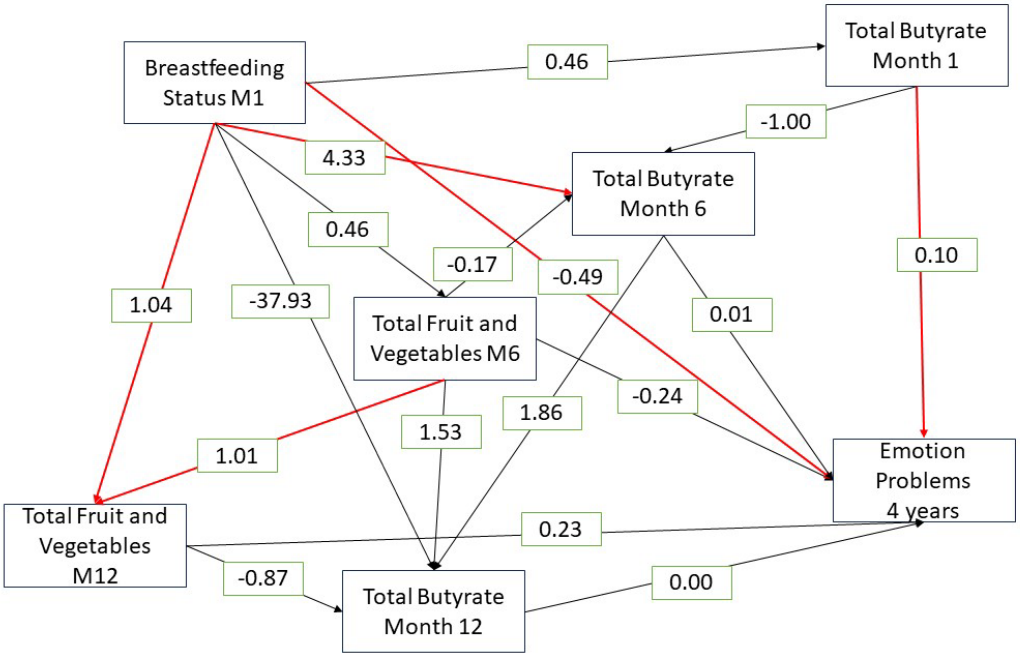
SEM 3. Total Difficulties, total butyrate producing bacteria and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p = 0.054, CFI = 0.968, and RMSEA = 0.710).



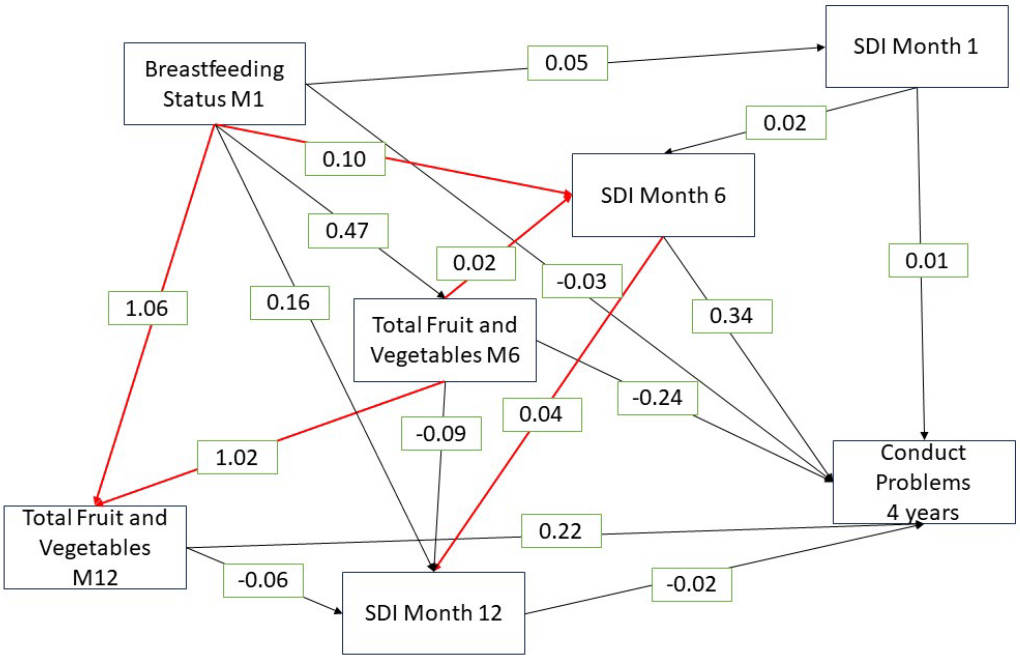
SEM 4. Emotional problems, relative abundance of *Bifidobacteria,* and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* = .337, CFI = 0.993, and RMSEA = 0.935).



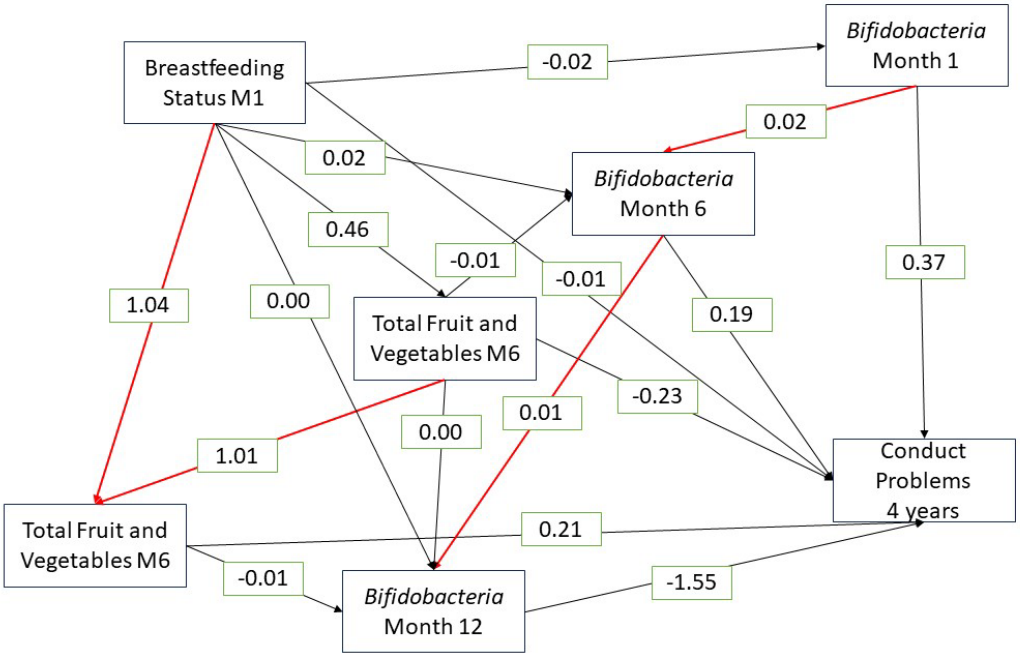
SEM 5. Emotional problems, total butyrate producing bacteria and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* = 0.062, CFI = 0.972, and RMSEA = 0.682).



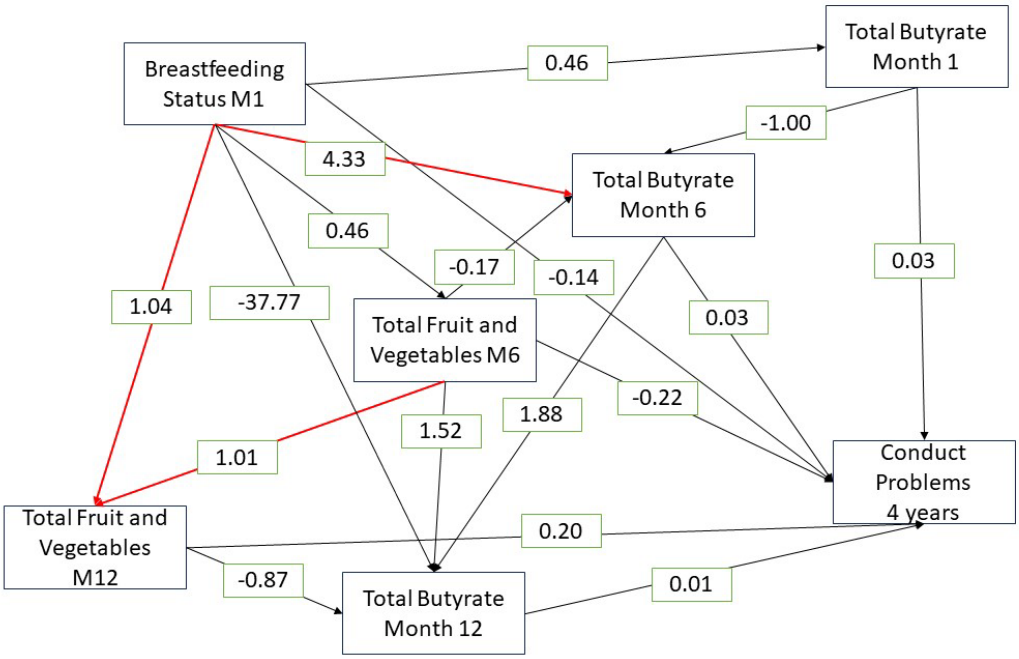
SEM 6. Conduct problems, SDI, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.048, CFI = 0.967, and RMSEA = 0.726).



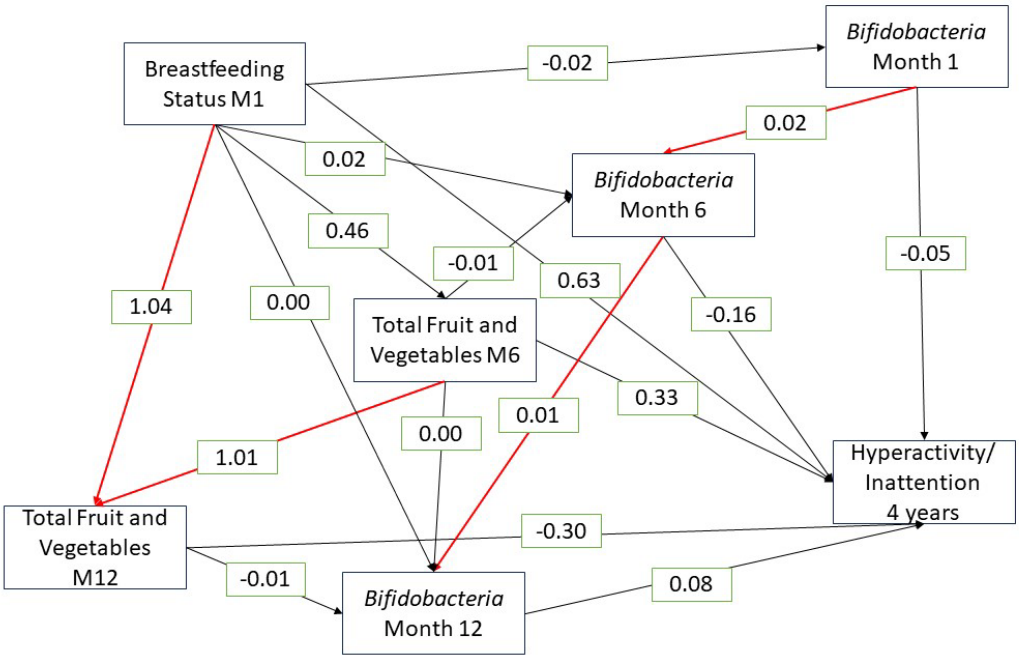
SEM 7. Conduct problems, relative abundance of *Bifidobacteria* and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.367, CFI = 0.995, and RMSEA = 0.942).



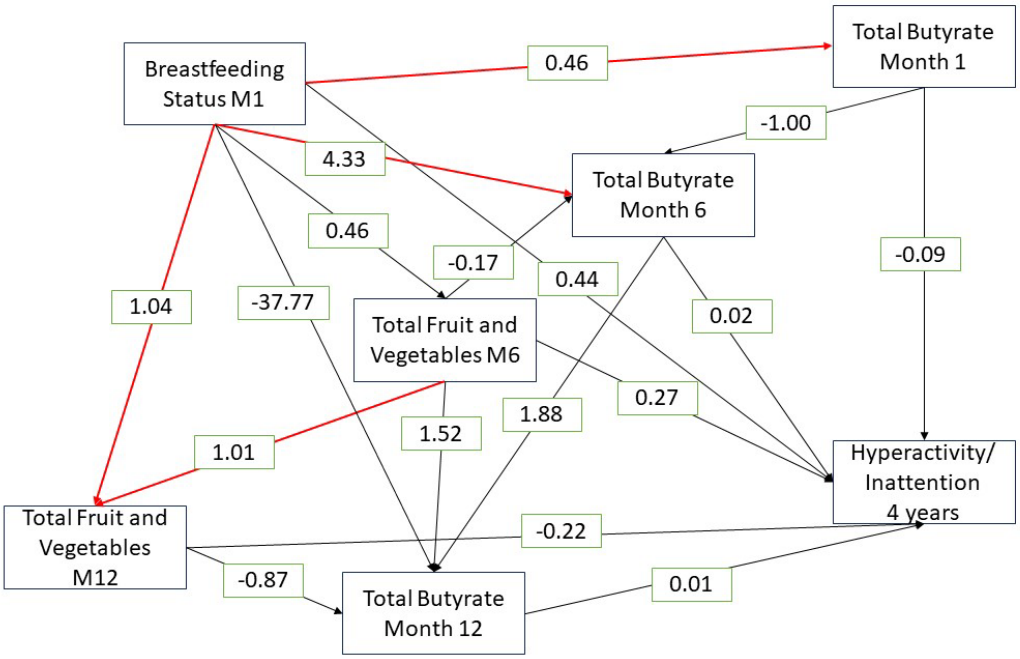
SEM 8. Conduct Problems, total butyrate producing bacteria, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p < .0068, CFI = 0.973, and RMSEA = 0.694).



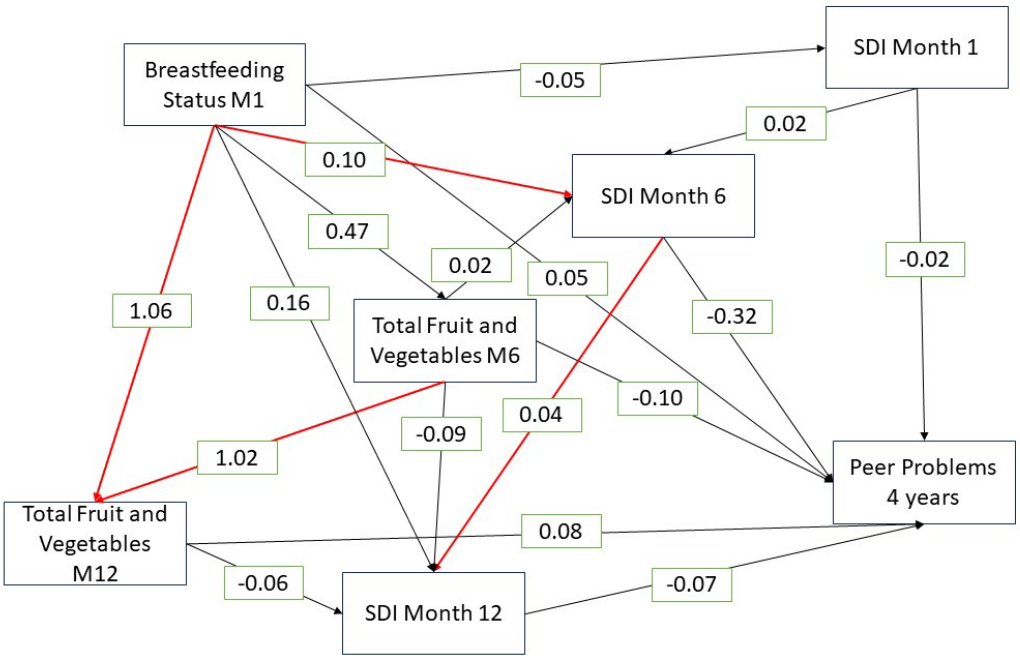
SEM 9. Hyperactivity/inattention, relative abundance of *Bifidobacteria,* and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p < .216, CFI = 0.985, and RMSEA = 0.903).



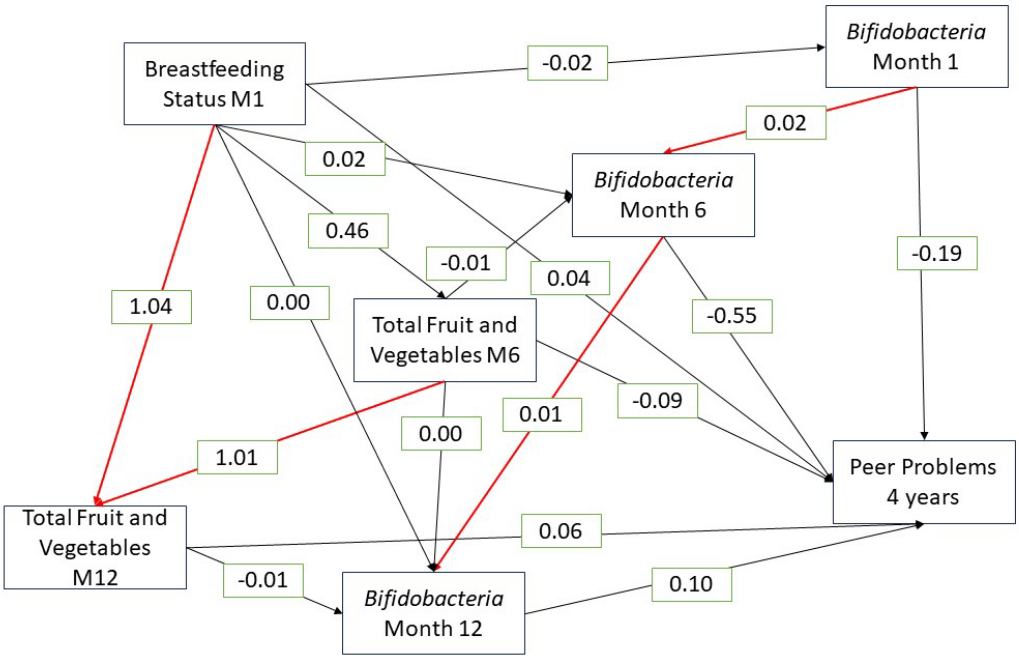
SEM 10. Hyperactivity/inattention, total butyrate producing bacteria, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p < .0.064, CFI = 0.970, and RMSEA = 0.732).



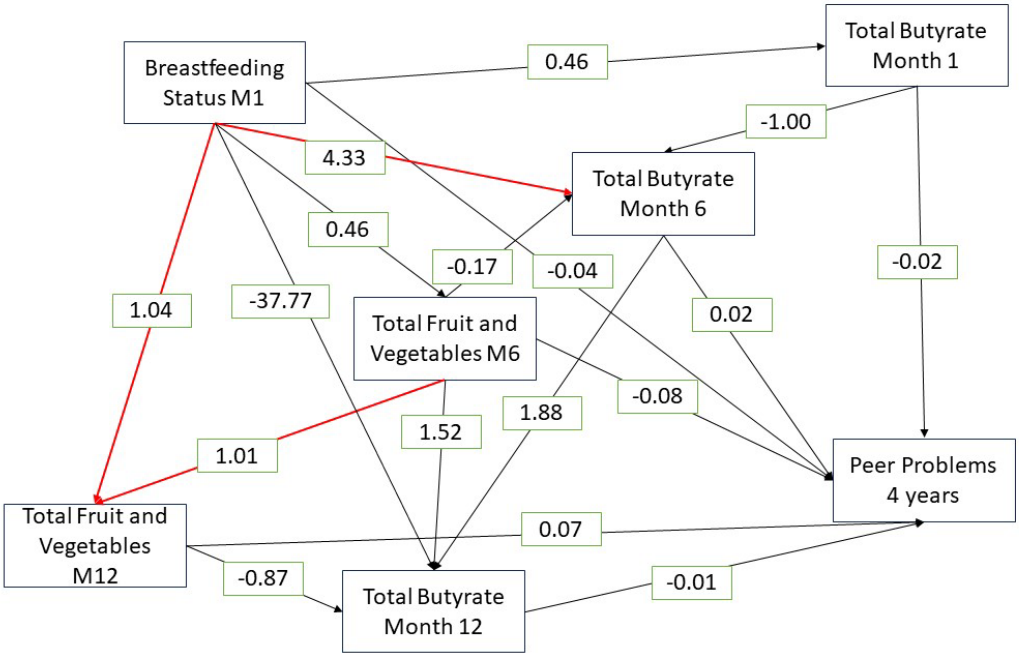
SEM 11. Peer problems, SDI, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 p =0.061, CFI = 0.966, and RMSEA = 0.810).



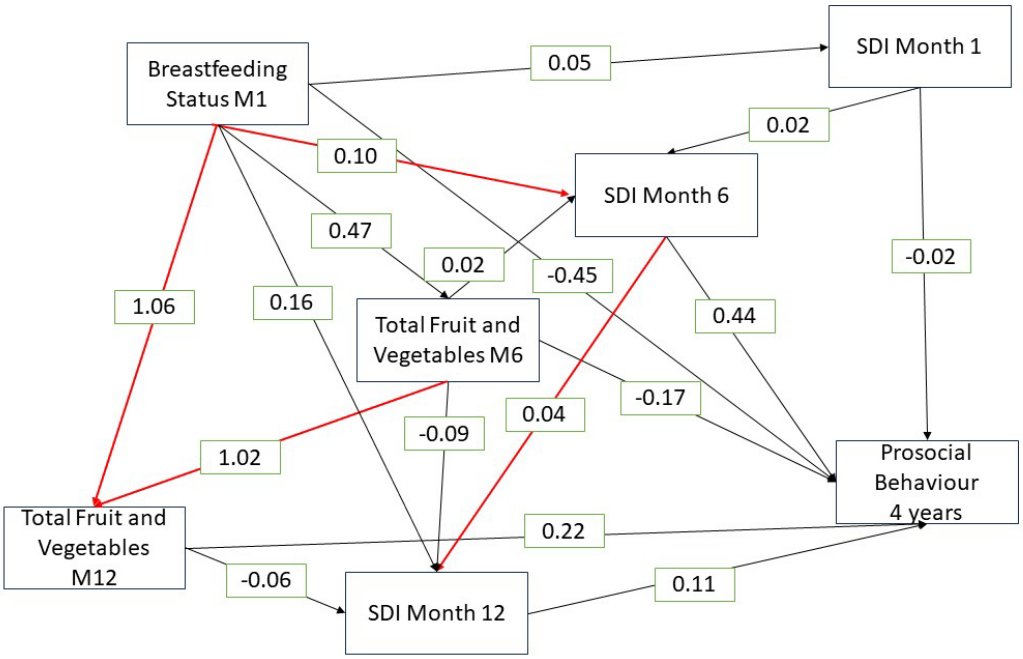
SEM 12. Peer problems, relative abundance of *Bifidobacteria*, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.061, CFI = 0.966, and RMSEA = 0.810).



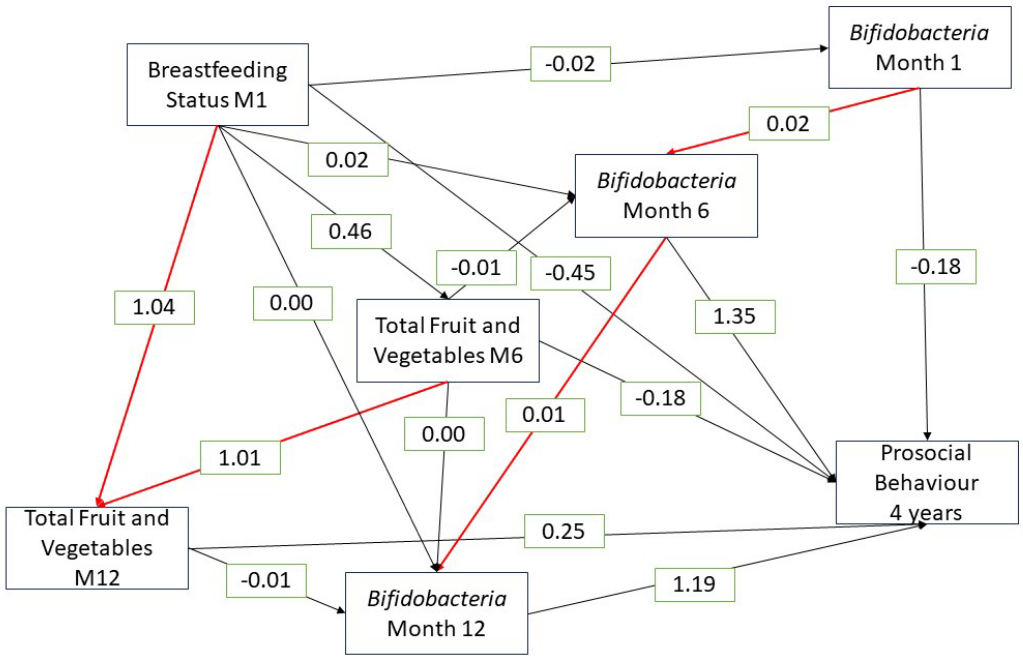
SEM 13. Peer problems, total butyrate producing bacteria, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.119, CFI = 0.977, and RMSEA = 0.833).



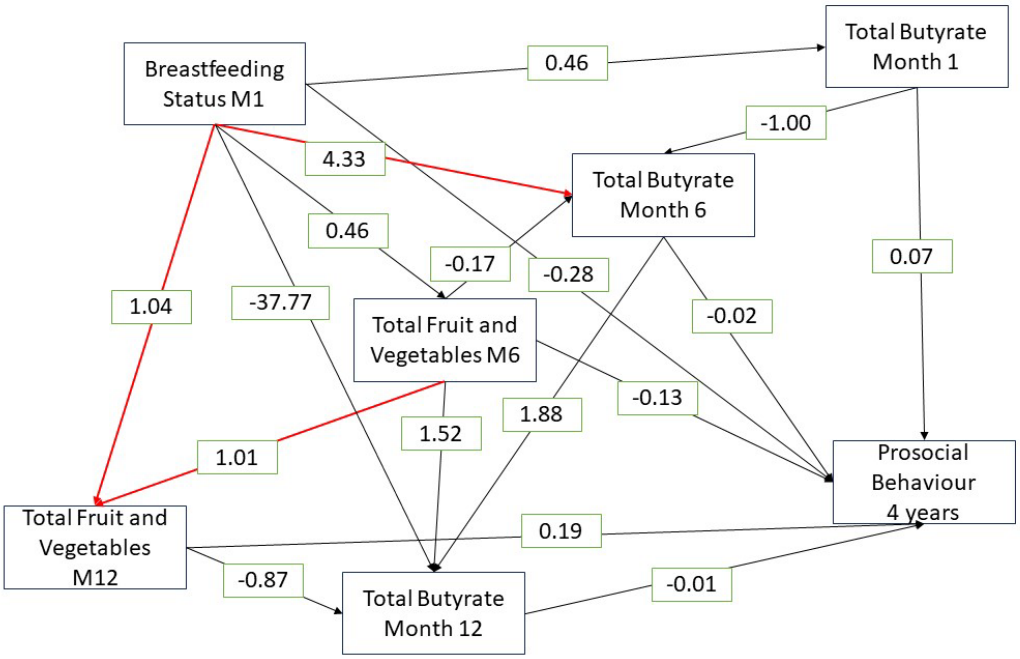
SEM 14. Prosocial behaviour, SDI, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.095, CFI = 0.974, and RMSEA = 0.824).



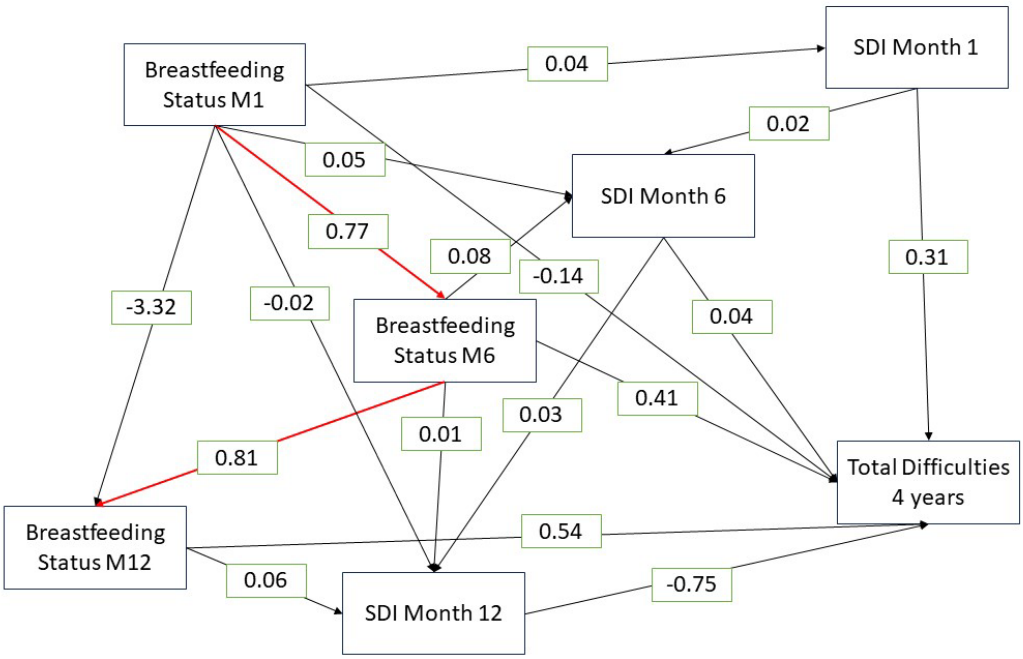
SEM 15. Prosocial behaviour, relative abundance of *Bifidobacteria*, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.656, CFI = 1.000, and RMSEA = 0.993).



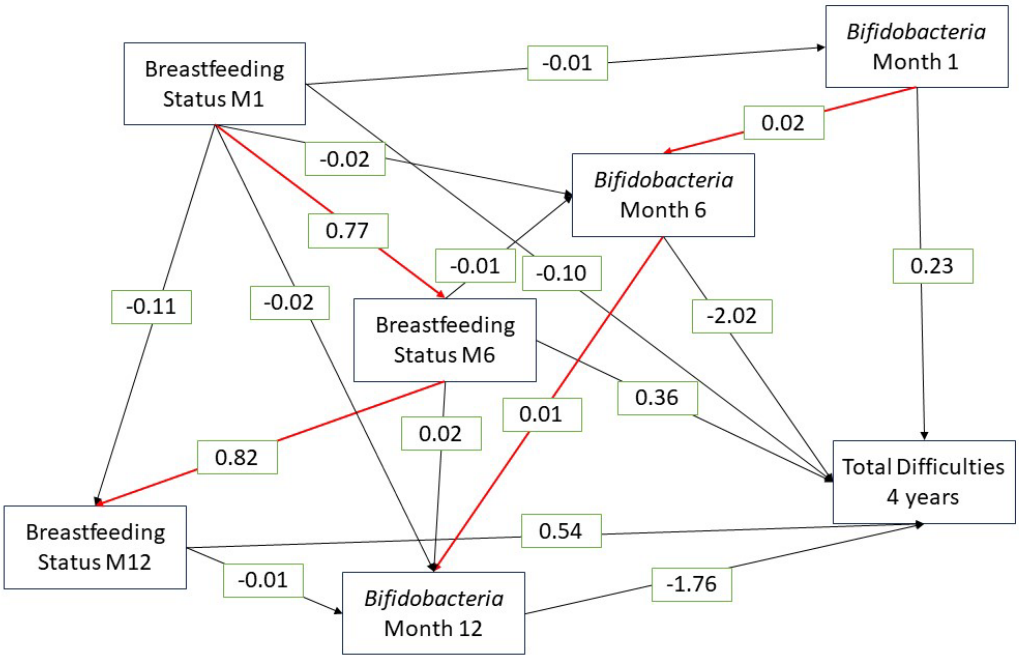
SEM 16. Prosocial behaviour, total butyrate producing bacteria, and Diet status, breastfeeding and Fruit, vegetable, and fibre intake (X^2 *p* =0.199, CFI = 0.984, and RMSEA = 0.899).



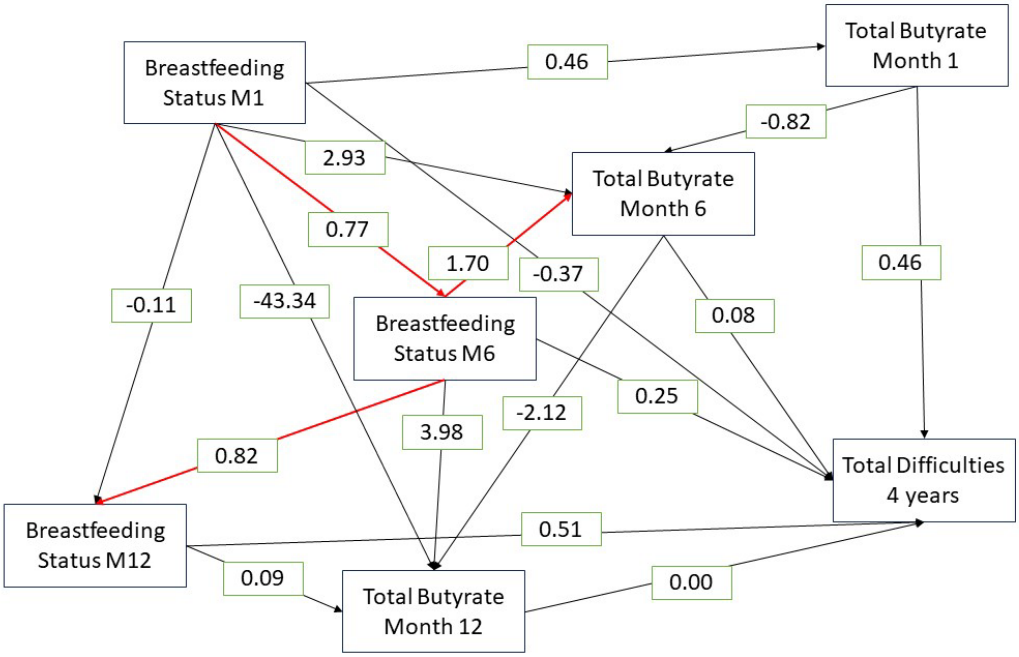
SEM 17. Total Difficulties, SDI, and breastfeeding status (X^2 *p* = 0.08, CFI = 0.940, and RMSEA = 0.773).



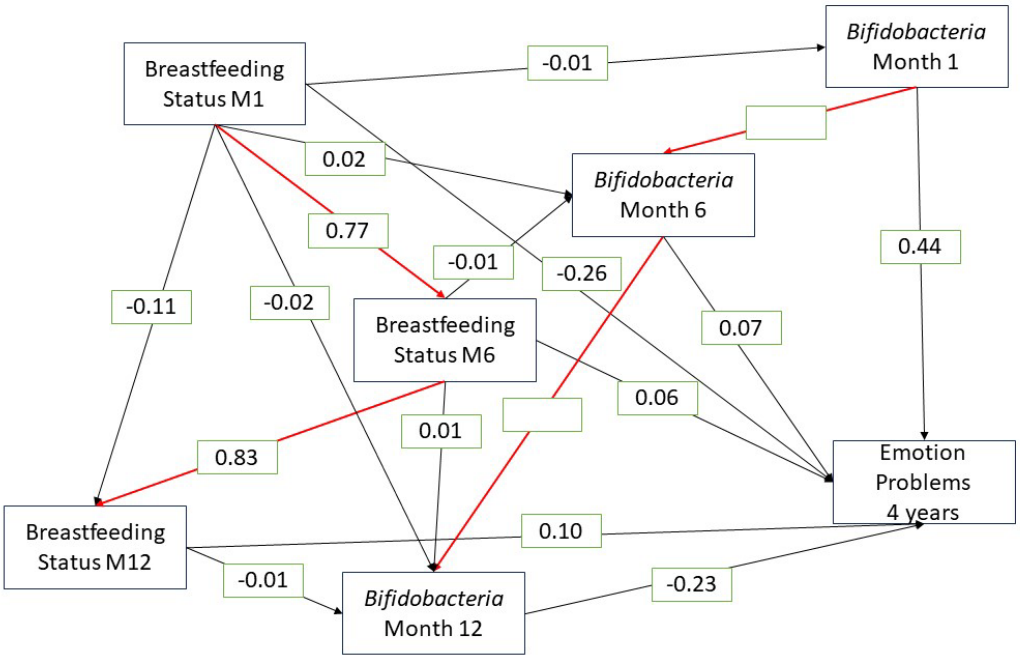
SEM 18. Total Difficulties, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 p = 0.402, CFI = 0.992, and RMSEA = 0.950).



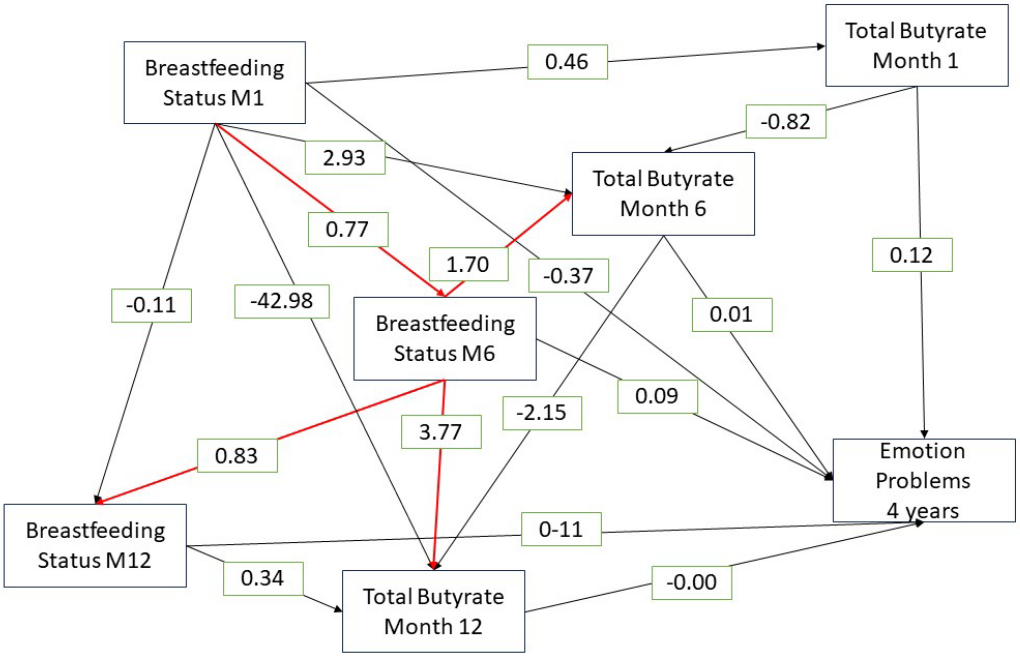
SEM 19. Total Difficulties, total butyrate producing bacteria, and breastfeeding status (X^2 p < .05, CFI = 0.908, and RMSEA = 0.617).



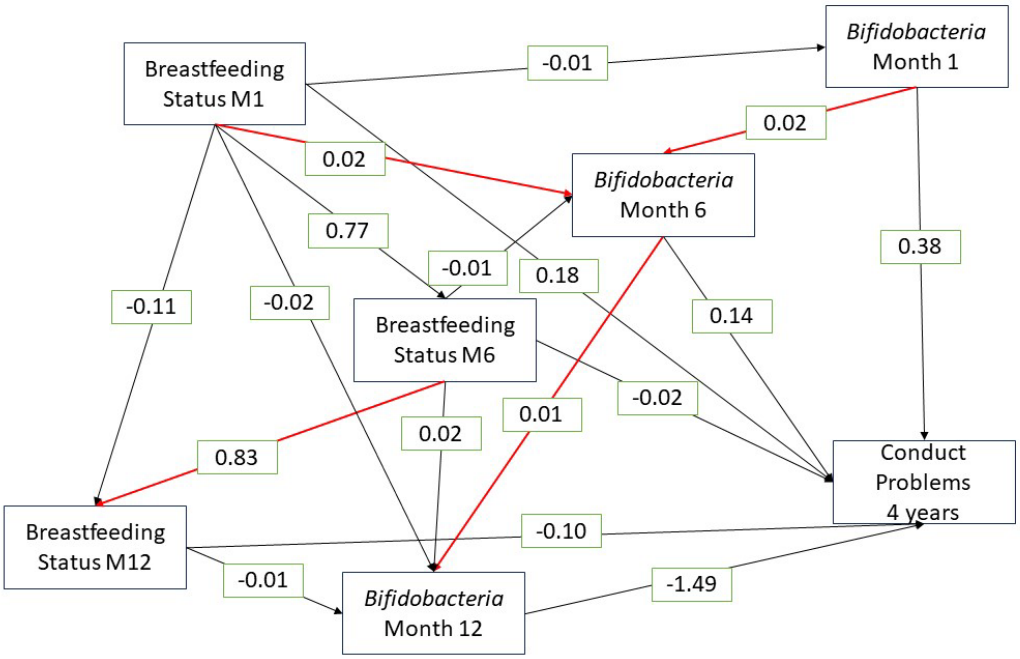
SEM 20. Emotional problems, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 *p* = 0.374, CFI = 0.990, and RMSEA = 0.928).



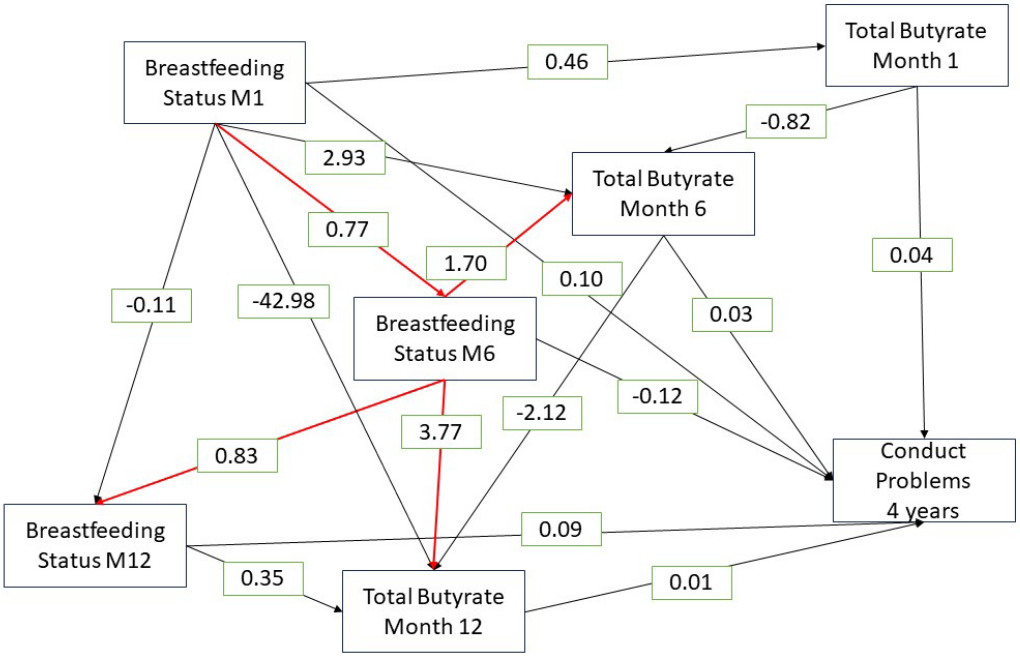
SEM 21. Emotional problems, total butyrate producing bacteria, and breastfeeding status (X^2 *p* = 0.193, CFI = 0.972, and RMSEA = 0.827).



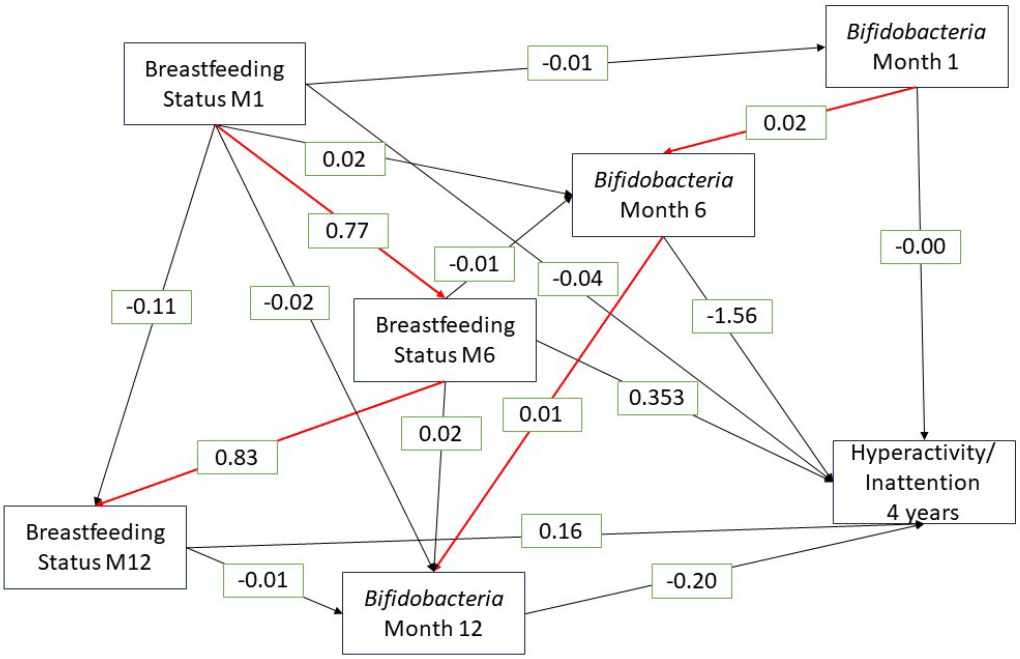
SEM 22. Conduct problems, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 p = 0.33, CFI = 0.985, and RMSEA = 0.912).



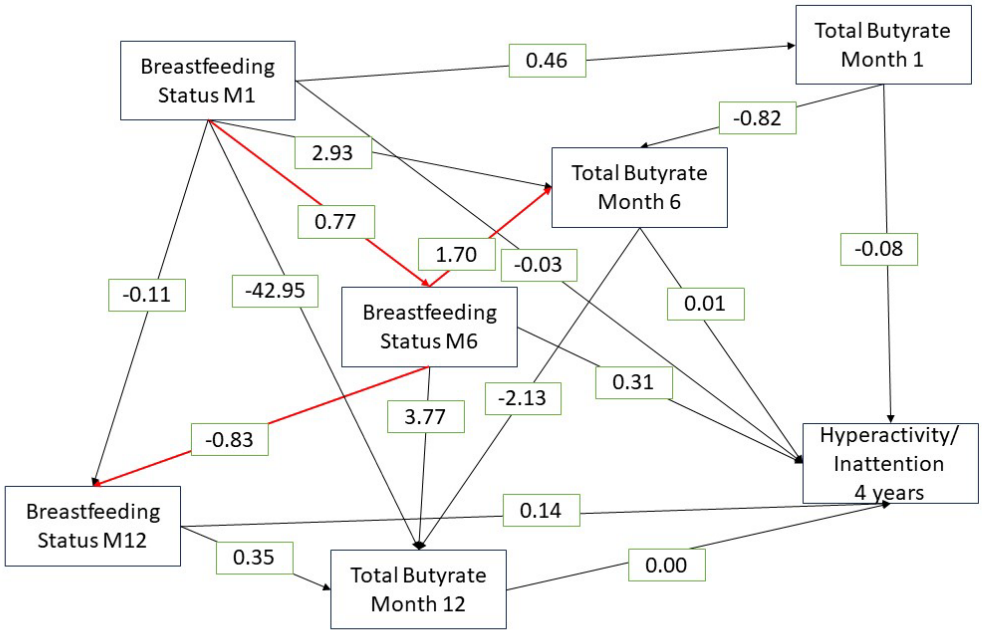
SEM 23. Conduct problems, total butyrate producing bacteria, and breastfeeding status (X^2 p = 0.145, CFI = 0.965, and RMSEA = 0.775).



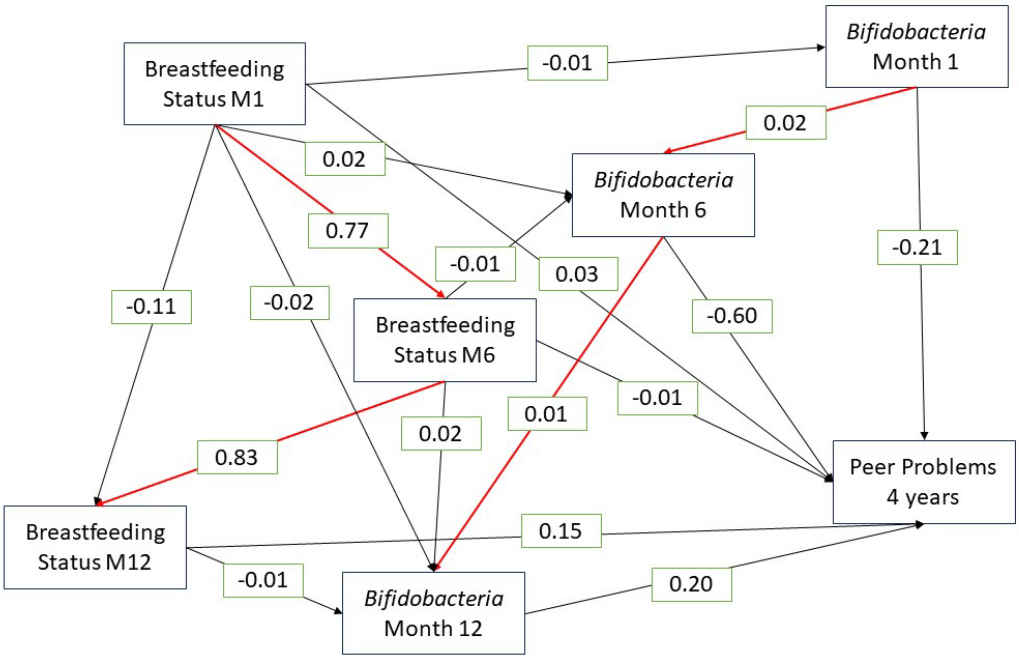
SEM 24. Hyperactivity/inattention, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 *p* = 0.168, CFI = 0.963, and RMSEA = 0.838).



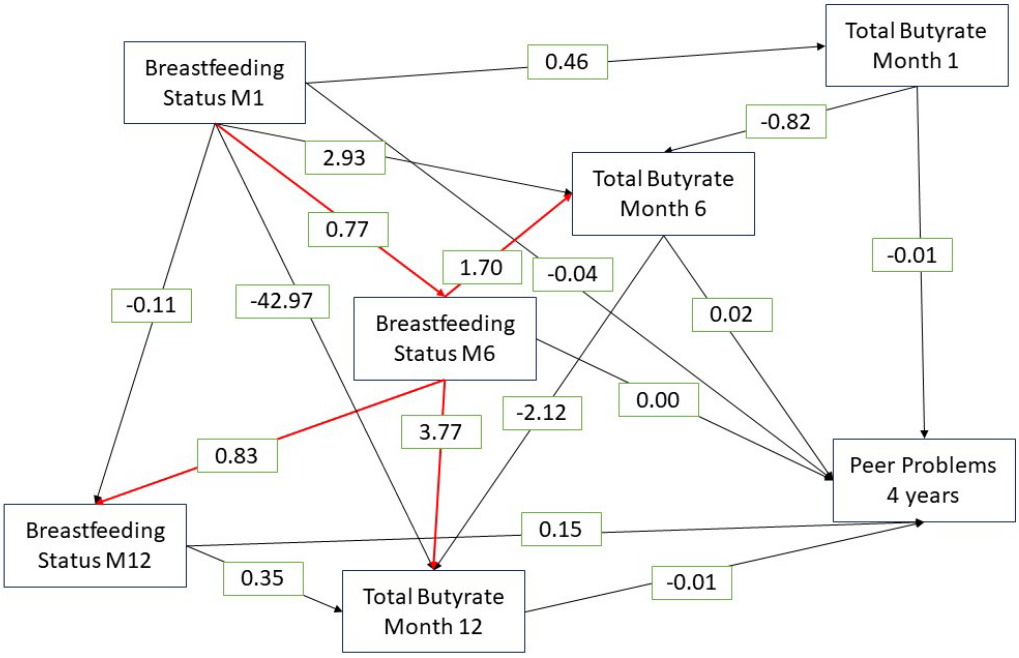
SEM 25. Hyperactivity/inattention, total butyrate producing bacteria, and breastfeeding status (X^2 *p* = 0.262, CFI = 0.980, and RMSEA = 0.872).



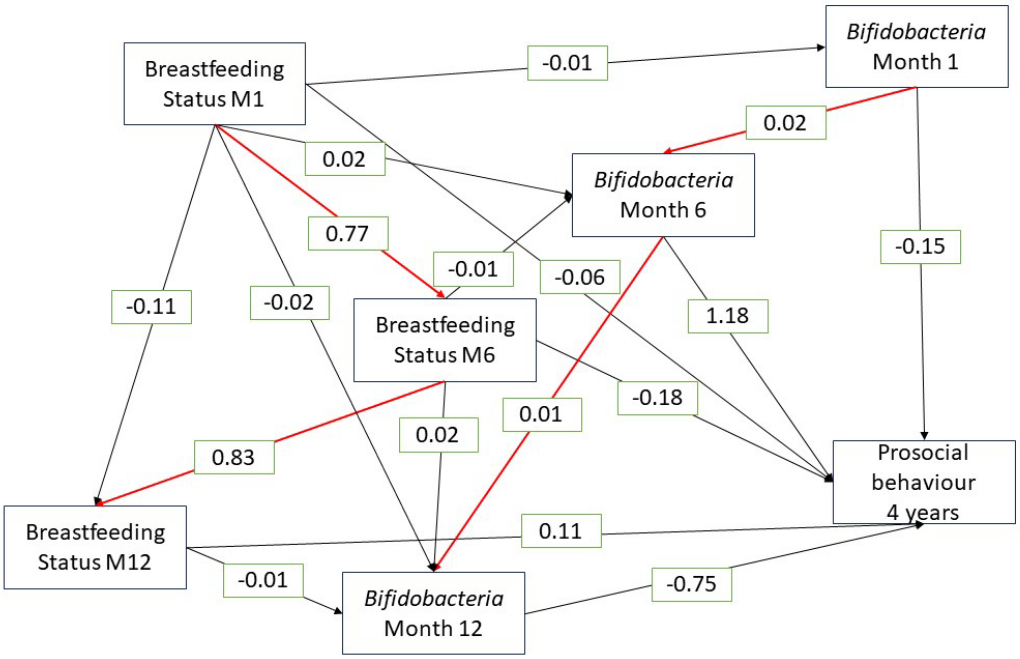
SEM 26. Peer problems, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 p = 0.394, CFI = 0.990, and RMSEA = 0.954).



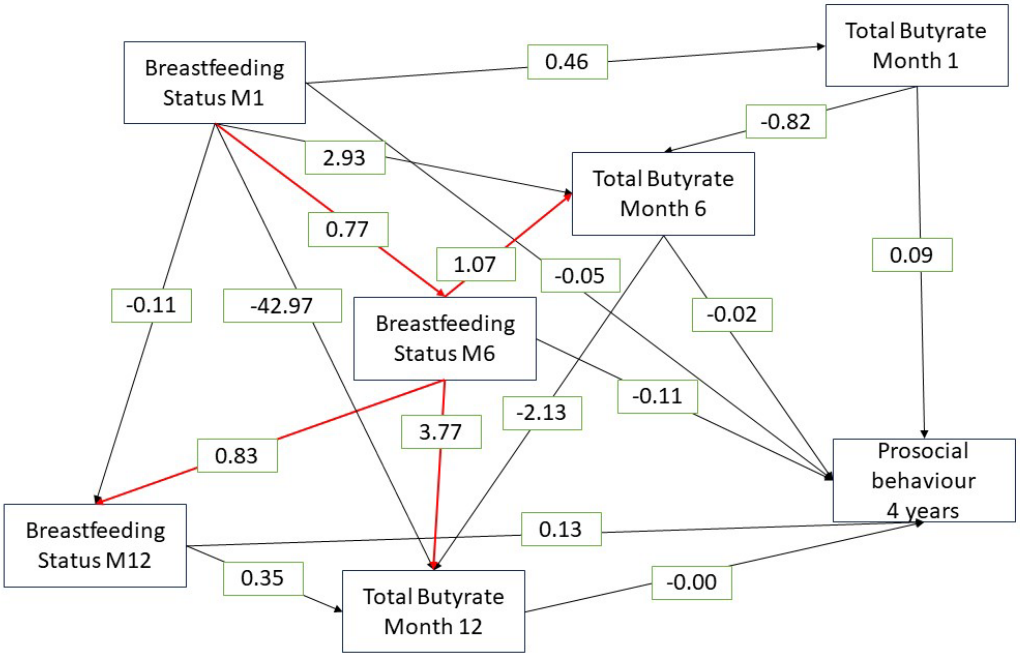
SEM 27. Peer problems, total butyrate producing bacteria, and breastfeeding status (X^2 p = 0.209, CFI = 0.971, and RMSEA = 0.877).



SEM 28. Prosocial behaviour, relative abundance of *Bifidobacteria*, and breastfeeding status (X^2 p = 0.475, CFI = 0.999, and RMSEA = 0.970).



SEM 29. Prosocial behaviour, total butyrate producing bacteria, and breastfeeding status (X^2 p = 0.195, CFI = 0.986, and RMSEA = 0.868).



Appendix 5

PCA coefficients for Months 6 and 12.

Table 3.5. PCA coefficients Month 6

Food or food Group (frequency)	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
Breastfeeding status	0.13	-0.08	0.38	0.04	0.07	-0.30
Fruits	0.06	0.15	-0.04	**0.40**	0.38	0.10
Vegetables	-0.02	0.35	-0.03	0.38	**0.51**	-0.31
Pulses	**0.46**	0.20	0.14	0.32	-0.15	0.26
Cereals	-0.05	0.25	-0.02	0.12	0.38	-037
Meat	0.10	0.38	0.25	**0.41**	0.06	**0.44**
Fish	0.08	0.24	0.31	-0.30	**0.40**	**0.51**
Pasta	**0.51**	0.08	0.00	0.37	-0.27	0.22
Rice	0.34	0.14	-0.08	0.24	-0.24	-0.18
Yoghurt	**0.62**	-0.06	-0.04	-0.04	-0.27	-0.21
Probiotic Yoghurts	0.28	0.24	0.13	0.05	-0.19	-0.08
Fried Food	**0.56**	-0.17	-0.01	-0.30	0.05	-0.24
Pre-prepared Food	0.29	-0.38	0.36	0.01	0.07	-0.12
Raw Food	0.10	**0.64**	**-0.45**	-0.27	-0.03	-0.03
Cooked Food	-0.14	**0.69**	**-0.49**	-0.21	-0.16	0.03
Organic Food	-0.02	-0.39	-0.12	-0.14	0.25	0.38
Nut	0.31	0.38	**0.47**	0.35	0.20	-0.07
Eggs	0.19	0.39	**0.58**	-0.33	-0.10	-0.06
Soy	**0.54**	-0.16	**-0.43**	-0.16	0.21	0.03
Sesame Products	**0.71**	-0.08	-0.22	-0.01	-0.14	0.11
Sesame Seeds	**0.44**	-0.17	-0.34	-0.03	0.33	0.08

Table 3.6. PCA coefficients Month 12

Food or food Group (frequency)	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Component 7
Breastfeeding status	-0.01	-0.31	**-0.42**	0.01	0.31	0.14	-0.35
Fruits	0.07	-0.04	0.13	**0.52**	-0.10	-0.14	-0.12
Vegetables	0.18	0.33	-0.30	-0.13	-0.26	0.18	-0.16
Pulses	**0.62**	0.21	0.00	0.01	-0.24	0.06	0.06
Cereals	**0.48**	0.10	-0.20	0.01	0.00	0.13	0.19
Meat	0.25	0.11	**-0.49**	0.21	0.30	-0.39	0.09
Fish	0.23	0.07	-0.11	**0.54**	-0.10	0.00	-0.26
Pasta	0.39	-0.01	0.18	0.25	0.18	0.27	0.21
Rice	**0.51**	0.09	-0.13	0.30	-0.07	0.15	0.28
Yoghurt	-0.04	-0.13	**0.50**	0.22	0.20	0.39	0.10
Probiotic Yoghurts	0.14	0.35	-0.17	-0.28	0.16	0.39	-0.10
Fried Food	0.05	-0.19	0.16	-0.08	0.12	**-0.40**	0.62
Pre-prepared Food	0.24	**-0.49**	-0.12	-0.32	-0.16	0.12	0.10
Raw Food	-0.34	**0.69**	-0.01	0.04	0.01	0.09	0.21
Cooked Food	-0.35	**0.74**	0.12	0.07	0.17	0.01	0.06
Organic Food	0.16	0.16	0.26	-0.11	**-0.52**	-0.33	-0.30
Nut	0.35	0.14	0.04	**-0.42**	0.34	0.05	-0.01
Eggs	0.36	0.30	0.04	-0.01	0.35	-0.33	-0.22
Soy	0.26	0.25	0.09	-0.26	-0.25	-0.23	0.15
Sesame Products	0.34	0.02	**0.56**	-0.06	0.00	0.22	-0.17
Sesame Seeds	0.27	-0.03	0.39	-0.09	**0.42**	-0.29	-0.29

Appendix 6

Scoring Strengths and Difficulties Questionnaire for ages 4-17

Therapist Use Only

Client code -

Scoring completed: Date __/__/____

SDQ completed by: Client / Parent/Carer / Teacher

For report: Initial / Interim / End of Therapy

Scoring Strengths and Difficulties Questionnaire for age 4-17 or 18+

SDQ is a globally recognised instrument for assessing the mental health status for children and young people. SDQ scoring provides a rough overview to help detect mental health issues, however it does not provide a clear-cut screening.

[For further information regarding SDQ visit: https://SDQscore.org/ or see: Goodman, A. & Goodman, R. (2009) 'Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health', *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), pp. 400-403.]

The 25 items in the SDQ comprises of 5 scales of 5 items each.

'Somewhat True' is always scored as 1, but the scoring of 'Not True' and 'Certainly True' varies with the item, as indicated in each scoring table (below).

Scoring Process:

1. Score each of the 5 scales, set out in the tables below (total range for each is 0 - 10).
2. Calculate the *Internalising* and *Externalising* scores, using the formula (page 3).
3. Work out the **Total Difficulties Score** (page 3).
4. For clients aged 4-17 years, use the Classification table (page 4), to plot information from SDQ symptom and impact scores (i.e. SDQs completed by parents/carers, teachers and clients themselves, where appropriate).
5. Complete the Reporting table (page 5), then copy this table across to relevant Initial, Interim and End of Therapy reports.
6. Where 2-sided SDQs have been completed, use the 'Impact Supplement' section to calculate scores (page 5).

Conduct problems Scale

		Not True	Somewhat True	Certainly True	**SCORE**
ITEM 5:	Often has temper tantrums or hot tempers (*I get very angry)*	0	1	**2**	
ITEM 7:	Generally obedient… *(I usually do as I am told)*	**2**	1	0	
ITEM 12:	Often fights with other children… *(I fight a lot)*	0	1	**2**	
ITEM 18:	Often lies or cheats *(I am often accused of lying or cheating)*	0	1	**2**	
ITEM 22:	Steals from home, school or elsewhere *(I take things that are not mine)*	0	1	**2**	
				TOTAL Conduct Problems Scale Score:	

Hyperactivity scale

		Not True	Somewhat True	Certainly True	**SCORE**
ITEM 2:	Restless, overactive… *(I am restless…)*	0	1	**2**	
ITEM 10:	Constantly fidgeting or squirming *(I am constantly fidgeting….)*	0	1	**2**	
ITEM 15:	Easily distracted, concentration wanders *(I am easily distracted)*	0	1	**2**	
ITEM 21:	Thinks things out before acting *(I think before I do things)*	**2**	1	0	
ITEM 25:	Sees tasks through to the end… *(I finish the work I am doing)*	**2**	1	0	

TOTAL Hyperactivity Scale Score:	

Emotional problems scale					
		Not True	Somewhat True	Certainly True	**SCORE**
ITEM 3:	Often complains of headaches… *(I get a lot of headaches…)*	0	1	**2**	
ITEM 8:	Many worries… (*I worry a lot)*	0	1	**2**	
ITEM 13:	Often unhappy, downhearted… (*I am often unhappy….)*	0	1	**2**	
ITEM 16:	Nervous or clingy in new situations… (*I am nervous in new situations…)*	0	1	**2**	
ITEM 24:	Many fears, easily scared *(I have many fears…)*	0	1	**2**	
				TOTAL Emotional Problems Scale Score:	

Peer problems scale					
		Not True	Somewhat True	Certainly True	**SCORE**
ITEM 6:	Rather solitary, tends to play alone *(I am usually on my own)*	0	1	**2**	

ITEM 11:	Has at least one good friend *(I have one goof friend or more)*	**2**	1	0	
ITEM 14:	Generally liked by other children *(Other people my age generally like me)*	**2**	1	0	
ITEM 19:	Picked on or bullied by other children… *(Other children or young people pick on me)*	0	1	**2**	
ITEM 23:	Gets on better with adults than with other children *(I get on better with adults than with people my age)*	0	1	**2**	
				TOTAL Peer Problems Score:	

Prosocial scale					
		Not True	Somewhat True	Certainly True	**SCORE**
ITEM 1:	Considerate of other people's feelings *(I try to be nice to other people)*	0	1	**2**	
ITEM 4:	Shares readily with other children… *(I usually share with others)*	0	1	**2**	
ITEM 9:	Helpful if someone is hurt… *(I am helpful is someone is hurt…)*	0	1	**2**	
ITEM 17:	Kind to younger children *(I am kind to younger children)*	0	1	**2**	
ITEM 20:	Often volunteers to help others… *(I often volunteer to help others)*	0	1	**2**	
				TOTAL Conduct Problems Scale Score:	

Internalising and *Externalising* scores:

The internalising score (range 0-20) is the sum of the emotional and peer problems scales.

Emotional Problems Score +	Peer Problems Score =	***Internalising* Score**

The externalising score (range 0-20) is the sum of the conduct and hyperactivity scales.

Conduct Score +	Hyperactivity Score =	***Externalising* Score**

N.B. Using these two amalgamated scales may be preferable to using the four separate scales in community samples, whereas using the four separate scales may add more value in high-risk samples.

Total Difficulties Score: This is generated by adding scores from all the scales except the prosocial scale. The resultant score ranges from 0 to 40. **Anything of 17 or above is HIGH.**

Conduct+	**Hyperactivity +**	**Emotional +**	**Peer =**	**TOTAL SDQ SCORE**

Categorisation bands for SDQ scores for age 4-17:

Use the following table to identify (CIRCLE) the correct classification for each score.

SDQ Scale	Close to Average (80% pop)	Slightly raised (/lowered) (10% pop)	High (/Low) (5% pop	Very high (/very low) (5% pop)
Parent/Carer completed SDC				
Emotional problems score	0-3	4	5-6	7-10
Conduct problems score	0-2	3	4-5	6-10
Hyperactivity score	0-5	6-7	8	9-10
Peer problems score	0-2	3	4	5-10
Prosocial score	8-10	7	6	0-5
Externalising score	0-7	8-10	11-13	14-20
Internalising score	0-3	4-7	8-10	11-20
Total difficulties score	0-13	14-16	17-19	20-40
Teacher completed SDQ				
Emotional problems score	0-3	4	5	6-10
Conduct problems score	0-2	3	4	5-10
Hyperactivity score	0-5	6-7	8	9-10
Peer problems score	0-2	3-4	5	6-10
Prosocial score	6-10	5	4	0-3
Externalising score	0-5	6-10	11-12	13-20
Internalising score	0-3	4-8	9-10	11-20
Total difficulties score	0-11	12-15	16-18	19-40

Self-completed SDQ				
Emotional problems score	0-4	5	6	7-10
Conduct problems score	0-3	4	5	6-10
Hyperactivity score	0-5	6	7	8-10
Peer problems score	0-2	3	4	5-10
Prosocial score	7-10	6	5	0-4
Externalising score	0-5	6-10	11-12	13-20
Internalising score	0-4	5-8	9-10	11-20
Total difficulties score	0-14	15-17	18-19	20-40

N.B. Although SDQ scores can be used as continuous variables, it is sometimes convenient to categorise scores. The initial bandings presented for the SDQ scores were 'normal', 'borderline' and 'abnormal'. These bandings were defined based on a population-based UK survey, attempting to choose cut-points such that 80% of children scored 'normal', 10% 'borderline' and 10% 'abnormal'.

More recently a four-fold classification has been created based on an even larger UK community sample. This four-fold classification differs from the original in that it (1) divided the top 'abnormal' category into two groups, each containing around 5% of the population, (2) renamed the four categories (80% 'close to average', 10% 'slightly raised, 5% 'high' and 5% 'very high' for all scales except prosocial, which is 80% 'close to average', 10% 'slightly lowered', 5% 'low' and 5% 'very low'), and (3) changed the cut-points for some scales, to better reflect the proportion of children in each category in the larger dataset.

Note that these cut-points have not been validated for use with the 18+ SDQ, so we suggest that it is safest to use continuous scores rather than categories for this measure

Reporting Table:

Scale	Score	Category	Population Average
	(Range 0- 10)		
Conduct Problems			

Hyperactivity Scale			
Emotional Problems			
Peer Relationships			
Prosocial Scale			
	(Range 0-20)		
Externalising			
Internalising			
	(Range 0-40)		
Total SDQ score *N.B. Anything of 17 or above is HIGH.*			

Impact Supplement:

When using a version of the SDQ that includes an 'Impact Supplement' (questions overleaf from the SDQ), the items on overall distress and impairment can be added up to generate an impact score that ranges from 0 to 10 for parent- and self-report, and from 0 to 6 for teacher-report.

Scoring the SDQ Impact Supplement	Not at all	Only a little	A medium amount	A great deal
Parent/Carer report:				
Difficulties upset or distress child	0	0	1	2
Interfere with HOME LIFE	0	0	1	2
Interfere with FRIENDSHIPS	0	0	1	2
Interfere with CLASSROOM LEARNING	0	0	1	2
Interfere with LEISURE ACTIVITIES	0	0	1	2

Teacher report:				
Difficulties upset or distress child	0	0	1	2
Interfere with PEER RELATIONS	0	0	1	2
Interfere with CLASSROOM LEARNING	0	0	1	2
Self-report:				
Difficulties upset or distress child	0	0	1	2
Interfere with HOME LIFE	0	0	1	2
Interfere with FRIENDSHIPS	0	0	1	2
Interfere with CLASSROOM LEARNING Interfere with LEISURE ACTIVITIES	0 0	0 0	1 1	2 2

Responses to the questions on chronicity and burden to others are not included in the impact score. When respondents have answered 'no' to the first question on the impact supplement (i.e. when they do not perceive themselves as having any emotional or behavioural difficulties), they are not asked to complete the questions on resultant distress or impairment; the impact score is automatically scored zero in these circumstances.

Appendix 7

PPI poster advertisement

Aston University

PEACh

Volunteers needed for Research study

Early childhood development: The relationship with the gut microbiome and diet

Tasks: A short zoom call, where you will:

1. Watch a short 5 minute video
2. Answer 13 follow up questions.

Duration: 20 minutes

Who can Participate:

- Parents of children aged between 0-10-years
- Expectant parents
- Fluent in English
- Access to a computer

You will receive a £10 gift voucher for participating

For more information or to participate in the study please contact:

Emma Alving-Jessep

e.alving-jessep@aston.ac.uk

Appendix 8

PPI study Questionnaire

Gut Microbiome and Behaviour Questionnaire

Please could you input the 5 digit participation code that was sent to you.

Thank you for watching our project proposal video. We would like you to imagine now that you have received this information during pregnancy and answer the following questions as if you were considering whether to take part in the study.

Question 1.

Was the information provided to you in the video clear and easy to understand?

- Completely
- Somewhat
- Not very
- Not at all

Question 2.
Is the reason for doing the study clear?

- Yes
- No

Question 3.

If you answer to question 2 was No, can you explain what other information you would need?

[Open answer]

Question 4.

Would you be willing to be contacted for just over 2-years for the duration of the project?

- Yes, definitely
- Maybe
- I don’t know

- Probably not
- Definitely not

Question 5.

Would you be willing to complete the questionnaires for the project?

- Yes, definitely
- Maybe
- I don’t know
- Probably not
- Definitely not

Question 6.

Would you be happy to donate a sample of poo from the mother?

- Happy
- Somewhat happy
- Neither happy not unhappy
- Somewhat unhappy
- Not happy at all

Question 7.

Would you be happy to donate samples of poo from the child?

- Happy
- Somewhat happy
- Neither happy not unhappy
- Somewhat unhappy
- Not happy at all

Question 8.

Would you be happy to donate samples of breastmilk?

- Happy
- Somewhat happy
- Neither happy not unhappy
- Somewhat unhappy

- Not happy at all

Question 9.

Would you be happy to record yourself and child during mealtime interactions?

- Happy
- Somewhat happy
- Neither happy not unhappy
- Somewhat unhappy
- Not happy at all

Question 10.

Do you think that the reward for taking part is enough?

- Yes
- No

Question 11.

Is there anything that would put you off taking part in the project?

[Open answer]

Question 12.

Is there anything that you think can be improved in the project?

[Open answer]

Question 13.

Please share with us any other comments or suggestions you have for us here.

[Open answer]

Appendix 9

PowerPoint slides presented in PPI Video presentation

Points to keep in mind

- What do you think could be improved in the project?
- Is there anything that would put you off taking part?
- What do you like or find interesting about the project?
- What would you find an acceptable reward for taking part?

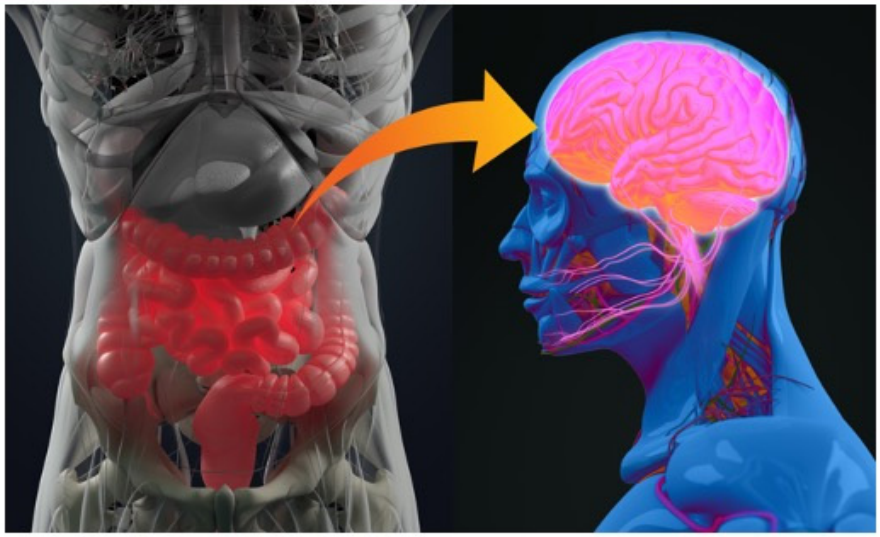
2

Patient, Participant Involvement (PPI)

- Imagine that you are being asked to take part in this study
- We want to hear how we can improve the study
- And increased people's willingness to take part

3

What are we researching?

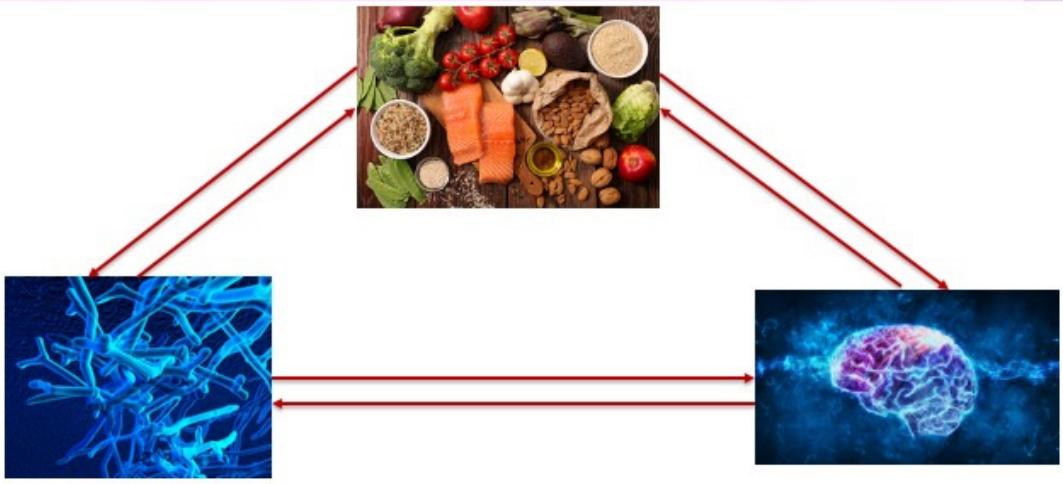


4

What are we researching?

5

What are we researching?



6

When we will be contacting you.

We will recruit pregnant women who have had a healthy 20 week scan.

The first participation contact for the study will be at 36-38 weeks of pregnancy.

7

When we will be contacting you.

The first participation contact for the study will be at 1 month after the birth of your child.

8

When we will be contacting you.

The Second participation contact for the study will be at 6 months after the birth of your child.

9

When we will be contacting you.

The Third participation contact for the study will be at 12 months after the birth of your child.

10

When we will be contacting you.

The Final participation contact for the study will be at 2 years after the birth of your child.

11

What will you be asked to do

- Complete questionnaires about your health, the health of your child, and about their behaviour

12

What will you be asked to do

- Complete a Diet diary for the Mother's diet and the child's diet

13

What will you be asked to do

- To take video recordings of yourself and your child during a meal time at 6 months, 12 months and 1 year.

14

What will you be asked to do

- Donate a sample of your own Poo at 36-38 weeks of pregnancy, and also your child's at each of the remaining visits

15

What will you be asked to do

- Donate a sample of your breastmilk at 4 weeks, 6 months, 12 months and 2 years, if you are breastfeeding.

16

What will you be asked to do

- To attend a lab visit at 12 months and 2 years of age, for the experimenter to assess your child's development.

17

Important information

- All information that is given to the project at Aston University will be kept confidentially and securely.
- We will not be using any of the samples donated to diagnose any health problems.
- All samples are stored according to the Human Tissue Act.
- Participation is not mandatory, you have the right to withdraw.
- We will not be giving individual assessments of a child's development, however a summary of overall project findings will be issued after analysis.

18

What will you get in return?

- For each of the 5 sessions that you have with the project you will receive a £20 gift voucher. As there are 5 sessions you will receive £100, if you complete all sessions.

- Additionally, there is a £100 voucher that will be won by one participant in a prize draw once the project has been completed.

19

Aston University
BIRMINGHAM UK

Thank you.

Please could you now complete the very short questionnaire.

20