Finding new answers in old trials with data linkage

A novel method to assess whether nutrient intake in infancy affects long-term cognition.

Maximiliane Lara Verfürden

A thesis submitted for the degree of
Doctor of Philosophy

Great Ormond Street Institute of Child Health (GOS ICH),
University College London (UCL)

May 2021

Declaration

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

Abstract

Background Due to high attrition in randomised controlled trials (RCTs), cognitive effects of infant formula modifications remain uncertain. The aim of this thesis was to test a new method to minimise attrition and, through doing so, to compare differences in academic performance between children previously randomised to either nutritionally modified or standard infant formula.

Methods Nine dormant infant formula RCTs conducted in England (1982-2001) were available for linkage to the National Pupil Database. Linkage was based on legal exemption from the need for participant consent. A trusted third party provided de-identified data for up to four candidate pupil matches per participant and agreement-metrics for all shared linkage variables. I completed the linkage of de-identified data, using auxiliary RCT variables and probabilistic methods. Six RCTs (n=1,563) were eligible for analysis, and a further three RCTs were used to assess linkage success and improve multiple imputation of missing data. Participant academic performance was measured using exam grades, with the primary outcome being General Certificate of Secondary Education (GCSE) Maths grades at age 16 years. Modified formula and standard formula groups were compared on an intention-to-treat basis, stratified by trial.

Results Within the six trials eligible for analysis, primary outcome data was available for 86% of all participants. Available outcome data was substantially higher than the average of 22% above age 2 years in previous consent-based cognitive follow-ups of the trials. There was no evidence of benefit for GCSE Maths performance for any type of modified formula. Secondary academic outcomes provided weak evidence of harm for one of the formula modifications.

Conclusions Unconsented linkage of dormant trials to administrative education data is feasible and leads to higher follow-up rates compared to traditional consented follow-up methods. None of the investigated nutritionally modified formula interventions improved academic performance.

Impact statement

Infant formula is consumed globally by over 60% of infants aged less than six months. Unlike other early interventions to support cognitive development, infant formula modifications are highly scalable. They are reviewed and regulated centrally, resulting in the potential to affect a large group of infants worldwide. In recent years, infant formula research has advanced substantially. This includes recommendations to increase nutrient supply for preterm and small-for-gestational-age infants, and changes to the lipid formulation and mineral content of formulas. However, high-quality evidence on the long-term risks and benefits of modifications for improving cognitive outcomes was lacking due to high rates of attrition in existing randomised controlled trials with follow-up to adolescence. However, as modifications were established practice, it was difficult to argue for equipoise to conduct new trials. The work presented in this thesis addresses this gap by linking dormant (i.e., historical) infant formula trials to administrative education data to compare differences in academic performance between children previously randomised to either nutritionally modified or standard infant formulas. I used section 251 (NHS Act 2006) support, instead of consent, to permit the linkage, because the potential for public benefit justified temporarily lifting the common law duty of confidentiality. This method of follow-up substantially reduced attrition compared to consent-based follow-up methods. I report evidence of no benefit and weak evidence of worse academic performance, associated with some formula modifications. These findings demonstrate that monitoring of long-term effects should be done routinely to detect safety issues and to ensure optimal development for all formula-fed infants. The use of unconsented linkage of dormant trials to administrative data makes it eminently feasible to provide new, timely and important answers from dormant RCT data that is already available, minimising costs, waiting times, and attrition bias.

Acknowledgments

The research that has gone into this thesis has been really enjoyable – largely as a result of the interactions I have had with my brilliant supervisors and colleagues. I feel extremely privileged to have received the trust of my supervisors Professor Ruth Gilbert, Professor Mary Fewtrell, and Professor John Jerrim. To each of them I owe a great debt of gratitude for their inspiration, patience, and invaluable support, academic and personal. I would also like to express my gratitude to Dr Katie Harron, Professor Bianca De Stavola, and Professor Tim Cole, who have spent hours of their time brainstorming ideas with me.

Thanks are also due to my examiners Dr Chris Gale and Dr Richard Silverwood, for their time spent examining my thesis and the rewarding discussion we had during my viva.

The Economic and Social Research Council UCL, Bloomsbury and East London Doctoral Training Partnership and the Great Ormond Street Hospital Charity were very generous in providing me with scholarship assistance, and the Population Policy and Practice Programme at the UCL Great Ormond Street Institute of Child Health has provided an excellent environment for my research. I feel very lucky to have spent many enjoyable hours at the institute before the Covid-19 campus closure.

Finally, I am so grateful to my wonderful kæreste Philip, my godparents Sonni and Chrischi, Annette van den Bussche, Claire Dale, and my friends – especially Poppy Mallinson and Rafael Albert –, who have all been extremely supportive, loving, and encouraging.

To my grandfather Heinrich, who would be very happy, and without whom none of this would have come to fruition.

Contents

List of Figures

CHAPTER 7

CHAPTER 8

List of Tables

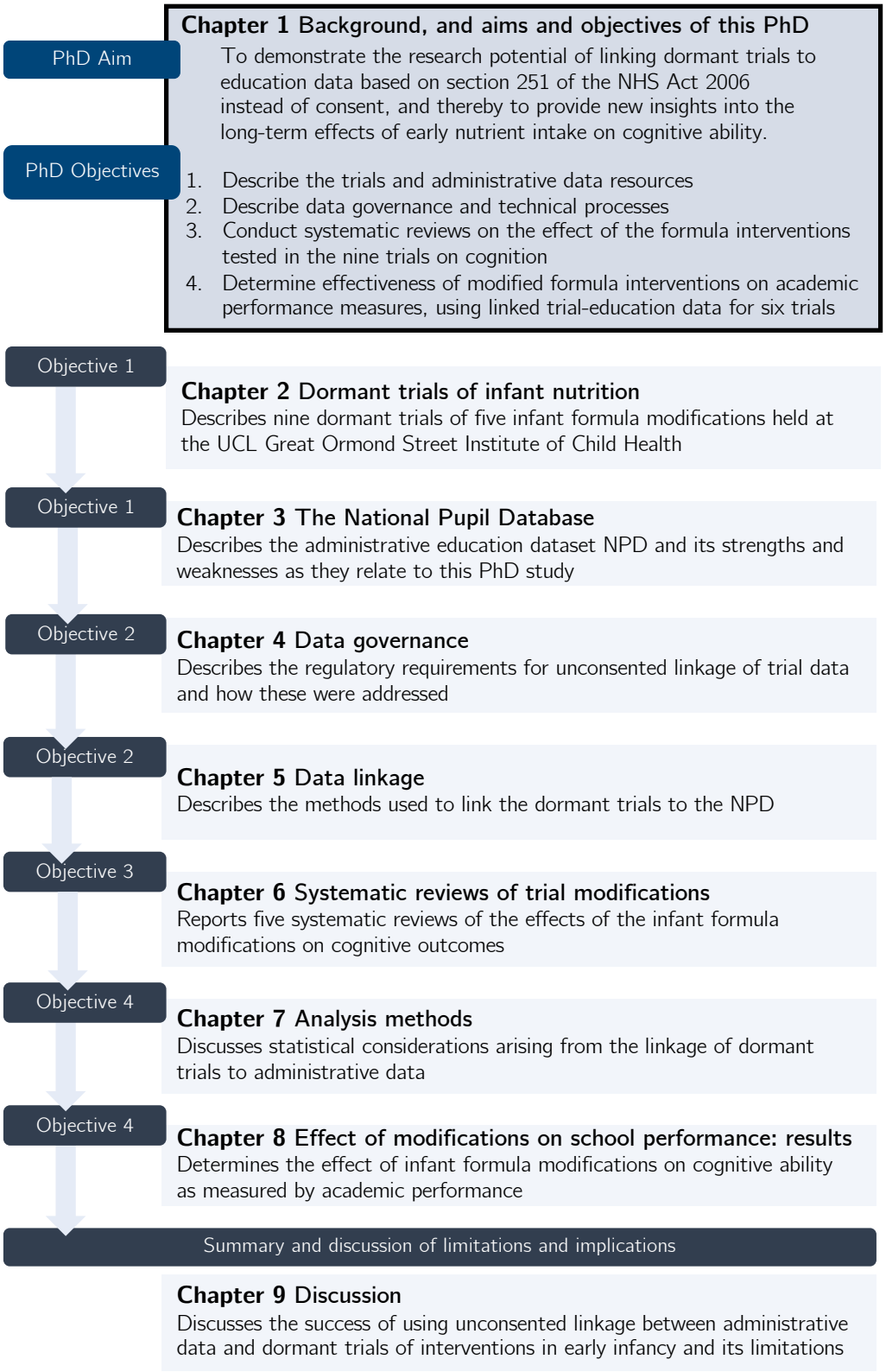
List of Definitions

Concept:	*Definition, specific to this thesis:*
Attrition	Loss of participants to follow-up over time.
Blinding	Keeping participants and/or personnel unaware of formula group assignment.
Dormant	A trial is referred to as dormant when primary outcomes have already been collected.
Deterministic linkage	Linkage algorithms based on pre-defined rules of agreement between linkage variables.
False match	A participant is linked to a pupil record that is not their own.
Infant formula	A food for feeding to babies and infants under 12 months of age, usually prepared for bottle-feeding from powder or liquid.
Linkage	Combining data from two or more datasets.
Linkage error	Missed matches or false matches.
M-probability	Probability of two records achieving a specific agreement level given they belong to the same child.
Match weight	Match weights represent the likelihood of records being a match.
Missed match	No matching pupil record is found for the participant.
Multiple imputation	Process of replacing missing data with multiple substituted values and combining them using Rubin's rules.
Parent	Where parents are mentioned in this thesis, this refers to the person(s) responsible for the care of the child, irrespective of genetic relationship.
Preterm	Babies born alive before 37 weeks of pregnancy.
Probabilistic linkage	Likelihood based linkage.
Section 251	Section 251 of the NHS act allows a temporary lift of the common law duty of confidentiality so that confidential patient information can be processed.
U-probability	Probability of two records achieving a specific agreement given they do not belong to the same child.
Unconsented	Using Section 251 as the legal basis for data processing.

Abbreviations

Abbreviation:	*Description:*
AA	Arachidonic acid
ALA	Alpha-linolenic acid
ALSPAC	Avon Longitudinal Study of Parents and Children
BSID	Bayley Scales of Infant Development
CA	Corrected age
CAG	Confidentiality advisory group
DARS	Data Access Request Service
DfE	Department for Education
DHA	Docosahexaenoic acid
DPA	Data Protection Act
DSA	Data sharing agreement
EFSA	European Food Safety Authority
ESPGHAN	European Society for Paediatric Gastroenterology Hepatology and Nutrition
ESRC	Economic and Social Research Council
FFT	Fischer Family Trust
GCSE	General Certificate of Secondary Education
GDP	Gross domestic product
GOS ICH	Great Ormond Street Institute of Child Health
HRA	Health Research Authority
IG	Information Governance
IQ	Intelligence Quotient
IRAS	Integrated Research Application System
IRON	RCT of term formula fortified with iron for terms (1993-94)
KS	Key Stage
LA	Linoleic acid
LCPUFA	Long-chain polyunsaturated fatty acids
LCPUFAP	RCT of preterm formula supplemented with long-chain polyunsaturated fatty acids for preterm infants (1993-96)
LCPUFAT	RCT of term formula supplemented with long-chain polyunsaturated fatty acids for terms (1993-95)
MAR	Missing at random
MCAR	Missing completely at random
MDI	Mental development index

MI	Multiple imputation
n	Number of participants in subgroup
N	Total number of participants
NEP-1	RCT of nutrient-enriched (mainly protein and calorie) term formula for preterm infants vs banked breast milk
NEP-2	RCT of nutrient-enriched (mainly protein and calorie) term formula for preterm infants vs standard term formula (1982-84)
NEP-PD	RCT of nutrient-enriched (mainly protein and calorie) term formula for preterm infants (as post-discharge formula) (1993-96)
NETSGA	RCT of nutrient-enriched (mainly protein and calorie) term formula for babies born at term, small for gestational age (1993-96)
NHS	(UK) National Health System
NPD	National Pupil Database
NUCLEO	RCT of term formula supplemented with nucleotides for term infants (2000-02)
PALM	RCT of term formula with sn-2 palmitate for terms (1995-96)
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
SD / σ	Standard deviation
SEN	Special Educational Needs
SGA	Small for gestational age
TTP	Trusted third party
UCL	University College London
UK	United Kingdom
WAIS	Wechsler Adult Intelligence Scale
WASI	Wechsler Abbreviated Scale of Intelligence
wga	Weeks gestational age
WISC	Wechsler Intelligence Scale for Children
WPPSI	Wechsler Preschool and Primary Scale of Intelligence
y	Years


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 1 Background, aims and objectives

1.1 Introduction

Infant formulas, intended for babies and infants under 12 months of age, are among the most heavily regulated products in the EU (EuroVoc, 2020). Despite this, knowledge about the safety and effectiveness of different infant formula compositions on cognitive function is uncertain (EFSA NDA Panel, 2014).

International regulatory bodies aim to ensure that infant formulas are safe, meet the nutritional requirements, and promote growth and healthy development of infants who are being fed formula as their sole source of nutrition (Koletzko et al., 2005). To ensure these benefits, only ingredients that serve a nutritional or other benefit should be added to infant formula, with the ultimate goal of bringing physiological (e.g., growth patterns), biochemical (e.g., faecal microbiota), and functional outcomes (e.g., cognitive function) in formula-fed infants closer to those of healthy and exclusively breastfed babies, who fare better than formula-fed babies on such outcomes and represent the gold standard in infant nutrition (Koletzko et al., 2005, Victora et al., 2016).

Measures of physiological, biochemical outcomes, and of most functional outcomes such as visual acuity and immune response, tend to be well-defined and validated (Sun et al., 2015). By contrast, measures of cognitive function are frequently criticised for being prone to bias (Colombo, 2018) and poorly predictive of real-world outcomes such as academic performance and employment (Sun et al., 2015).

Are there better ways to measure cognitive ability, and what are the implications for public health? This chapter explores the factors that contribute to the knowledge gap, argues why scarcity of knowledge on cognitive effects is an important public health problem, and proposes a solution to address this gap: linking dormant infant nutrition trials to administrative education data in adolescence without the need for participant consent.

1.2 The importance of cognitive ability in infant formula research

1.2.1 Relevance for public health

Globally, fewer than 40% of babies below the age of 6 months are exclusively breastfed (Victora et al., 2016), with numbers even lower in the UK (McAndrew et al., 2012). That means a considerable proportion of infants depend on infant formula to be safe and beneficial. If formula modifications could bring cognitive outcomes of formula-fed babies closer to those of healthy, exclusively breastfed babies, the effect on a population level would be substantial. For example, the Intelligence Quotient (IQ) difference between breastfed and formula-fed children is reported to be about 6 points (Kramer et al., 2008). A one-point national difference in IQ is associated with a 0.11% increase in annual GDP per capita (Jones and Schneider, 2006) – a significant economic impact. On an individual level, low cognitive ability is a recognised risk factor for fewer occupational opportunities and lower socioeconomic status, and it predicts poor adult health and decreased life-expectancy (Hanley et al., 2010, Deary and Batty, 2007). Having valid and reliable data on cognitive effects of infant formula modifications is critical to inform evidence-based policy decisions to prevent sub-optimal cognitive ability in formula-fed infants. The effect of infant formula composition is, therefore, of considerable public health interest.

1.2.2 Definition of cognitive ability

Cognitive ability may be defined as *"mental capability that (...) involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience"* (Gottfredson, 1997). It is often used interchangeably with the terms intelligence, mental ability, and cognitive function (Deary and Batty, 2007). People differ in how accurately and fast they perform cognitive work, and researchers have long been interested in identifying the drivers of these differences (Deary, 2020). One factor that is likely to affect cognitive ability is diet, particularly during the first year of life.

1.2.3 Nutritional programming of cognitive ability

Humans undergo a period of rapid brain development in early infancy when they are thought to be particularly sensitive to external stimuli such as nutrient supply (Lucas, 1998, Lucas and Sampson, 2006, Lucas, 1991). The idea that such early stimuli could cause long-term, potentially irreversible, effects on the structure or function of the brain has been around for centuries (Spalding, 1873). Coined by the controversial* neuro-endocrinologist Dörner (1976), the concept of *programming* was taken up and popularised in the field of infant nutrition, notably through Barker and Lucas. Barker (1986, 1990, 1992) conducted a series of observational studies in which he linked infant size, a marker of in-utero nutrition, to heart disease in adulthood. Lucas' group then tested whether similar associations could be observed for nutritional status in infancy. The group conducted the first randomised controlled trials of early nutrition, suggesting potential developmental effects of enriching formula for preterm babies with extra nutrients (Lucas, 1991).

How might early nutrition cause changes in cognitive ability? From the time of conception, the brain develops rapidly. It overproduces neural connections (synapses) from shortly before birth until early childhood. **Fig. 1.1** illustrates this process for different brain regions and their corresponding functions. At the earliest age, development is greatest for basic functions and sensory processing. After that, simple lower-order cognitive functions, such as attention and memory, develop. This is followed by the development of the ability to behave according to set goals and learn rules (Best and Miller, 2010). In early to mid-childhood, higher-order abilities such as strategic decision making and problem-solving are developing (Colombo, 2018). In the early stages of brain development, the abundance or absence of certain nutrients (e.g., protein, calories, or essential fatty acids) might cause changes in physiology, metabolism, and cell development that result in effects on lower-order cognition (attention,

* Dörner argues that manipulating sex hormone levels of pregnant women can "prevent their children from becoming homosexual". His research became the scientific basis for his 1970s campaign to prevent homosexuality.

memory). These could then cascade into higher-order effects (language, problem-solving), resulting in long-term consequences for cognitive ability. Firm evidence that the availability or absence of certain nutrients in infancy results in the adaptation of irreversible cognitive trajectories would be of considerable public health importance and justify immense investments into optimising early nutrition for preventative purposes (Lucas, 1998). There is evidence that malnutrition-induced stunting is negatively correlated with cognitive development (Grantham-McGregor et al., 2007). However causal nutritional programming effects on cognitive ability have not been conclusively established for any nutrient due to challenges in measuring cognitive ability throughout different points in the child's life course.

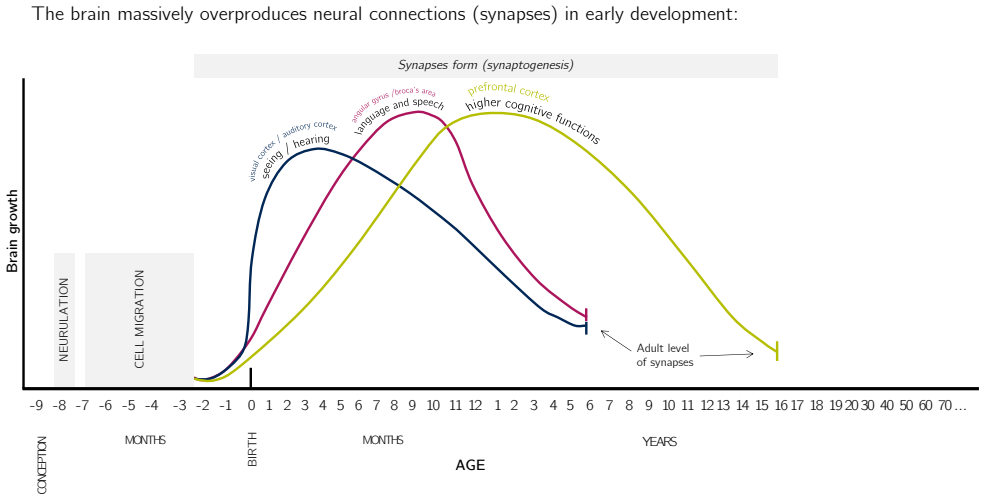


Fig. 1.1: Overproduction of synapses in different brain regions indicates sensitive periods in brain development. Adapted from Thomson, R. A., & Nelson, C. A. (2001). Developmental science and the media: early brain development. American Psychologist 56(1), 5-15, with permission from the American Psychological Association

1.2.4 Challenges of measuring cognitive ability in infant nutrition studies

Cognitive effects of infant nutrition are likely to be confounded by social factors, such as poverty or maternal education, which simultaneously affect a child's diet and cognitive ability (Lucas and Sampson, 2006). Thus, the most valid and

reliable way of determining whether a cognitive effect is attributable to a new ingredient added to infant formula is through randomised controlled trials (RCTs) in humans. Yet so far, RCTs primarily relied on short-term cognitive ability measures, most commonly the Bayley Scale of Infant Development (BSID) (**Table 1.1**). Choosing short-term measures is reasonable, given the financial limitations of large trials and likely loss of participants over time (Colombo, 2018). However, researchers now believe that such early measures are not adequate to serve as primary endpoints for trials (Sun et al., 2015, Colombo and Carlson, 2012). While the BSID is intended to reflect developmental attainment, the measure reports only slight differences between breastfed and formula-fed infants (Andres et al., 2012) – groups shown to differ substantially in their cognitive ability in later childhood (Kramer et al., 2008). Measures of infant development, such as the BSID, are easily limited by circumstance: scores are assigned by the outcome assessor and may be influenced by the assessor's impression of the accompanying parents' education or social standing, or even by the degree of cooperativeness that the infant shows on that specific day.

Moreover, early developmental measures are also poorly predictive of later cognitive function in childhood or adolescence (Sun et al., 2015). Predictive value is important in the context of reports that some infant formula modifications could carry long-term risks. For example, an RCT, which randomised 1,120 healthy term infants to either high-iron formula or low-iron formula, observed adverse effects on cognition in the high-iron group at age ten years (Lozoff et al., 2012) and again at 16 years (Gahagan et al., 2019) but not in earlier follow-ups (Walter et al., 1998). Similarly, a trial investigating the effect of long-chain polyunsaturated fatty acid fortification of formula reported no effect on cognition at age 18 months (Lucas et al., 1999) but significant adverse effects on IQ at age seven years (Lucas et al., unpublished). Long-term follow-up is, therefore, clearly indicated for both scientific and safety reasons.

Wechsler IQ tests are the most frequently used tests of cognitive ability in long-term follow-up studies and are typically administered from age three years onwards (**Table 1.1**, on page 28). They are also better at predicting

academic achievement.† In fact, the original raison d'être for standardised cognitive ability tests was to identify pupils who were at risk of poor academic performance (Binet and Simon, 1916). The Wechsler tests cover various cognitive domains such as verbal comprehension, perceptual reasoning, working memory, and processing speed. Tasks are adapted to the age range at which they are administered. Test scores are standardised, so the average of the relevant age range is 100 points, with a standard deviation of 15 points.

However, IQ scores might not be fit for purpose either, as doubts exist about their ability to measure the individuals' *capacity* to perform (Rao and Georgieff, 2000). A wide body of literature has shown that standard cognitive outcomes such as IQ tests are considered "low-stakes" for participants, which can lead to low test-effort (Attali, 2016), potentially introducing substantial measurement error into the results (Akyol et al., 2018, Duckworth et al., 2011).

Furthermore, small improvements in IQ scores are less tangible and, by themselves, unlikely to have any major long-term consequences. In contrast, small changes in academic performance, another measure of cognitive ability (**Table 1.1**), can lead to a difference in final grades. Recent research illustrated how falling just above or just below critical grade boundaries can have significant consequences for academic progression (Machin et al., 2020), making academic performance a more policy-relevant outcome. Academic grades have also been shown to be strong predictors of later labour market outcomes (Hayward et al., 2014), making it easier to translate the impact of infant formula modifications on the economy.

Apart from the high costs associated with long-term follow-up (Llewellyn-Bennett et al., 2018), the main challenge of measuring long-term outcomes is participant attrition. "*The whole [infant nutrition] field is bedevilled by the fact that formula-fed infants are particularly difficult to follow up long-term*" was

† Correlations between WISC at 8 years and GCSE grade attained at age 16 years in the ALSPAC Cohort: Maths: 0.6294 (n=6,043) English: 0.566 (n=6,167) MORRIS, T. 10 May 2018 RE: Correlation of IQ and school attainment in "How well can we predict educational outcomes? Examining the roles of cognitive ability and social position in academic achievement" (Personal communication to Verfuerden M.)

printed in a summary report of a large 2010 Nestlé Nutrition workshop (Lucas et al., 2010).

Participants may be lost because they withdraw, die, emigrate, or become untraceable. Attrition is more common in nutrition trials than in therapeutic trials because, in therapeutic trials, participants with a specific condition are often motivated to contribute to the advancement of their treatment (Fewtrell et al., 2016). Furthermore, studies in infants initially rely on the consent of their parents. Once the children are old enough to consent for themselves, they might have other priorities than to make time for continued participation in a randomised controlled trial. Attrition of study participants has an impact on the robustness of research findings by affecting the validity, reliability, and generalisability of the results: participants who consent to be followed-up are likely to be healthier and are also more likely to differ in other (measured and unmeasured) factors from the participants who are lost to follow-up; this can lead to false-positive or false-negative findings (Fewtrell et al., 2016). As sample sizes decrease from attrition, trials also lose their statistical power. Consequently, attrition can increase the risk of failing to demonstrate existing associations between the formula composition and later outcomes (Rothman et al., 2008).

Table 1.1: Measures of cognitive ability measures discussed in this thesis

Measure	Age range	Description	Common challenges
Bayley Scores (BSID)	1-42 months	Standardised test to identify children with a developmental delay. Produces two scores: mental and motor score, with the mental score being my primary focus.	Noisy and poorly predictive of later cognitive ability outcomes. Does not identify differences where differences are expected
Wechsler preschool and primary Scale of Intelligence (WPPSI)	3-7.5 years	The most used tests of cognitive ability.	Low stake so might not reflect capacity to perform, context-dependent, high cost of follow-up.
Wechsler Intelligence scale for children (WISC)	6-16 years		
Wechsler Adult Intelligence scale (WAIS)	16-90 years		
Wechsler Abbreviated Scale of Intelligence (WASI)	6-90 years		
Academic performance (in the UK)	5-18 years	Independently marked grades on high-stakes exams	Complexity of information governance and data access

1.2.5 Uncertainty in current policy recommendations

Globally, compositional standards for infant formulas are specified by the Codex Alimentarius and, in the EU, regulated by the recently updated European Directive. Changes in regulations tend to be based on the scientific opinions of independent expert bodies, such as the European Food Safety Authority (EFSA) and the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN). EFSA and ESPGHAN base their scientific opinions on the composition of breast milk, estimated nutrient requirements in infancy, and published evidence from randomised controlled trials (RCTs) for short- and long-term benefits and harms (EFSA Panel on Dietetic Products and Allergies, 2014). However due to the absence of robust evidence on many functional outcomes,

the majority of infant formula composition standards relies on expert consensus alone. This means there is the potential for considerable harm not only through suboptimal regulations, but also because such regulations may limit equipoise and could thereby prevent future trials.

Effects on cognitive ability have been proposed but not confirmed for a range of formula modifications, including protein and calorie enrichment for preterm babies (Embleton, 2013, Embleton et al., 2021) and small-for-gestational-age babies (Lin et al., 2019), addition of long-chain polyunsaturated fatty acids (LCPUFA), which was recently mandated in the EU (EFSA NDA Panel, 2014), and iron fortification (EFSA NDA Panel, 2014). The next section introduces a potential solution to address the paucity of high-quality evidence on long-term cognitive outcomes of infant formula modifications.

1.3 Reactivating dormant infant formula trials with administrative education data

1.3.1 A potential solution to address the evidence gap

Data from randomised controlled trials with limited attrition are needed to create robust infant policy recommendations. As evidence on long-term cognitive trajectories is scarce, studies that track cognitive ability over time, using several different measures, would be especially valuable. As it takes an average of 17 years for new research evidence to turn into clinical practice (Morris et al., 2011), methods that could address remaining uncertainties sooner rather than later could prevent suboptimal cognitive ability for a large number of children. The potential solution that I propose in this thesis tackles the challenges of attrition, timeliness, and cost simultaneously: the reactivation of *dormant* infant formula trials through linkage to administrative education data on the basis of section 251 NHS Act 2006 instead of participant consent.

Dormant trials are existing trial cohorts where the primary outcome has already been measured, often without plans of further follow-up (Henry and Fitzpatrick, 2015). Advances in the availability and quality of administrative datasets have opened up the possibility of reactivating dormant infant formula

trials with linkage to national education data to answer open questions on long-term cognitive effects. Reactivation is possible where participant identifiers (e.g., contact details) have been retained, and governance arrangements allow secure linkage without consent to limit attrition from non-response. In the UK, regulatory bodies allow the use of unconsented linkage for research on the grounds of substantial expected benefits to the public and evidence that consent-based follow-up would not be feasible or would not produce valid answers to an important research question (Cross et al., 2020).

Unconsented linkage of trials to administrative data also enables researchers to ascertain long-term outcomes now, rather than having to start a new study and wait for participants to reach a particular age. Participant cohorts old enough to have passed through the whole academic trajectory have data on a range of academic outcomes that could then be compared to any previously measured in-trial cognitive outcomes such as IQ or Bayley score to assess correlations.

Linkage to administrative data would also be cost-effective: a recent systematic review evaluated the cost of different methods of post-trial follow-up for 65 studies (Llewellyn-Bennett et al., 2018) and found that linkage to administrative records outperformed more traditional follow-up methods (i.e., postal correspondence, face-to-face appointments, review of paper-based medical records, and telephone interviews).

To my knowledge, so far, there has been no systematic effort to reactivate dormant trials using unconsented linkage – neither for infant formula trials nor any other trials. A series of nine dormant infant formula trials held at the UCL Great Ormond Street Institute of Child Health retained participant identifiers and is therefore eligible for linkage to administrative data. I used this series of trials to determine the effect of nutritionally modified infant formula on cognitive ability as measured by academic performance. This PhD study is intended as a proof-of-concept study to demonstrate the processes and added value of linking trials without consent to administrative data to measure long-term effects. In addition, findings from this PhD study may be used to inform infant feeding policy by generating more robust evidence on long-term cognitive effects of certain infant formula modifications.

1.4 PhD aims and objectives

The aim of my thesis was twofold:

1: To demonstrate the processes and added value of trial linkage, legally based on section 251 of the NHS Act 2006 instead of consent, to administrative education data to measure long-term cognitive ability ...

and thereby

2: ... to determine the effect of nutritionally modified infant formula on cognitive ability as measured by academic performance from a series of dormant infant formula trials held at the UCL Great Ormond Street Institute of Child Health.

In line with these aims, my thesis addressed the following objectives:

Aim 1

1) Describe the trial and administrative data resources:
    i) Dormant infant formula trials
    ii) National Pupil Database (NPD)

2) Describe data governance and technical processes:
    i) Information governance requirements
    ii) Digitisation of participant identifiers
    iii) Data linkage

Aim 2

3) Conduct systematic reviews on the effect of the nutritionally modified formulas on cognition

4) Determine the effectiveness of modified infant formulas for improving academic performance measures, using linked trial-education data for six dormant infant formula trials

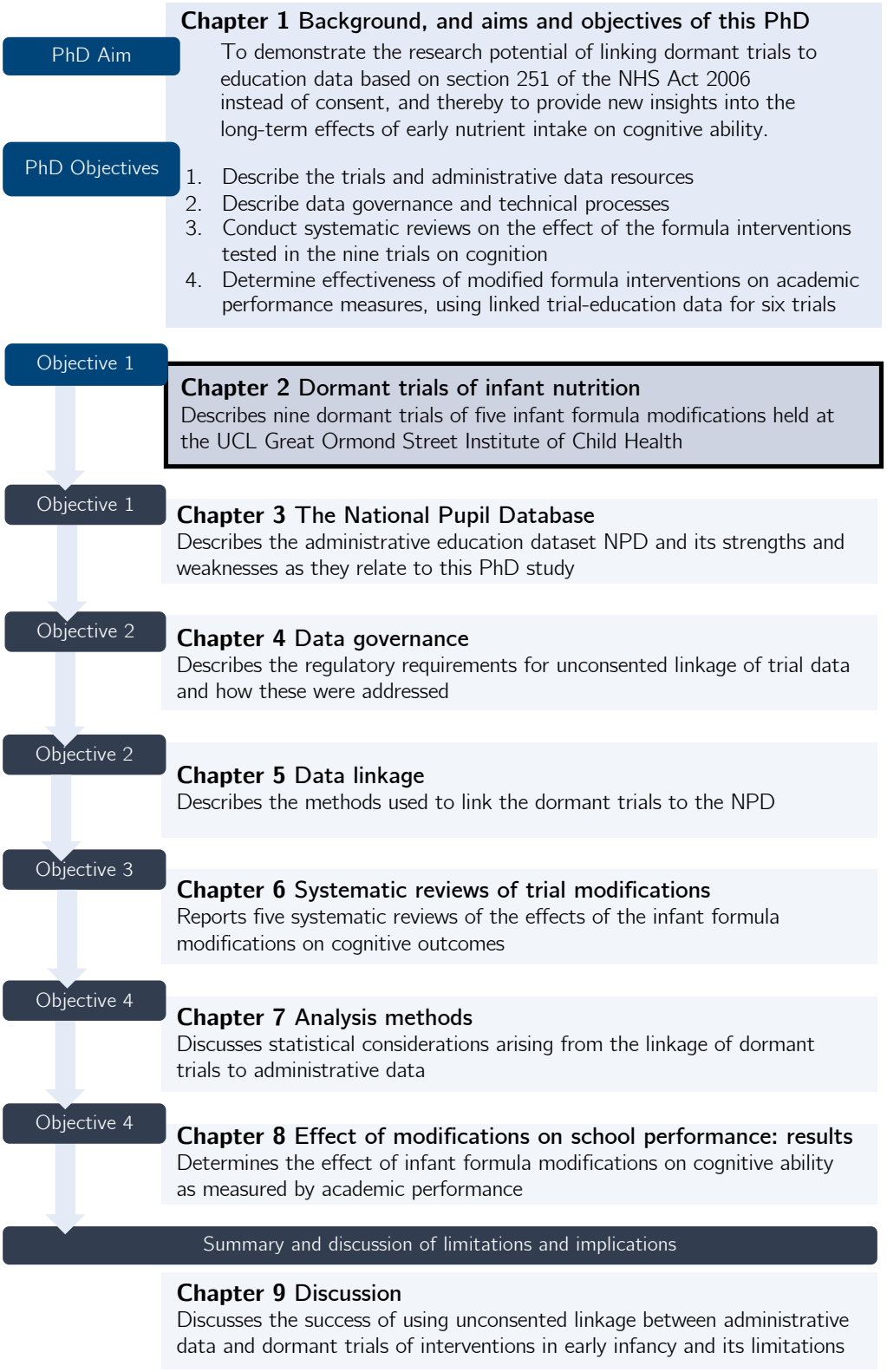
1.5 Thesis structure

The next chapter (Chapter 2) describes the nine dormant infant nutrition trials used in this thesis. The administrative education dataset (the National Pupil Database) is described in Chapter 3. The ethics and data governance aspects of using unconsented linkage to follow up trial participants are discussed in Chapter 4, while Chapter 5 describes the technicalities of the linkage processes. Chapter 6 presents five systematic reviews, which formally assess what is already known about the cognitive effects of the different infant formula modifications investigated in the dormant infant nutrition trials. The statistical issues and methods for the effectiveness analysis are discussed in Chapter 7. The effectiveness of modified infant formulas for improving academic performance measures, using linked trial-education data for six trials, is presented in Chapter 8. A summary and discussion of this thesis are provided in Chapter 9, including implications for practice and further research.

1.6 Key points from Chapter 1

- On average, physiological, biochemical, and functional outcomes of formula-fed babies are worse than of those of breastfed babies.
- Because a large proportion of babies worldwide depend on formula milk, formula modifications that benefit cognitive ability could have a substantial impact at the population level.
- Infant formula policy recommendations are based on the composition of breast milk, estimated nutrient requirements in infancy, and published evidence from randomised controlled trials (RCTs) for short- and long-term benefits.
- Attrition of study participants is a worse problem for cognitive measures compared to most measures of physiological, biochemical outcomes, and functional outcomes such as visual acuity and immune response. This is because early cognitive measures are less well-validated and have poor predictive ability of cognitive ability in adolescence and adulthood.

- As a result, knowledge about the long-term effect of different infant formula compositions on cognitive function is highly uncertain and compositional recommendations might be at risk of not promoting optimal future health.
- Several nutrients have the theoretical potential for affecting cognitive ability, including protein and calorie enrichment for preterm babies and small-for-gestational-age babies, supplementation of formula with long-chain polyunsaturated fatty acids (LCPUFA), and supplementation of formula with iron.
- This thesis proposes a potential solution to address attrition in research investigating infant formula effects on cognitive ability. The solution involves reactivating dormant infant nutrition trials by linking them to administrative education data legally based on support under section 251 of the NHS Act 2006 instead of consent. This could limit participant attrition from non-response and provide timely answers on cognitive effects at a lower cost compared to traditional follow-up methods.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 2 Data source: dormant trials held at the UCL Great Ormond Street Institute of Child Health

2.1 Chapter structure and content

As discussed in Chapter 1, RCT evidence on cognitive effects of infant formula modifications is uncertain. To address this uncertainty, I proposed to link dormant infant formula trials to administrative education data through the route of section 251 NHS Act 2006 support. There are several pre-requisites to enable such a linkage and obtain valid estimates of long-term cognitive effects. Among them is that the dormant trials are of high internal and external validity and that they have retained high-quality identifiers enabling linkage to administrative data.

This chapter presents work towards objective 1: *"to describe the trial and administrative data resources"*. I describe a series of nine dormant trials selected based on their high internal and external validity and retention of identifiers, with the aim to demonstrate the processes of the linkage and, for a subset of trials, to determine the effect of nutritionally modified infant formula on cognitive ability as measured by academic performance.

2.2 Role of individual trials for this thesis

A significant proportion of infant nutrition trials conducted in industrialised countries in the past 40 years were led by researchers across the UK, now based at the UCL Great Ormond Street Institute of Child Health (UCL GOS ICH). For this PhD project, nine of these, now dormant, trials were selected from the UCL GOS ICH archives to determine the long-term cognitive effects of nutritionally modified infant formula. The nine dormant trials were chosen because they have the following characteristics in common: i) they tested infant formula modifications that are widely available; ii) they were all conducted in England and have retained personal identifiers of sufficient quality, enabling linkage to English administrative school data.

Differences in long-term cognitive effects between modified and standard formula groups could be investigated for six trials in this thesis: NEP-PD, NETSGA, LCPUFAP, LCPUFAT, IRON, and PALM (see **Table 2.1**, page 38 for a key to the trial abbreviations). Based on biological plausibility, only the

first five trials tested modifications hypothesised to affect cognitive ability. The PALM trial tested a modification that was not expected to have any cognitive effects and therefore served as a negative control for the analysis (see section 7.8, page 178 for an explanation of negative controls).

Only a small proportion of participants in the NEP-1 and NEP-2 trials, born between 1982-84, had records in the NPD as most had already completed their schooling before school data in England was systematically collected. Participants in the NUCLEO trial were too young to have data on the primary outcome (information on the primary outcome in section 7.4.1, page 20) at the time of linkage. I, therefore, did not report cognitive results for the NEP-1, NEP-2, and NUCLEO trials. Participants in the three trials were instead used to evaluate the linkage algorithm (the algorithm is described in section 5.2.2.2, from page 85) and contribute additional information to improve the multiple imputation model (multiple imputation is discussed in section 7.6.2, from page 170). Although cognitive ability outcomes are not available for these trials in this thesis, other measures could become available in the future, such as GCSE records for the NUCLEO trial or the individual learning record or earnings data for NEP-1 and 2. Therefore, the data from the NEP-1, NEP-2, and NUCLEO trial was prepared, so that linkage of these three trials to available administrative datasets is imminently possible.

The trials were diverse in terms of their populations (preterm born babies, babies born small-for-gestational-age, and healthy term babies), the number of follow-up assessments, and the years in which they recruited participants (**Table 2.2**). This diversity was an asset because it allowed me to explore the research potential of unconsented data linkage across a range of different settings, making the findings of this thesis more readily generalisable. Finally, all nine trials suffered from high participant drop-out rates, even though most trials involved dedicated paediatric investigators, known to the participants and their parents, inviting them to follow-up. Response rates have been universally below 15% by age 17 years – even in those trials where contact for follow-up was preceded by checking addresses through the NHS tracing service to ensure that addresses were correct.

Table 2.1: Trial abbreviations used in this thesis

Trial abbreviation	*Formula modification*	*Years of recruitment*
NEP-PD	Nutrient-enriched (mainly protein and calorie) term formula for preterm infants (as post-discharge formula)	1993-96
NETSGA	Nutrient-enriched (mainly protein and calorie) term formula for babies born at term, small for gestational age	1993-96
LCPUFAP	Preterm formula supplemented with long-chain polyunsaturated fatty acids for preterm infants	1993-96
LCPUFAT	Term formula supplemented with long-chain polyunsaturated fatty acids for terms	1993-95
IRON	Term formula fortified with iron for terms	1993-94
PALM	Term formula with sn-2 palmitate for terms	1995-96
NEP-1	Nutrient-enriched (mainly protein and calorie) term formula for preterm infants vs banked breast milk	1982-84
NEP-2	Nutrient-enriched (mainly protein and calorie) term formula for preterm infants vs standard term formula	1982-84
NUCLEO	Term formula supplemented with nucleotides for term infants	2000-02

2.2.1 Data sources and preparation

The data for the dormant trials was provided by Professor Mary Fewtrell on behalf of the UCL GOS ICH Nutrition Group. Clinical data (minus identifiers) was provided in the form of 18 CSV files. As 67% of variables were unlabelled, the variables of interest (randomisation and assessment dates, baseline demographic variables, cognitive test scores, and treatment group) were identified with the help of two trial investigators: Professor Mary Fewtrell and Kathy Kennedy. I cleaned and validated the data, resulting in one dataset with one row per participant. Validation involved trying to reproduce published tables of baseline characteristics and cognitive outcome results from the files I received, clarifying with the original trialists wherever there were discrepancies. This process took approximately ten months. The Stata code for this is publicly available online: https://github.com/MaxVerfuerden/PhD.

The identifier variables (names, childhood addresses, and dates of birth) were only available in paper-based format from consent and contact forms. The process of digitally recording and of validating these identifiers took one year and three months and is discussed in Chapter 5.

2.3 Description of trial designs, methods, and outcomes

2.3.1 Overview and trial designs

Information about trial design was extracted from previous publications and complemented by contacting the original trialists. All studies were parallel randomised controlled trials, investigating the superiority of one nutritionally modified infant formula over another, otherwise identical, standard formula, or banked breast milk (**Fig. 2.1**). The scientific rationales for the formula modifications and evidence context for each modification are discussed in Chapter 6.

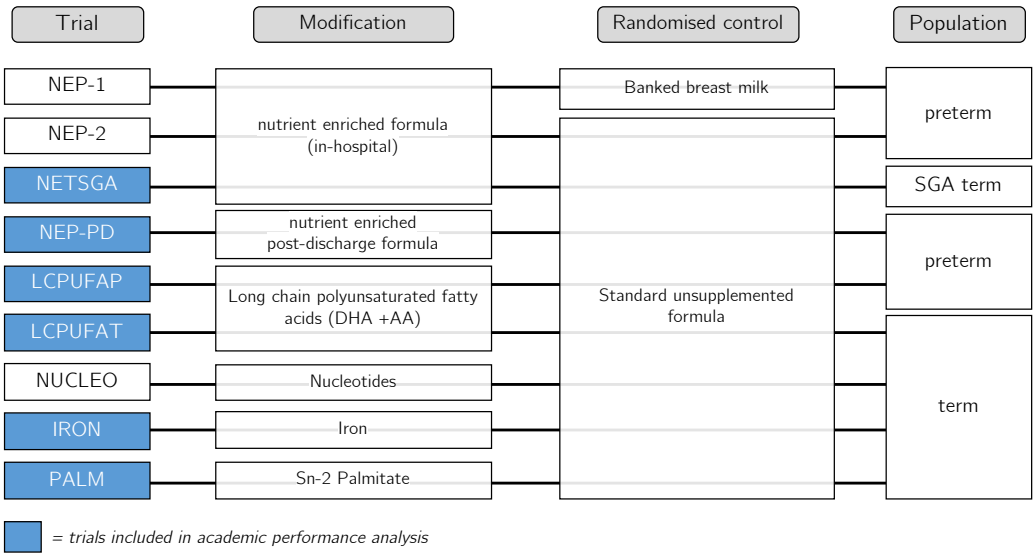


Fig. 2.1: Formula modifications, comparators, and participant populations in the dormant infant nutrition trials

Table 2.2 on page 43 outlines the characteristics of the nine dormant infant formula trials, and **Fig. 2.2** on the next page shows the number of participants followed-up throughout childhood (focusing on cognitive measures only). The trials recruited participants born between 1982 and 2002 in hospitals of seven English cities. In total, 2,551 babies were randomised in a 1:1 ratio to modified and control formula groups. In all trials but the *NEP-PD* and the *IRON* trial, the formula was given from birth, and mothers had already decided not to

breastfeed. In the *NEP-PD* trial, which was conducted in babies born preterm, the formula was supplied from 1 week before discharge from the hospital or when their bodyweight reached 2kg, irrespective of prior feeds. In the *IRON* trial, the formula was supplied at age nine months in a group of infants who were either previously breastfed or formula-fed and whose mothers now intended to feed cow's milk. Within trials, 366 children were randomised together with their twin or triplet siblings. A total of 169 children took part in both the *LCPUFAP* infant formula trial and the *NEP-PD* post-discharge formula trial, however randomisation schedules were separate for both trials, and trial periods were not overlapping.

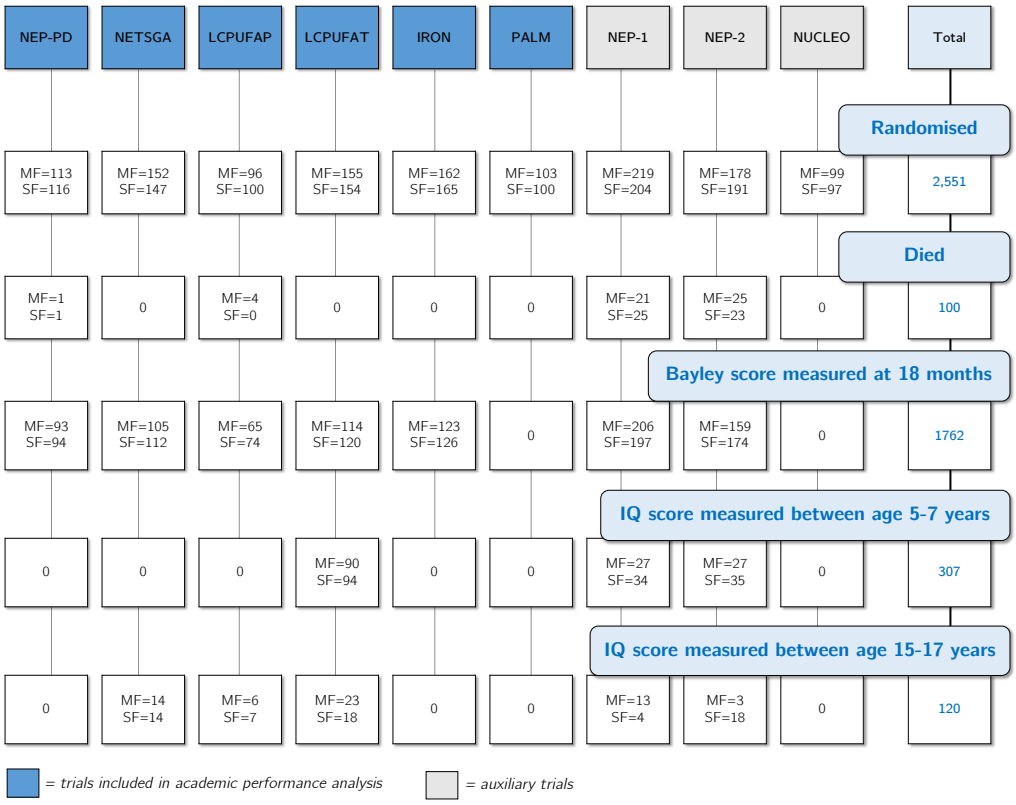


Fig. 2.2: Consort flow diagram showing number of participants followed-up throughout childhood (cognitive measures only), by trial and trial arm

2.3.2 Composition of infant formulas in modified and standard groups

The formula compositions are detailed in appendix p 2, and the nutritional rationale and background of the formula modifications are described in Chapter 6 (systematic reviews).

2.3.3 Participant demographics at randomisation

At randomisation, information on birth weight, gestational age, maternal age, infant sex, maternal smoking, and maternal education was collected in all trials. All of these factors might have a plausible independent impact on cognitive ability (Sammons et al., 2014, Botting et al., 1998, Abel et al., 2017). Therefore, imbalances between groups could introduce bias in the analysis of cognitive outcomes. **Table 2.3** and **Table 2.4** show that there were no large absolute differences between the modified formula and control formula groups with respect to these characteristics in any of the trials.

2.3.4 Randomisation, allocation concealment, blinding, and formula sponsorship

Although conducted between the 1980s and early 2000s, the RCTs meet today's research quality standards. All details on randomisation, allocation concealment and blinding can be found in appendix p 11. In brief, allocation was concealed using appropriate measures available at the time, and the randomisation sequence was generated externally. All trials except the NEP-1 and NEP-2 trials[‡] (not involved in the cognitive analysis) blinded families and personnel to formula allocation. I remained blind to group allocations until my analysis plan was peer-reviewed and accepted for publication (Verfürden et al., 2020).

[‡] The NEP-1 and NEP-2 trials were not blinded to clinical staff in the neonatal units as the formula modification was clearly distinguishable from the standard, and knowledge about the type of intervention was considered a safety measure to enable appropriate clinical management. Personnel at follow-up were blinded to the allocation.

In infant formula trials, formulas are usually obtained from formula manufacturers as they are typically not commercially available at the time of the trial. It is common that manufacturers donate formulas to the research group and support formula studies technically (e.g., with blinding to group allocation) and financially. **Table 2.2** on the next page details the companies that supplied the study formulas and the nature of their involvement in the trials. None of the formula manufacturers were involved in follow-ups beyond the primary outcome.

2.3.5 Evidence of cognitive effects from previous in-trial measures of cognitive ability

Within the dormant trials discussed in this thesis, evidence of cognitive effects for any of the formula modifications is inconsistent. Seven of these trials have previously measured cognitive ability (NEP-1, NEP-2, NEP-PD, LCPUFAP, LCPUFAT, IRON). Cognitive ability was measured using the Bayley Mental Development Index in infancy, Wechsler IQ in childhood, and Wechsler IQ in adolescence. Not all trials that measured cognitive ability assessed it at all ages. **Table 2.5** shows previous findings expressed as within-trial standardised mean differences between modified and control formula groups for the whole cohort (disregarding any effects found in subgroups). Nutrient-enriched formula had a consistent positive effect on cognition in infancy, childhood, and adolescence compared to banked breast milk in preterm infants (NEP-1 trial) and term formula in SGA term infants. However, most effect sizes were small, and none were statistically significant at the 95% level. Against the expectation of the investigators, LCPUFA supplemented term formula for term infants showed a consistent adverse effect on cognitive ability in infancy, childhood, and adolescence compared to unsupplemented term formula. However, only the effect on IQ in adolescence (-0.41 SD) was statistically significant, and 87.7% of participants were lost to follow-up, indicating a high risk of bias. None of the other trials showed consistent, large, or statistically significant effects on cognitive ability. **Table 2.5** also shows that the loss of participants increased significantly with participant age, with an average participant loss of 18.9% in infancy and 88.9% in adolescence.

Table 2.2: Characteristics of the dormant infant formula trials (continued on next page)

Trial	**NEP-PD**	**NETSGA**	**LCPUFAP**	**LCPUFAT**	**IRON**	**PALM**	**NEP-1**	**NEP-2**	**NUCLEO**
Study design									
Recruitment	1993-96	1993-96	1993-96	1993-95	1993-94	1995-96	1982-84	1982-84	2000-02
Place of recruitment	Cambridge Ipswich Nottingham Leicester	Cambridge Nottingham Leicester	Nottingham Leicester	Nottingham Leicester	Norwich Nottingham Leicester	Cambridge	Sheffield Norwich	Cambridge Ipswich Kings-Lynn	Nottingham Leicester
Population	Preterm infants: <1750g bw <37w ga	SGA term infants: bw <10th centile ≥37w ga	Preterm infants: ≤1850g bw ≤37w ga	Term infants: ≥37w ga	Term infants: bw > 2500g, ≥37w ga age: 9m	Term infants: ≥37w ga, bw>5th centile	Preterm infants: <1850 g	Preterm infants: <1850 g	Term infants: ≥37w ga
Intervention timing	1w < discharge – 9m	Birth – 12w	Birth – 3w	Birth – 6m	9m – 18m	Birth – 12w	Birth - 30d	Birth - 30d	Birth - 20w
Avg. number of attended follow-ups per participant (SD)	5.0 (1.5)	4.8 (1.9)	1.7 (0.8)	4.1 (1.8)	5.5 (1.0)	3.1 (0.8)	4.0 (1.3)	2.8 (1.1)	2.7 (0.9)
In-trial cognitive outcomes collected	Bayley (18m)	Bayley (18m) IQ (16y)	Bayley (18m) IQ (16y)	Bayley (18m) IQ (5y) IQ (17y)	Bayley (18m)	n/a	Bayley (18m) IQ (8y) IQ (16y)	Bayley (18m) IQ (8y) IQ (16y)	n/a
Risk of bias									
Random sequence generation:	✔	✔	✔	✔	✔	✔	✔	✔	✔
Allocation concealment:	✔	✔	✔	✔	✔	✔	✔	✔	✔
Blinding:	✔	✔	✔	✔	✔	✔	✘	✘	✔
Formula sponsor	Farley (now Heinz)	Farley (now Heinz)	Milupa (now Danone)	Nestlé	Wyeth (now Nestlé)	Nutricia (Danone)	Farley (now Heinz)	Farley (now Heinz)	Heinz
Role of formula sponsor	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,	Supply of formula,

	financial assistance	financial assistance	financial assistance	financial assistance	financial assistance	financial assistance	financial and technical assistance, co-author	financial and technical assistance, co-author	financial assistance
Notes									
Will be included in main analysis	✔	✔	✔	✔	✔	✔	✖	✖	✖
Expected cognitive effect?	✔	✔	✔	✔	✔	✖	✔	✔	✖
Role in this thesis	Included in all analyses	Included in all analyses	Included in all analyses	Included in all analyses	Included in all analyses	Negative control for academic outcome analysis	Inform linkage algorithm	Inform linkage algorithm	Contribute data to MI process.
First publication	(Lucas et al., 2001)	(Morley et al., 2004)	(Fewtrell et al., 2002)	(Lucas et al., 1999)	(Morley et al., 1999)	(Kennedy et al., 1999)	(Lucas et al., 1989)	(Lucas et al., 1989)	(Kennedy et al., 1999)

Table 2.3: Characteristics of participants in the dormant trials as measured at randomisation (1/2)

	NEP-PD		**NETSGA**		**LCPUFAP**		**LCPUFAT**		**IRON**	
	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard
Randomised, n	113	116	152	147	96	100	155	154	162	165
Average birth weight, min max (grams)	1378 (775-2160)	1359 (630-2020)	2532 (1400-3160)	2602 (1770-3160)	1329 (640-1850)	1352 (740-1800)	3648 (2950-4900)	3540 (2680-4930)	3493 (2495-5103)	3465 (2466-4706)
Average gestational age, min max (weeks)	30.7 (26-36)	30.8 (25-36)	39.0 (37-42)	39.4 (37-42)	30.3 (24-36)	30.3 (25-36)	40.1 (37-42)	40.0 (37-42)	39.8 (36-43)	39.9 (35-43)
Mother's age (years)	28.2 (16-41)	28.5 (17-44)	26.8 (15-42)	26.4 (14-42)	26.1 (16-39)	26.7 (17-39)	27.5 (17-44)	27.0 (18-41)	27.7 (17-40)	27.5 (15-39)
Infant sex										
Male, n (%)	53 (47%)	57 (50%)	74 (49%)	68 (46%)	42 (44%)	53 (53%)	82 (53%)	83 (54%)	82 (51%)	81 (49%)
Female, n (%)	60 (53%)	58 (50%)	78 (51%)	79 (54%)	54 (56%)	47 (47%)	73 (47%)	71 (46%)	79 (49%)	84 (51%)
Mother smoked during pregnancy										
No, n (%)	67 (61%)	74 (68%)	79 (55%)	67 (50%)	55 (57%)	60 (60%)	117 (77%)	110 (74%)	116 (73%)	111 (69%)
Yes, n (%)	42 (39%)	36 (32%)	64 (45%)	66 (50%)	41 (43%)	40 (40%)	35 (23%)	39 (26%)	44 (27%)	51 (31%)
Missing	4	6	9	14	0	0	3	5	2	3
Mother has degree										
No, n (%)	106 (94%)	97 (88%)	143 (94%)	139 (96%)	52 (91%)	47 (90%)	140 (92%)	145 (96%)	140 (88%)	147 (90%)
Yes, n (%)	7 (6%)	13 (12%)	9 (6%)	6 (4%)	5 (9%)	5 (10%)	13 (8%)	6 (4%)	20 (12%)	16 (10%)
Missing	0	6	0	2	39	48	2	3	2	2

Table 2.4: Characteristics of participants in the dormant trials as measured at randomisation (2/2)

	PALM		**NEP-1**		**NEP-2**		**NUCLEO**	
	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard
Randomised, n	103	100	219	204	178	191	99	97
Average birth weight, min max (grams)	3575 (2640-4730)	3479 (2520-5400)	1389 (739-1844)	1392 (663-1847)	1419 (697-1847)	1398 (786-1842)	3455 (2210-4830)	3459 (2170-5360)
Average gestational age, min max (weeks)	40 (37-42)	39.9 (37-42)	30.9 (25-39)	31.0 (25-38)	31.1 (25-38)	31.1 (26-39)	39.4 (37-42)	39.2 (37-42)
Mother's age (years)	26.6 (15-40)	27.8 (17-42)	26.8 (15-42)	26.5 (16-42)	26.9 (15-42)	26.4 (15.8-40.1)	27 (16-44)	27 (16-40)
Infant sex			¶		¶			
Male, n (%)	66 (64%)	52 (52%)	112 (53%)	99 (51%)	88 (51%)	94 (51%)	60 (61%)	53 (55%)
Female, n (%)	37 (36%)	48 (48%)	100 (47%)	95 (49%)	86 (49%)	92 (49%)	39 (39%)	44 (45%)
Mother smoked during pregnancy								
No, n (%)	65 (63%)	74 (74%)	90 (56%)	88 (61%)	94 (75%)	90 (62%)	70 (71%)	58 (60%)
Yes, n (%)	38 (37%)	26 (26%)	72 (44%)	57 (39%)	31 (25%)	56 (38%)	28 (29%)	38 (40%)
Missing	0	0	57	59	53	45	1	1
Mother has degree								
No, n (%)	*	*	174 (85%)	167 (87%)	119 (79%)	138 (82%)	91 (92%)	86 (92%)
Yes, n (%)	*	*	32 (15%)	24 (13%)	31 (21%)	30 (18%)	8 (8%)	8 (8%)
Missing	76	70	13	13	28	23	0	3

** suppressed due to small cell sizes.*

¶ some participants in the NEP-1 and NEP-2 trials had missing infant sex and/or birth weight/ gestational age, presumably because they died shortly after randomisation, therefore, the number of males + females does not always add up to the total number randomised.

Table 2.5: Standardised mean differences between modified formula and control formula groups in previously collected cognitive endpoints and % lost to follow-up at each endpoint, by trial

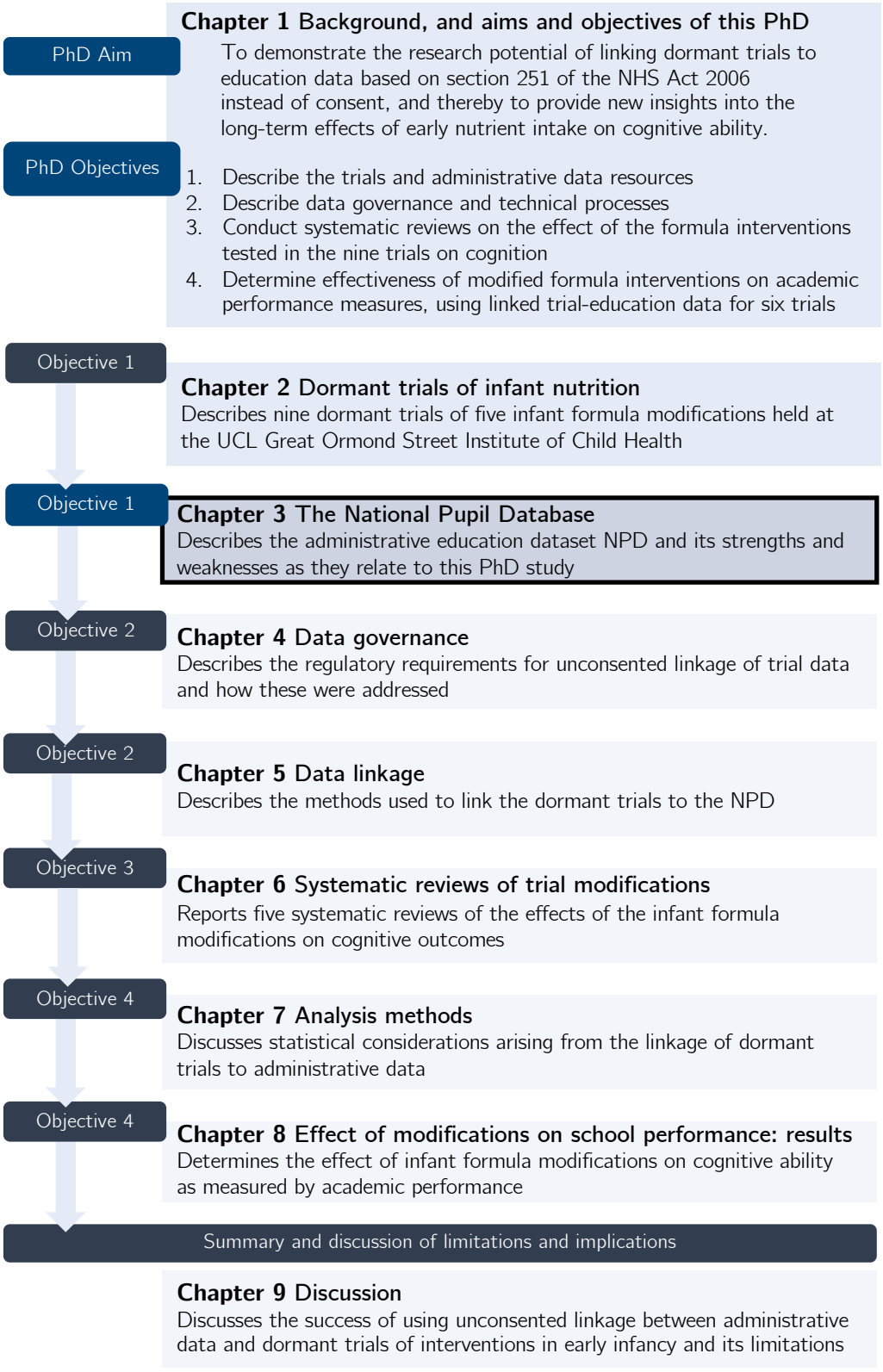
Endpoint	**NEP-PD**	**NETSGA**	**LCPUFAP**	**LCPUFAT**	**IRON**	**PALM**	**NEP-1**	**NEP-2**	**NUCLEO**
Bayley MDI (age 18 m)	0.09 SD lower (95%CI -0.36,0.18)	0.10 SD higher (95%CI -0.16,0.36)	0.08 SD higher (95%CI -0.27,0.43)	0.04 SD lower (95%CI -0.29,0.21)	0.05 SD lower (95%CI -0.29,0.19)	-	0.10 SD higher (95%CI -0.10,0.29)	0.09 SD higher (95%CI -0.13,0.30)	-
% lost to follow-up	**18.3%**	**27.4%**	**29.1%**	**24.3%**	**19%**	**-**	**4.7%**	**9.8%**	**-**
IQ measurement (age 5-8 y)	-	-	-	0.43 SD lower (95%CI -0.98,0.11)		-	0.05 SD higher (95%CI -0.15,0.25)	0.10 SD higher (95%CI -0.11,0.30)	-
% lost to follow-up	**-**	**-**	**-**	**40.5%**		**-**	**7.3%**	**3.8%**	**-**
IQ measurement (age 16-17 y)	-	0.18 SD higher (95%CI -0.56,0.93)	1.17 SD lower (95%CI -0.27,0.43)	0.41 SD lower (95%CI -0.71, -0.12)		-	0.39 SD higher (95%CI -0.17,0.96)	0.17 SD lower (95%CI -0.62,0.29)	-
% lost to follow-up		**90.6%**	**93.4%**	**87.7%**			**87%**	**85.6%**	

* **higher SD**= better cognitive outcome for modified formula group compared to standard group; **lower SD** = better cognitive outcome for standard formula group compared to modified group; difference is statistically significant at the 5% level if confidence interval (CI) does not include 0.

2.4 Key points from Chapter 2

- The nine dormant infant formula trials were chosen because they have the following characteristics in common: i) they tested infant formula modifications that are widely available and ii) they were all conducted in England and have retained personal identifiers of sufficient quality, enabling linkage to English administrative school data.
- The trials fulfilled different roles in this thesis. Cognitive effects were hypothesised and investigated for the NEP-PD, NETSGA, LCPUFAP, LCPUFAT and IRON trial. The other trials were negative control trials that were either not hypothesised to have a cognitive effect or were not expected to link in sufficient numbers to the school data. Their purpose was to allow the detection of suspected and unsuspected sources of error and bias in both the linkage processes and in the observed associations between modified formula and academic performance.
- All nine trials were parallel randomised controlled trials, investigating the superiority of one nutritionally modified infant formula over another, otherwise identical, standard formula, or banked breast milk.
- All trials suffered from high participant drop-out rates over time.
- The risk of bias from randomisation, allocation concealment, and blinding was judged to be low in the trials selected for cognitive analysis.

The next chapter, Chapter 3, introduces and describes the second data resource used in this thesis: The National Pupil Database (NPD).


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 3 Data source: The National Pupil Database

3.1 Chapter structure and content

In the previous chapter, I described the context and characteristics of nine dormant infant formula trials. This chapter seeks to address the second part of objective 1: *"to describe the trial and administrative data resources"*. I examine the National Pupil Database (NPD), the data source I used to extract the outcome data for my analyses. I argue why administrative data, particularly the NPD, is well-suited for tracing the cognitive development of children who have been part of a clinical trial in their infancy. I outline and justify which part of NPD data I requested for my analysis, provide background information on how these variables were collected and validated by the data controllers, and describe how I addressed resulting data quality issues. The closing section of this chapter discusses how the NPD and the trial data fit together.

3.2 Using administrative data for research

The NPD is an English administrative data resource that holds longitudinal school-level and pupil-level information and is curated by the UK government's Department for Education (DfE). As an administrative data resource, the NPD holds information primarily collected for administrative purposes: resource allocation, policy development, operational management, and statistics such as academic performance rankings (Department for Education, 2017b). While this implies certain methodological challenges, administrative research data also holds distinct advantages compared to follow-up methods involving primary data collection.

3.2.1 Administrative data vs traditional direct contact to retrieve outcome data

There are several possible ways to retrieve long-term data on cognitive ability for dormant trials. **Table 3.1** on the following page gives a head-to-head comparison of administrative data vs traditional direct participant contact. It

highlights that using administrative data to retrieve outcome data has several advantages over seeking direct contact.

Table 3.1: Routes to retrieve outcome data: comparison of administrative data vs direct participant contact

Aspect	**Administrative data**	**Direct participant contact**
Type of contact information needed	Identifiers coinciding with the period of data collection in admin data	Up-to-date information
Researcher time	Medium	High
Cost of data acquisition	Medium	High
Risk of accidental disclosure of sensitive participant outcome data	Low	High
Expected follow-up rate	Depends on legal basis for follow-up (high for unconsented)	Low
Risk of selection bias	Depends on legal basis for follow-up (low for unconsented)	High
Sources of information bias	Data entry errors, missing data	Social desirability bias, recall bias, missing data
Meta-data on data collection and validation procedures	Low transparency	High transparency
Complexity of process	High	Low

3.2.1.1 Problems with retrieving outcome data directly from participants

It is not feasible to retrieve outcome data directly from participants to ascertain the long-term effects of formula modifications on cognition. First, to contact former trial participants directly, up-to-date contact information is needed. However, this data is often inaccessible due to, for example, relocation or, especially in older preterm-born cohorts, early death. This method, therefore, also introduces the risk of potentially upsetting families of deceased children. To update participants' contact information, approval from an additional data

controller, such as the NHS Personal Demographics Service (PDS), would be needed. This would increase the complexity and timeframe of the project. The accuracy of contact data held by PDS also depends on the participants' engagement with health services. This might introduce selection bias. It could skew the sample towards those who tend to engage frequently with health services and their cognitive ability might differ systematically from those who cannot be reached. Secondly, the risk of accidental disclosure of participant information is higher with direct participant contact. This is because more data, including more sensitive up-to-date contact data, is processed on paper or via email, which can be breached, accidentally lost, or destroyed. This stands in contrast to the tightly regulated digital infrastructure required by UK data controllers and ethics committees when dealing with administrative datasets for research (described in more detail in Chapter 4, page 63).

Thirdly, when the follow-up for this PhD study was planned, the age of trial participants ranged from 16 to 35 years. Many participants and their families may not recollect being participants in the original study. Consequently, building up trust to enable re-engagement with researchers to share sensitive information would be difficult. This process would be costly, lengthy, and still likely to result in low response rates (Fewtrell et al., 2016). In addition, older participants might not remember their exam results or could be inclined to under-report bad exam grades.

Finally, costs of active follow-up, such as those generated by setting up appointments and sending testing material, are extremely high, as is researcher time required to engage with trial participants individually.

3.2.1.2 Limitations of using the NPD

The main limitation is that, as an administrative dataset, the NPD does not primarily collect data for research purposes. This means that careful attention needs to be applied during the data cleaning process, considering how and when the data was collected and what implications changes in data collection have for the research variables. Fortunately, the NPD has well-documented data collection and validation procedures, which I discuss in section 3.5 on page 57.

The complexity of information governance around administrative data is another limitation. Enabling linkage between datasets held by different institutions requires close communication with data holders and ethics committees and requires a legal basis to process the data without unnecessarily compromising the response rate. I describe this process in Chapter 4 from page 63.

3.3 NPD: structure and data contents

The NPD consists of several sub-datasets, referred to as *modules*. These modules started and stopped data collection in different years, with some being merged into each other. In-depth information about the NPD and its modules can be found in a data profile published by Jay and colleagues (2019) and the NPD data tables (Department for Education, 2020a).

Among the modules, and relevant for my research study, are the census modules and attainment modules. All pupil-level records are linkable across time and modules with the help of unique pupil identifier numbers.

3.3.1 Data sources

NPD data is compiled from several places. These include state schools, local authorities, and exam bodies. The data is entered into local systems as it is generated, for example, when exam results become available. The modules are collected and released at different time intervals. Information on the timings and data sources for each module is available publicly (Department for Education, 2017b).

3.3.2 Census modules

There are several census modules, which provide data at pupil-level. Relevant for this PhD project is that these modules contain variables that are used for linking individual pupil records to external data such as names, dates of birth, and postcodes. These were used to facilitate linkage to the trial data. In addition,

they detail the eligibility for special educational needs (SEN) support, which was one of the outcomes investigated in this thesis (**Table 3.3**, page 57). A full overview of census modules is available online (Department for Education, 2021).

3.3.3 Attainment modules

The NPD holds extremely rich data on academic attainment. Exams attended, levels of qualification achieved, and the actual marks are all stored in attainment modules. The modules are divided into the *Early Years Foundation Stage* (ages 3-5 years) and five different *Key Stages* (ages 5-18 years). The Key Stage 2 attainment module contains pupil-level information on the compulsory exams sat in the last year of primary school when pupils are around 11 years old. Key Stage 4 contains pupil-level data on the *General Certificate of Secondary Education* (GCSE), which all students sit around the age of 16 years. Mathematics and English language and literature are compulsory subjects. Pupils need to pass five or more *GCSEs* before they can move on to study *A-levels*, which are a requirement for students in the UK to enrol at university. In general, it is very rare in the UK for children to skip or be held back a year. This makes it possible to predict in which year the participants sat which exams (see **Fig. 3.1**).

3.4 Outcomes extracted from the NPD

NPD data for this research project was requested from the Fischer Family Trust (FFT) with permission from the UK's Department for Education. The FFT is a nongovernmental organisation that holds NPD data and analyses it on behalf of schools. This PhD project was the first project to directly retrieve data for research purposes through FFT rather than through the DfE (more information on rationale and data flows in Chapter 4, from page 64). Importantly, the FFT version of the NPD is not identical to that held by the DfE. It contains additional attainment score variables, adjusted for grade inflation, making it more suitable for analyses that span multiple academic years.

3.4.1.1 Choice of outcome variables

The primary outcome for this study was the mean difference in GCSE Maths exam standard deviation (SD) scores between modified formula and control formula groups within each trial. The calculation of the scores is discussed in Chapter 7 (from page 173). The choice of using GCSE Maths scores at age 16 years as my primary outcome was based on the following considerations:

i. The measure is available for a high number of trial participants, based on the years in which the trial participants passed through the academic trajectory and the years in which the NPD collated data (**Fig. 3.1**)
ii. GCSE Maths exams are compulsory, nationally administered and considered a 'high-stakes' exam, decreasing the risk of missing data.
iii. GCSE scores have higher predictive value for future employment and academic opportunities than other compulsory and nationally administered exam scores at Key Stage 2 (Hayward et al., 2014).
iv. Mathematics scores are better correlated with IQ outcomes compared to English scores (**Table 3.2** below).
v. Exam results for Mathematics are commonly considered to be less subjectively graded than English language and literature (Rhead and Black, 2018)

Table 3.2: Correlation of academic performance and IQ score in children from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort

	GCSE Mathematics grade at age 16 years	GCSE English grade at age 16 years
IQ at age 8 years	0.6294 (n=6,043)	0.566 (n=6,167)

Data supplied by Tim Morris (School of Geographical Sciences, University of Bristol, Bristol, UK) based on analyses for Morris T, Dorling D, Davey Smith G. How well can we predict educational outcomes? Examining the roles of cognitive ability and social position in academic performance. Contemp Soc Sci. 2016;11(2-3):154–168. doi:10.1080/21582041.2016.1138502

In addition to GCSE Mathematics exam SD-scores, the NPD was used to extract outcome data for several prespecified secondary and exploratory analyses. An overview of all outcome data extracted from the NPD is given in **Table 3.3** on the following page.

Table 3.3: Overview of outcome variables extracted from the NPD

Outcome	Variable	Format
Primary outcome	GCSE Mathematics score	A*=58 points to U=0 points, adjusted for grade inflation
Secondary outcome	GCSE English language score	A*=58 points to U=0 points, adjusted for grade inflation
Secondary outcome	KS2 Mathematics score	1 to 100 points, adjusted for grade inflation
Secondary outcome	KS2 English reading score	1 to 50 points, adjusted for grade inflation
Secondary outcome	5+ A*-C grades at GCSE	yes/no, derived
Secondary outcome	Ever received special educational needs support	yes/no, derived

3.4.1.2 Identifier variables used for linkage to trial data

The NPD holds several variables that can be used to link it to the infant formula trials. This includes the pupil reference number, which is used to link records from the same pupil across the NPD, as well as pupil name, date of birth and pupils' home postcodes during the period of data collection. These are discussed in more detail in Chapter 5 on page 80.

3.5 NPD: Coverage and data validation procedures

3.5.1 General coverage

Data on pupil numbers and attainment have been submitted systematically to the DfE and incorporated into the NPD since 1996. In England, it is a legal requirement that all children receive education and that state schools, local authorities, and awarding bodies report data on pupil numbers and attainment to the DfE. While most children in England aged 5-16 years receive their education in state schools, the NPD does not typically contain data on periods where children are enrolled in privately-funded schools or schooled at home (Department for Education, 2017a). It does, however, receive data from exam boards for the compulsory GCSEs, meaning that this data is included in the NPD even if participants never interacted with state-funded schools. In any given

academic year, data from over 99% of all children of compulsory school age are included in the NPD database (Jay et al., 2019).

3.5.2 Data validation

All data is subject to manual and automatic quality control checks upon submission from the source organisations. These include automatic de-duplication procedures, the use of visual tools such as scatterplots to identify any outliers in year-on-year trends and to inspect the data for improbable scenarios such as a school having no children with special educational needs. A summary of validation rules that are applied by the DfE each year are available online (Department for Education, 2020b).

3.5.2.1 Specific data quality considerations

Several data quality considerations are relevant for this PhD study; these are discussed in the table below.

Table 3.4: Implications of data quality issues and proposed solutions

Issue	Potential implication	Solution
Grade inflation	Can affect year-on-year comparisons between modified formula and standard formula groups	Use inflation adjusted variables created by FFT
GCSE attainment scales for exams have changed over time	Can affect year-on-year comparisons between modified formula and standard formula groups	Harmonise scales
Participants who are expected to link do not link to the NPD	Can result in participants being excluded from analysis	Impute exam results and do sensitivity comparing against complete case analyses
Missing data within pupil records	Depending on the variable that is missing this could result in participants being excluded from analysis	Impute missing data and do sensitivity analyses comparing against complete case analyses

3.6 Overlap of NPD data with trial data

In order to gain an overview of the outcome data that is available for each trial, it is instructive to illustrate where trial data overlap with NPD data. As described above, some NPD modules started or stopped collecting data over the years, resulting in some outcomes being unavailable depending on the age of the trial participants in that calendar year. **Fig. 3.1** on the next page shows the birth year of trial participants on the left-hand side, and stretching out to the right is the corresponding participant age by school year (grey numbers inside cells). The area where the cells are shaded dark blue shows the school year at which the Key Stage modules (top of the figure) collected outcomes. This explains why some participants in the correct age group for a module (NEP-1 and NEP-2 participants at age 16 years) were not expected to have any outcome data. The maximum number of participants with outcomes from the Key Stage modules of interest is shown at the bottom right. In practice, this number will be smaller due to factors such as death, emigration, or never-interaction with schools or exam boards collecting attainment data for the NPD. I estimated these factors to account for the loss of about 3-5% of participants, depending on the trial (appendix p 14).

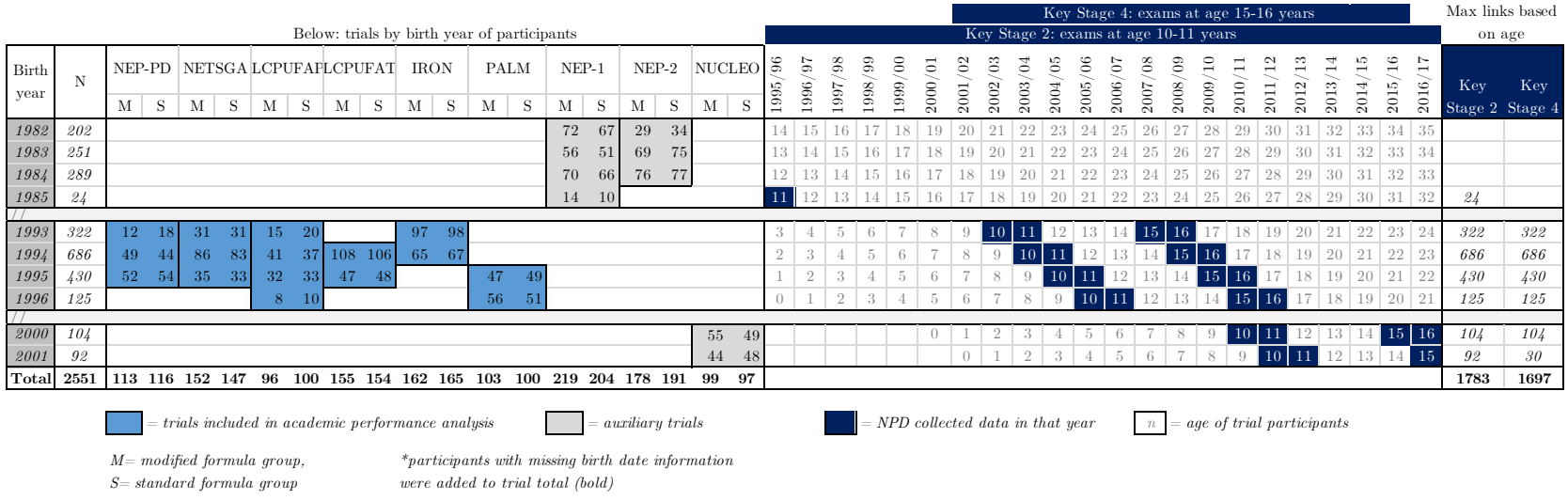
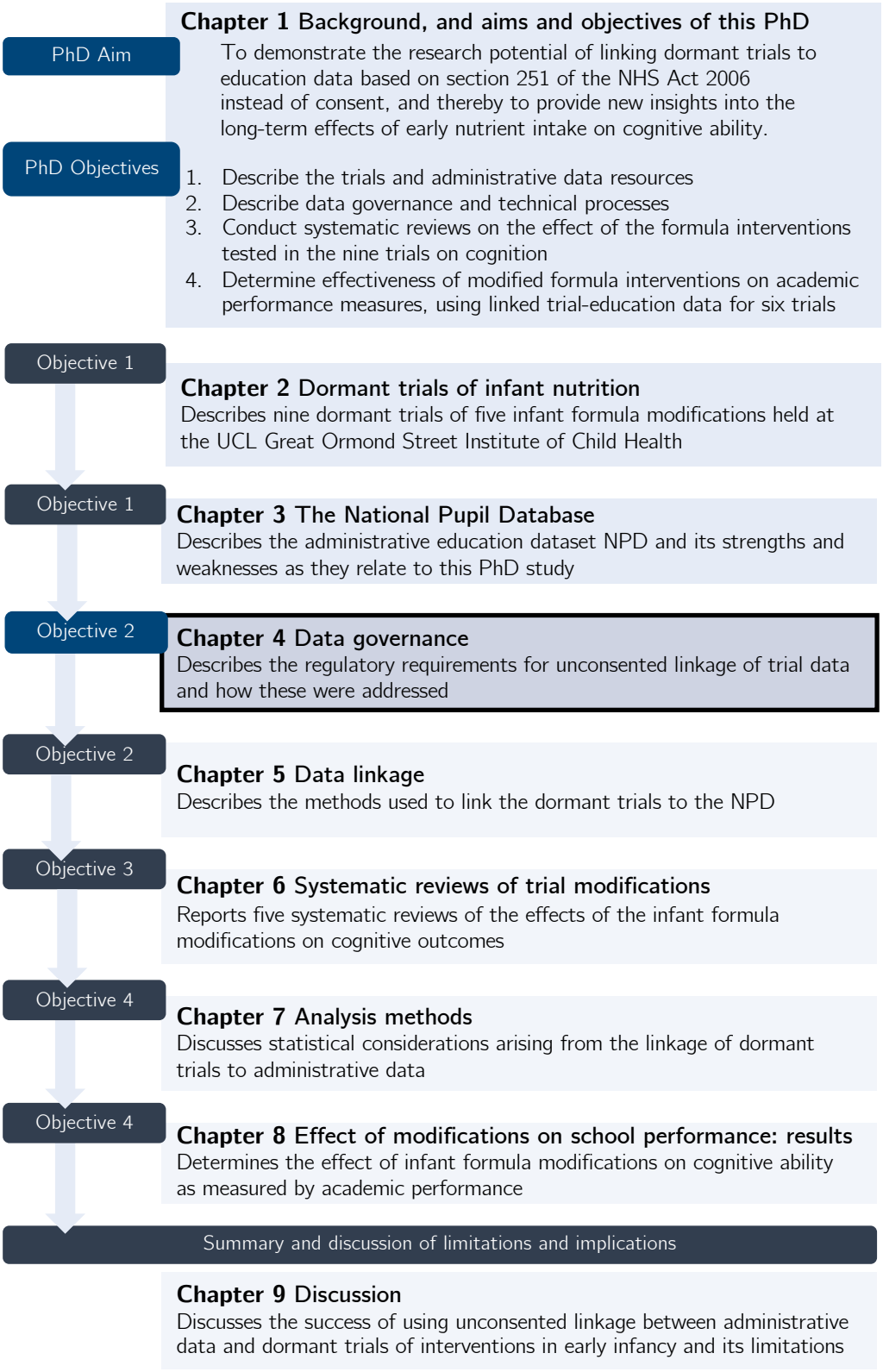


Fig. 3.1: The maximum number of participants who can be linked to exam results at each Key Stage depends on whether the module collected data in the academic year in which participants sat the exam.

3.7 Key points from Chapter 3

- The NPD is an administrative dataset; as such, it contains information collected primarily for administrative (not research) purposes.
- The NPD consists of several modules with different data collection periods. Among these are the KS2 and KS4 (GCSE) modules, which contain data on nationally administered compulsory exams sat at age 11 and age 16 years, respectively.
- I will use the KS2 and KS4 modules to ascertain cognitive outcomes as measured by academic performance.
- My primary outcome is the grade for the GCSE Maths exam, sat at age 16 years.
- For this PhD study, key benefits of ascertaining outcomes through unconsented linkage to the NPD include that the NPD data provide near 100% coverage of the UK pupil population, especially for GCSE grades which are submitted by the central exam bodies, not the schools themselves; that blinding of outcome assessors is maintained; and that the NPD is likely to capture participants who were lost-to-follow-up in previous surveys.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 4 Data governance

4.1 Chapter structure and content

The previous two chapters have described the two data resources used for this thesis – the dormant trials and the NPD. This chapter presents work towards objective 2: "*to describe data governance and technical processes*". I outline the information governance requirements for unconsented linkage between dormant trial and NPD data and how these were addressed in this study. The chapter covers four themes: (1) information governance requirements; (2) data flows; (3) NPD access and approvals; and (4) project timeline and costs.

4.2 Information governance requirements

Information governance describes the way in which organisations process information. It encompasses data collection, data security arrangements, data use, data sharing, data archiving, and the destruction of data. The concept of personal data is central to the governance of linking trials to administrative data. The UK Data Protection Act defines personal data as follows:

Data which relate to a living individual, who can be identified:

(a) From those data, or

(b) From those data and other information, which is in the possession of, or is likely to come into the possession of, the data controller.

In this PhD study, personal data refers to identification data, such as names and addresses, but also to the participant data that was collected in the trials (e.g., birth weight) and pupil data that was collected in the censuses, which in combination might be identifying.

The UK has comprehensive and heavily enforced legislation in place to protect an individual's right to data privacy (discussed below), and researchers, as well as data-holding organisations, need to consider whether data sharing is accordance with those legislations.

Where personal data from multiple data providers – including outside the health field – is combined, such as in this PhD study, providing evidence of good

research practice can be complex because there is no common independent auditor checking compliance with provider-specific data governance rules. The proposal to link the trials to the NPD without consent was therefore reviewed in a robust and lengthy process involving local and national committees and both the DfE and NHS institutions, which scrutinised legal, ethical, and technical aspects of the study (**Fig. 4.1**).

Linkage requires the use of identifiers, which constitute personal and confidential data. To use these identifiers without consent, and as a pre-requisite for research ethics approval, I sought exemption from section 251 NHS Act 2006 from the Confidentiality Advisory Group (CAG) (the next section details this process). Given approval by CAG, Research Ethics approval was granted by the London City & East Research Ethics Committee and the DfE's Data Management Advisory Panel. These applications also required evidence of compliance with all principles of the Data Protection Act (discussed in section 4.2.2, page 67) as well as evidence of public acceptability of the research (discussed in section 4.2.3, page 69).

4.2.1 CAG and section 251 of the NHS Act 2006 as legal basis

There are several legal pathways that can be used as the basis for processing personal and confidential patient data obtained through health care for research. Informed consent is one of those pathways, section 251 support another. I sought support for processing the data under section 251 of the National Health Service Act 2006 from CAG, which was granted.

Several criteria needed to be fulfilled to achieve CAG approval. First was that the research project must demonstrate substantial expected medical benefits to the public. I argued that there is likely substantial expected medical public benefit from the research conducted in this PhD study: there is evidence of cognitive harms emerging in older age groups from certain infant formula modifications (Lanigan and Singhal, 2009, Lozoff et al., 2012, Makrides et al., 2010) despite that, some of these modifications have recently been mandated EU-wide (EU Commission, 2016). The dormant RCTs in this PhD study offer a unique data resource to investigate potential long-term harms and benefits. They

include the first trial of a nutritional intervention in infancy designed to test efficacy and safety and collectively represent the largest number of participants randomised to modified infant formulas. Their diverse designs and involvement of different types of nutrients in different subgroups (e.g., preterm, term, and babies born small for gestational age) informs the generalisability of results as well as methods for using linked trial-administrative data. Findings generated from this PhD also have the potential to be relevant for infant nutritional practices beyond the UK.

The second condition for CAG approval was evidence that consent-based follow-up would not be sufficient to answer the research question at hand. I argued that the method used in this PhD study could produce valid and important results where previous consented follow-up could demonstrably not. This was evidenced by the low participant retention in previous consent-based follow-ups. To maximise follow-up, the infant formula trials used in this PhD had previously involved dedicated paediatric investigators known to the participants and their parents. In addition, contact for follow-up was preceded by checking addresses through the NHS tracing service to ensure that addresses were up to date. Despite these efforts, response rates for long-term outcomes had fallen universally below 15% by age 17 years (see **Table 2.5**, page 47). This caused two problems: a major loss of statistical power and a biased subgroup of more healthy participants (Fewtrell et al., 2016, Fewtrell et al., 2008), which could not be used to validly answer questions about effects on cognitive ability.

Finally, based on six consultations of trial participants I conducted for this PhD study (section 4.2.3, page 69), I believe that it is plausible that the trial participants would support maximising the value of the data they have already contributed to the trials. My consultations indicated that participants were reassured that any potential risks to their privacy were mitigated by the study design and the security and governance measures in place at the various data processors. Actively tracing all participants to obtain consent would impose an even greater intrusion on their privacy and place an additional burden in terms of time and effort on the participants and on public resources for research funding.

4.2.2 Data Protection Act 1998

To be granted ethical approval, I had to outline how this study will comply with the principles of the Data Protection Act 1998[§]: (1) Fair processing; (2) Used for specified purposes; (3) Minimum necessary for the purpose; (4) Accuracy; (5) Kept for minimum time necessary; (6) In accordance with the rights of the data subject; (7) Security and confidentiality protection; (8) Not disclosed outside the European Union.

Below, I discuss principles **(1)**, **(5)**, and **(7),** which required particular consideration:

(1) Fair processing requires that certain information about how the data is processed is made available to participants. This includes but is not limited to the identity of the data controller (in this study, the UCL GOS ICH), specific reference to the purpose and legal basis for processing, as well as data security measures put in place. Fair processing notices are challenging in situations where there is no obvious forum for participant contact. In this PhD study, fair processing was addressed by setting up a participant facing trial webpage on the UCL GOS ICH website, which is the institution where the last participant contact has taken place. The website also links to one of the principal investigators who was last in contact with the participants. The website states the purpose and scope of the follow-up study and provides participants with a contact for any questions and the option to dissent.

(5) Navigating compliance with data retention policies between the organisations involved in the project also required special consideration. According to the Data Protection Act 1998, any research data is expected to be destroyed at the end of a project. This project was anticipated to last five years.

[§] I applied for ethics approval in 2017. In 2018 the Data Protection Act 1998 became the UK's implementation of the General Data Protection Regulation (GDPR) leading to several changes in the regulation including more stringent fair processing / privacy notices. This was anticipated and integrated in the ethics application for this PhD study. As of 01.01.2021 this legislation is superseded by the so-called *UK GDPR* laid out in the European Union (Future Relationship) Bill. No changes therein have an impact on this PhD study.

However, UCL requires research data to be archived for a minimum of 15 years after publication for audit purposes. CAG granted permission to hold the digitally recorded personal identifiers (see section 5.3.2, page 87) for 15 years to link them to health data in the future. If access to the identifiers beyond the purposes of archiving is required at a later stage, then new ethics applications and data sharing agreements will have to be drawn up.

(7) Providing assurance that all data processors have appropriate data security arrangements in place proved difficult. This challenge was addressed by asking the Fischer Family Trust (FFT), the organisation that conducted the initial part of the linkage (section 4.4 page 72 for data flows) and is usually not involved in health research, to create a detailed data-sharing agreement with UCL (appendix p 35) and to undertake an IG toolkit assessment (now called Data Security and Protection toolkit). The toolkit is an online tool issued by the Department of Health and requires an annual published self-assessment of compliance against the National Data Guardian's ten data security standards.

Table 4.1: Section 251, subsections 1, 4, 10 and 12 of the National Health Service Act 2006

(1) The Secretary of State may by regulations make such provision for and in connection with requiring or regulating the processing of prescribed patient information for medical purposes as he considers necessary or expedient—

(a) in the interests of improving patient care, or

(b) in the public interest.

(4) Regulations under subsection (1) may not make provision requiring the processing of confidential patient information for any purpose if it would be reasonably practicable to achieve that purpose otherwise than pursuant to such regulations, having regard to the cost of and the technology available for achieving that purpose.

(10) In this section "patient information" means—

(a) information (however recorded) which relates to the physical or mental health or condition of an individual, to the diagnosis of his condition or to his care or treatment, and

(b) information (however recorded) which is to any extent derived, directly, or indirectly, from such information, whether or not the identity of the individual in question is ascertainable from the information.

(12) In this section "medical purposes" means the purposes of any of—

(a) preventative medicine, medical diagnosis, medical research, the provision of care and treatment and the management of health and social care services, and

> *(b) informing individuals about their physical or mental health or condition, the diagnosis of their condition or their care and treatment.*

4.2.3 Acceptability: stakeholder views on the study design

As a pre-requisite to obtain CAG and ethics approval, it was necessary to demonstrate that I sought the views of the trial participants and that my research is understandable and acceptable to the trial participants. I had the opportunity to consult six participants from the NEP-1 and NEP-2 trial, now aged 30+ years, who were invited back to UCL GOS ICH to participate in a brain scan study. I met them separately (two in person and four over the phone) and started the conversation with two broad questions:

1. "At the institute, we are thinking of ways to make most of the data that has already been collected, such as using information from old trials for other research projects. How do you feel about that and why?"
2. "How would you feel if we tried to make most of the data by linking it to health or education records?"

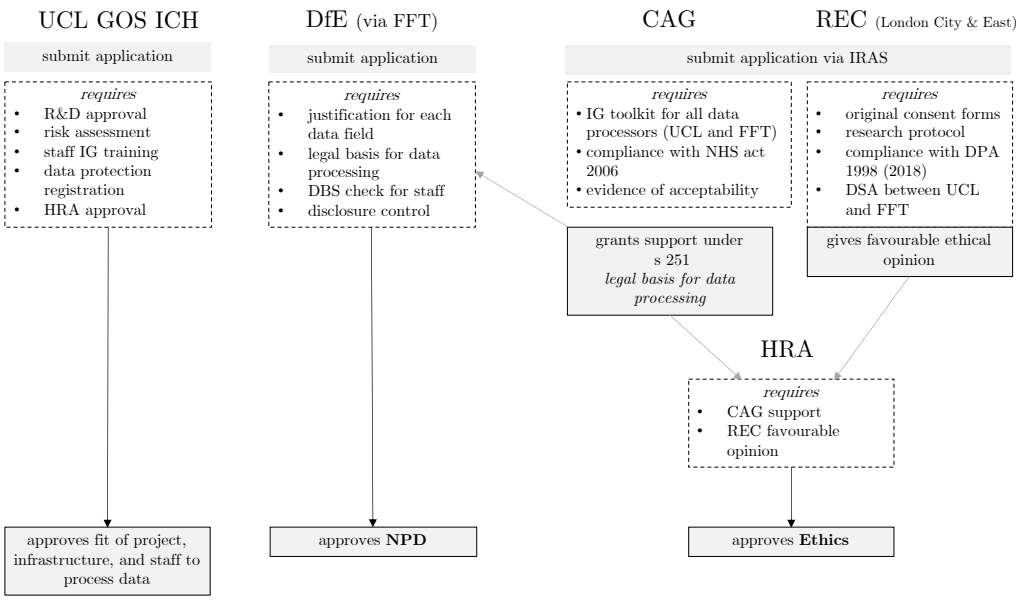
If prompted, I clarified these questions and further explained the outline of the planned study. The stakeholders voiced a high level of support for our approach. Among the reasons for support, they stated that they expected findings to be useful to advance patient care. The stakeholders also supported adding value to their own previous inputs through data linkage and acknowledged the difficulty to trace participants after several years. I felt it was also generally understood that participant data would be analysed in de-identified, digital form. It was stated repeatedly that their support for research is not limited to a specific research question; rather, that their support is towards research for the benefit of patients in general. However, almost all stakeholders said independently from each other that they would make a distinction between data used for academic research and that used for for-profit research.

While I recognise that I consulted only a small number of participants, who were arguably biased towards a positive attitude to research, they were the

only available representatives of the original participants. These consultations were important for two reasons: (1) they provided some degree of confidence to me, to CAG, and to the ethics committee that the methods of my study were acceptable to the participants, (2) and they highlighted the importance of emphasising the broad research benefits of my study design (beyond specific research questions) in lay material.

Table 4.2: Quotes from the stakeholder consultations

"I want health professionals to look at all the data that has accumulated on me to identify any patterns that can help other people. Honestly researchers should use as much as they have because I want to help you guys and you know... future patients. Anything that helps people on the long run. I do not think that when I was born my mother knew that I will be born preterm – I think much has happened since then. Research has an impact. People can now identify signs for that and do something."
"Things can go missing in paperwork, probably much easier than when data is on the computer. I think there is still a heavy reliance on paperwork. For example, every time I visit my Cardio nurse she is sitting in an office surrounded by hand-written stuff."



UCL GOS ICH *UCL Great Ormond Street Institute of Child Health*; CAG *Confidentiality Advisory Group*; DBS *Disclosure and Barring Service*; DfE *Department for Education*; DPA *Data Protection Act*; DSA *Data sharing agreement*; FFT *Fischer Family Trust*; HRA *Health Research Authority*; IG *Information Governance*; IRAS *Integrated Research Application System*; NPD *National Pupil Database*; R&D *Research and Development*; REC *Research Ethics Committee*; NHS *National Health System*; s 251 *Section 251 of the NHS Act 2006*;

Fig. 4.1: Information Governance requirements and data application processes

4.3 NPD: access and approvals

Sharing of pupil-level data between the Department for Education and external organisations is subject to strict approval procedures. Procedures and justifications were laid out in data-sharing agreements between the DfE, FFT, and the UCL Great Ormond Street Institute of Child Health (UCL GOS ICH) (appendix p 35). Requirements include that suitable data security arrangements (Information Governance Toolkit – an information governance tool used by the Department of Health) at all data processing institutions are in place and that each person accessing the NPD data must undergo an official criminal record check through the *Disclosure and Barring Service* (DBS).

4.4 Data flows

4.4.1 The trusted third-party mechanism

The laws surrounding the processing of personal data require mechanisms to mitigate any risks of data disclosure and to maintain functional anonymity of the data. On their own, identification data (that is, names, addresses, dates of birth etc.) hold no information on a person beyond the fact that this person is present in a dataset. Identification data should therefore be handled separately from other personal data, in this case, the linked RCT-NPD data, to minimise the risk of disclosure of personal data about an individual (termed the separation principle). One widely used mechanism used when sharing data between different organisations is the use of trusted third parties (TTP). A TTP allows the separation of certain data flows to ensure that no party is able to see more personal information than they already hold (Harron et al., 2015). It can ensure that parties never have access to participant identification data (e.g., names and postcodes) and the linked data (e.g., trial data and test scores) at the same time, which is a key privacy concern. A TTP carries out the linkage using only pseudonymised unique identification numbers (that have no real-world meaning) and the identification data (e.g., names and postcodes). The ideal process of involving a TTP in data separation is illustrated in **Fig. 4.2** on the next page.

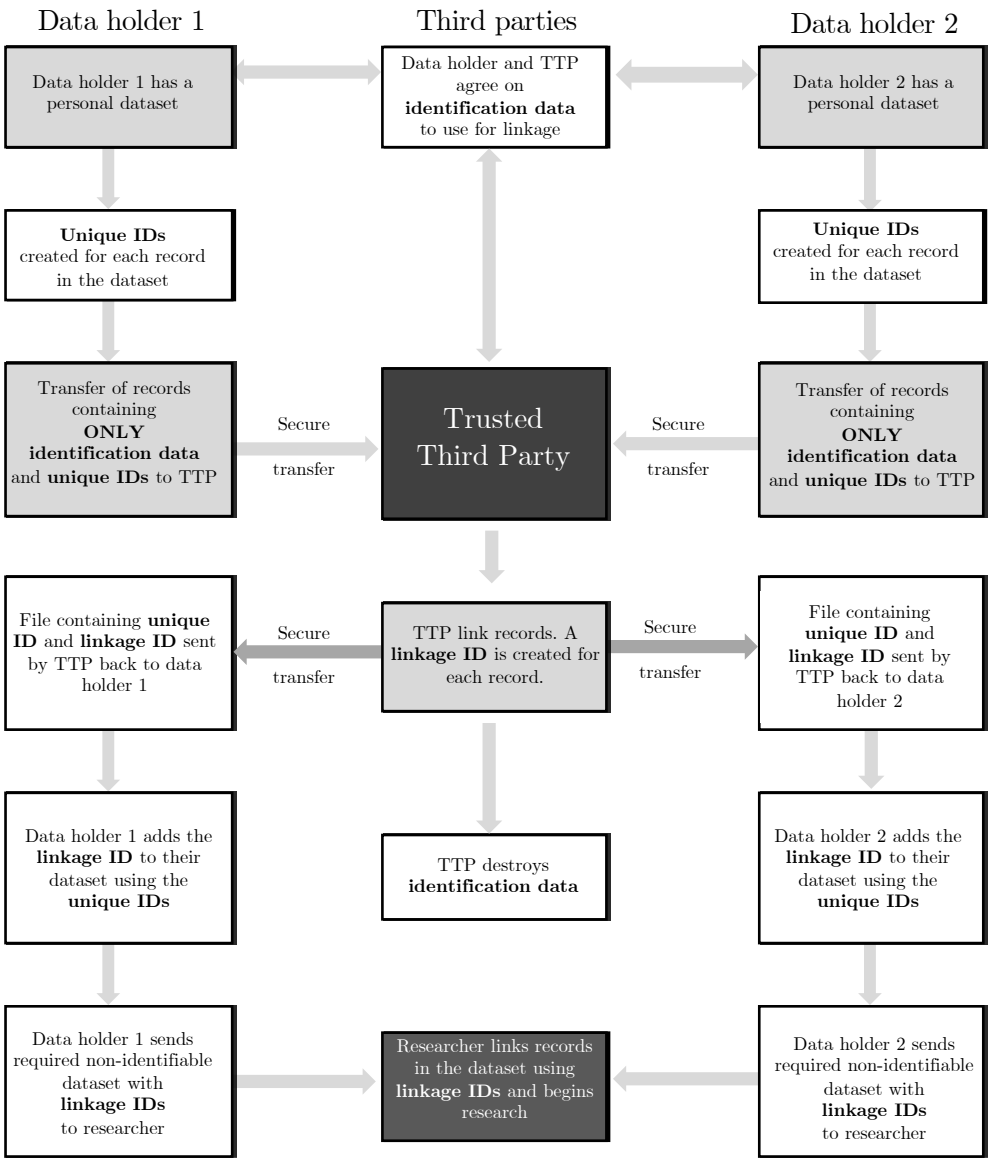
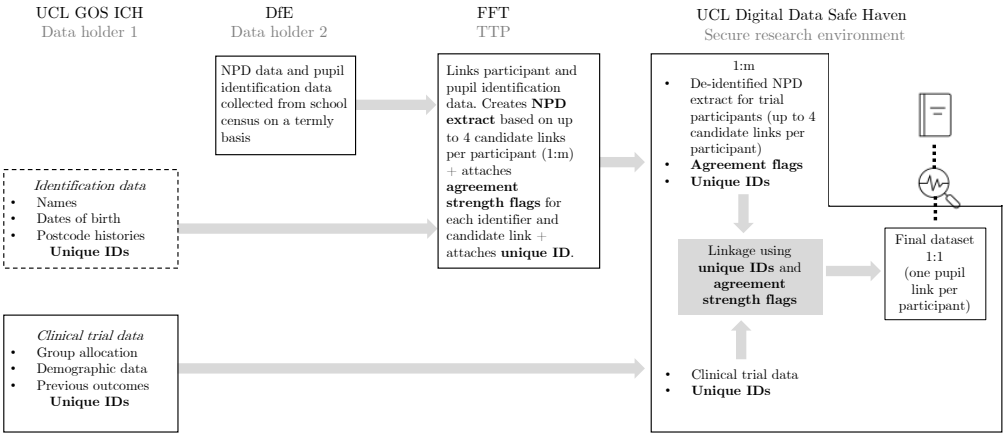


Fig. 4.2: Ideal-world data flows when using a trusted third party to minimise disclosure of personal data. Figure reproduced from Dibben, C., Elliot, M., Gowans, H., Lightfoot, D. and (2015). The data linkage environment. In Methodological Developments in Data Linkage (eds K. Harron, H. Goldstein and C. Dibben). with permission from John Wiley and Sons) License Number: 5004271126622, License date: Feb 08, 2021

In practice, complete separation of data flows is not always achievable, and the TTP can sometimes hold a dual role by holding personal data extracts as well as conducting the linkage, which involves having access to parts of the attribute data. In this PhD study, the appointed TTP was the Fischer Family Trust, a non-profit organisation that usually creates bespoke NPD extracts and analyses for English schools and therefore holds NPD data with permission from the Department for Education.

4.4.2 Description of data flows to link dormant trials and NPD



UCL GOS ICH *UCL Great Ormond Street Institute of Child Health*; DfE *Department for Education*; FFT *Fischer Family Trust*; NPD *National Pupil Database*; TTP *Trusted Third Party;* 1:m *One (participant record) to many (pupil candidate records);* 1:1 *One (participant record) to one (pupil record)*

Fig. 4.3: Data flows in this PhD study

Fig. 4.3 above shows the data flows for this PhD study. There were three organisations involved in the linkage: UCL, the DfE and the FFT. At UCL, I first attached unique participant IDs to both the participant identification data and the clinical trial data. Then, the unique participant IDs + identification data (names, addresses etc.) were securely transferred to the FFT (without the clinical data). At the FFT, participant identification data was used to identify matching pupils. This was done in a series of steps, involving comparisons of first name(s), last name(s), date of birth, and postcode between participant and pupil identifying data. The goal of data linkage is to identify true matches. However,

even in true matches, not all identifiers always agree exactly, for example because of different name spellings. The matching process therefore generated multiple potential candidate links. Beforehand, I agreed with the FFT to be supplied with up to four of these potential candidate links per participant. It was also agreed that I would be supplied with information on the strength of agreement for each identifier (e.g., participant first name matches pupil first name exactly or participant first name truncated at any hyphen matches pupil first name, see **Table 5.3** on page 95 for an overview of all possible agreement levels). FFT then extracted and attached the exam score data for all candidate links. The data was then de-identified and, together with the unique participant IDs and information on the strength of agreement for each identifier, securely transferred to the UCL data safe haven (an IG toolkit approved digital research environment). The de-identified clinical trial data files from the principal trial investigators were also uploaded to the UCL data safe haven. In the data safe haven, I completed the linkage of trial data with the NPD extract so that the final dataset contained only the best matching participant-pupil pairs (see detailed discussion of linkage in CHAPTER 5). All data analyses were also conducted in the UCL data safe haven. Any results, such as tables, figures and Stata scripts, downloaded from the UCL data safe haven were documented and checked for potentially identifying data and small cell sizes (which can be indirectly identifying (Office for National Statistics, 2016)) by an independent data manager.

4.5 Time frame and costs

The project benefited hugely from my research group's expertise in using linked administrative data for research. I received guidance on the process and the wording of the applications to maximise the understanding required by the review panels. Still, unconsented linkage of dormant trials to administrative data has no precedent and was therefore a proof-of-concept study. Hence, the preparation of data and ethics approvals needed significant time and other resources. **Table 4.3** on page 77 shows that digitisation of participant identifiers (see Chapter 5) took about one year and three months, acquisition and cleaning

of the clinical data files spanned a period of ten months, the NPD approval took nine months, and ethics support for this study was granted after one year and 11 months.

The FFT charged £14,160 for the bespoke NPD extract, and the digitisation of identifiers required two data entry assistants and therefore incurred £37,966 in additional staff costs, resulting in total costs of £52,126. While the digitisation included a further 1,500 participants from other trials not included in this PhD study, the linkage cost at FFT was independent of the number of participants linked. Therefore, if all 4,051 participants had been sent for linkage, this study would still have incurred costs of approx. £13 per participant – a fraction of the costs reported in Llewellyn-Bennett et al. (2018). Here, costs were reported to be $2,126 (approx. £1,500) per participant for a four-year follow-up through a combination of telephone and face-to-face appointments, and $500 (approx. £355) per participant for a one-year follow-up through postal correspondence, telephone, and medical records.

Table 4.3: Timeline of data preparation, acquisition, and approvals

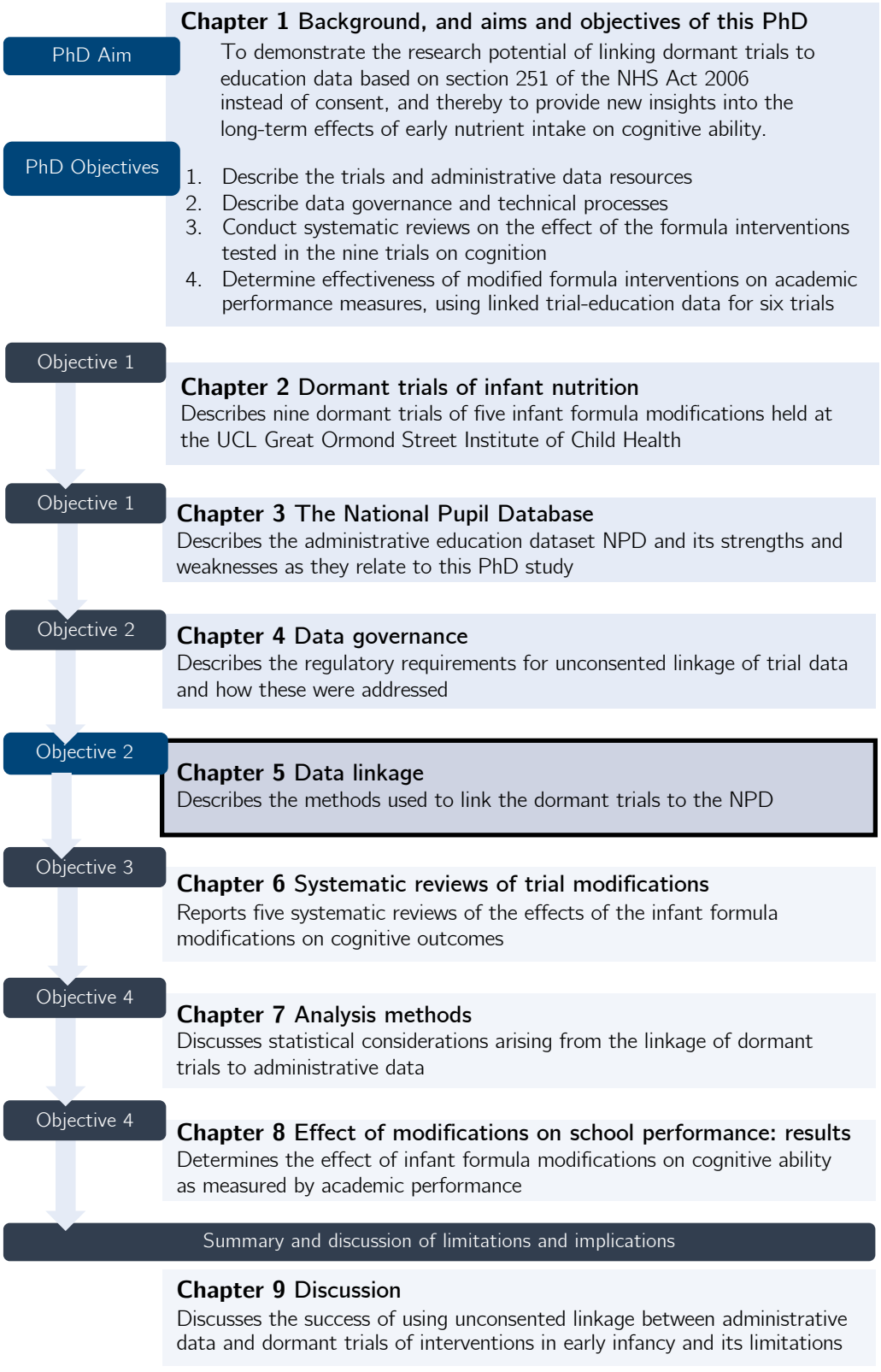
Event	Actors	Date
Database set up in data safe haven to digitise personal identifiers	MV	Sep. 26, 2016
Consultations with former trial participants	MV, MF, Participants	Oct. 24, 2016
Clinical data cleaning	MV	Nov. 21, 2016 to Oct. 13, 2017 (10 months)
UCL ethics approval	MV	Dec. 02, 2016
Data entry	MV, KK, NP, MR	Dec. 02, 2016 to Mar. 9, 2018 (15 months 1 week)
UCL study website and privacy notice online	MV	Dec. 08, 2016
REC and CAG forms submitted	MV, RG	Mar. 15, 2017
London - City & East Research Ethics Committee approval	REC, MV, RG	Apr. 22, 2017
NPD application submitted	MV, RG, LD	Jul. 04, 2017
PhD funding period starts	MV, ESRC	Oct. 01, 2017
FFT agree to collaborate, DSA signed	FFT, MV, RG	Feb. 02, 2018
DfE approve linkage by FFT	DfE	Apr. 04, 2018
CAG approval	CAG	Oct. 12, 2018
HRA approval	HRA	Feb. 11, 2019 (22 months, 3 weeks)
NPD data extract received	FFT	Feb. 28, 2019
Linked data cleaning	MV	Feb. 28, 2019 to Jan. 31, 2020 (11 months)
Statistical analysis plan submitted for publication	MV, KH, MF, JJ, RG	May 1, 2020
Data analyst unblinded	MV	May 21, 2020

CAG *Confidentiality Advisory Group*; DfE *Department for Education*; DSA *Data Sharing Agreement*; ESRC *Economic and Social Research Council*; FFT *Fischer Family Trust*; HRA *Health Research Authority*; JJ *Prof John Jerrim*; KH *Dr Katie Harron*; KK *Kathy Kennedy*; LD *Prof Lorraine Dearden*; MF *Prof Mary Fewtrell*; MR *Marina Ruiz*; MV *Maximiliane Verfürden*; NPD *National Pupil Database*; REC *Research Ethics Committee*; RG *Prof Ruth Gilbert*.

4.6 Key points from Chapter 4

- In the UK, the right to data privacy for individuals is tightly regulated. Researchers seeking to use personal data and organisations that hold personal data need to comply with the relevant legislation when sharing data. Relevant legislation for this PhD study includes the NHS Act 2006 and the Data Protection Act 1998 (now UK GDPR).
- Instead of consent, the legal basis for processing of personal data for this PhD study was support from the Confidentiality Advisory Group under section 251 of the NHS Act 2006.
- Financial resources invested in this project amounted to £52,126, significantly lower than active follow-up approaches typically cost.
- Several mechanisms were used to bring the overall risk of personal data disclosure to an acceptably small level. These include the drafting of detailed data-sharing contracts, background checking of all individuals involved in the linkage, and the separation of data flows to ensure that no party can see more personal information than they already hold.

The next chapter will present the remaining work towards objective 2 by outlining the processes of participant identifier digitisation and data linkage between dormant trial and NPD data.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 5 Linkage of trial data to NPD data

5.1 Chapter structure and content

In the preceding chapters, I described the data resources for this thesis: the trial data (Chapter 2) and the NPD data (Chapter 3), as well as the information governance requirements for processing these data resources (Chapter 4). This chapter outlines in detail the steps required to prepare and link the data resources together, addressing objective 2: "*describe data governance and technical processes*". The first section of this chapter discusses the importance of linkage quality and describes how linkage error can be evaluated. The second section of this chapter gives a step-by-step overview of the linkage strategy that was used in this thesis, including the digitisation of participant identification data. In the closing section, I describe the linked dataset as it will be used in Chapter 7 and Chapter 8.

5.2 The importance of linkage quality

Before describing the linkage strategy applied in this thesis, it is necessary to first understand the importance of linkage error and describe how it can be evaluated.

5.2.1 Implications of linkage error

Linkage accuracy (the absence of linkage error) is a pre-requisite for producing reliable analysis results for the dormant infant formula intervention trials. The two signs of linkage error, false links and missed links, can manifest in analyses as missing data, unrepresentative sampling, misclassification, and measurement error (Harron et al., 2017a, Harron et al., 2014, Harron et al., 2012, Harron et al., 2017b, Doidge and Harron, 2019).

5.2.1.1 Missed links and false links

When linking trial participants to the NPD, *missed links* occur in trial participants who have an NPD record but whose trial and NPD records are not linked. In practice, I cannot distinguish between cases 1) where there is no link, but there should be a link, and the cases 2) where there is no link because the participant did not survive, emigrated, or never interacted with schools or exam boards that submit data to the NPD. However, the proportion of missed links in category 2) should be small (appendix p 14). Missed links are essentially missing data, leading to a loss of statistical power. They can also introduce bias by making the linked sample unrepresentative of the originally randomised sample if certain groups are more likely to be missed than others (Harron et al., 2014, Harron et al., 2012).

False links are the result of participants being erroneously linked to NPD records that are not their own, for example, because the quality of identifiers is low (**Table 5.1**). False links can lead to measurement error (information bias) (Hagger-Johnson et al., 2015) and loss of power through introducing erroneous variation. An example of a false link would be if a participant, who in reality has

average grades, would be erroneously linked to a pupil record of a child with above-average grades.

In this PhD study, every party involved in the linkage was blinded to whether the participant was part of the control or intervention group. In addition, proportions of missing identifier data (names, addresses etc.) were low and distributed fairly equally across modified formula and standard formula groups (**Fig. 5.8**, page 94). I, therefore, expected linkage error to occur at random and at roughly the same rate across randomised groups. As a consequence, I expect missed links and false links to bias intervention effect estimates towards the null (Rothman et al., 2008).

Table 5.1: Possible scenarios following linkage of two records between RCT and NPD data

	Same child	Different children
Records link	True link	False link *(Information bias and loss of power)*
Records do not link	Missed link *(Selection bias and loss of power)*	True non-link

5.2.1.2 Linkage structure

The term linkage structure refers to the relationship between the analysis sampling frame and the datasets that are being linked; it determines who is included in the analysis. For any two sets of data, Doidge and Harron (2019) have identified eleven possible linkage structures, defined by the possible ways two datasets can intersect to create an analysis sampling frame. Out of these, the combined trial and NPD data can either form an '*imperfect nest*' or an '*intersection*'. **Fig. 5.1** illustrates these two types of linkage structures:

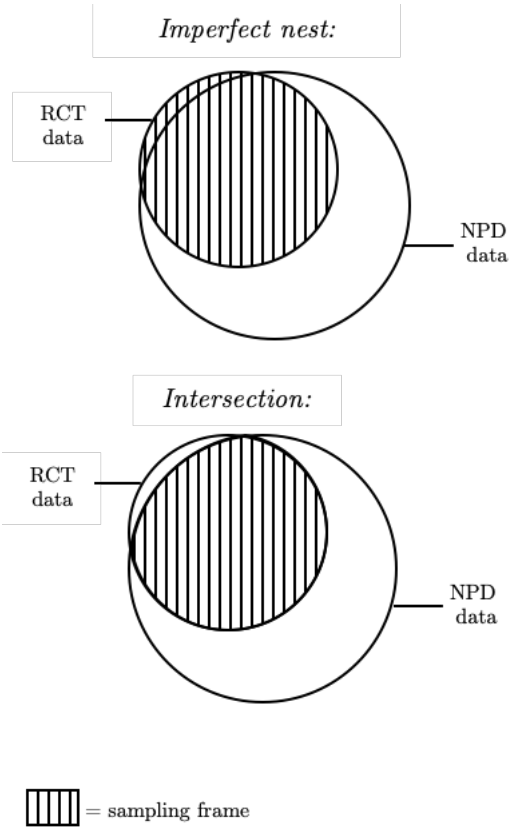


Fig. 5.1: Sampling frame of randomised controlled trial (RCT) and National Pupil Database (NPD) data depend on how missing data is handled. Patterned area denotes the proportion of participants included in the analysis.

What determines whether the linked trial-NPD data has an imperfect nest structure or an intersection structure? In my case, the structure is determined by the way missed links are handled because this has an effect on the analysis sampling frame (patterned area in **Fig. 5.1**). To create an imperfect nest structure, the missing NPD data from a missed link would need to be imputed before the analysis. As a result, information from the whole trial sample can be included in the analysis. By contrast, an intersection structure would arise if I used a complete-case approach to analyse the data: only participants with complete outcome and covariate data can contribute information to the analysis. Comparing the two linkage structures, it is easy to visualise that a complete-case analysis (i.e., an intersection structure) does not make use of all available information and, as a result, can give rise to selection bias if the linked participants are not representative of the participant population at

randomisation. I therefore used multiple imputation to handle missing data in the analyses, creating an imperfect nest sampling frame, and used complete-case analyses as a sensitivity analysis. Multiple imputation and sensitivity analyses are discussed in Chapter 7, from page 161.

5.2.2 Methods used to evaluate linkage quality

Several methods were used to evaluate the linkage quality of trial data to the NPD in this thesis:

5.2.2.1 Using agreement strength to approximate gold-standard data

In an ideal world, data analysts could use subsets of the linked data where true links are known (*gold-standard data*) in order to estimate linkage sensitivity and specificity and adjust analysis results accordingly (Harron et al., 2012). Yet, genuine gold-standard data is rarely available in practice (Harron et al., 2017b) and was also not available for this PhD study. To approximate a gold-standard dataset, information on the strength of agreement between personal identifiers in the trial and NPD data was requested from the FFT, who conducted the first step of the linkage in this study (**Fig. 4.3**, page 74). Agreement information helped to identify participant-pupil pairs who were especially likely to be true links. The type of identifiers and levels of agreement are discussed below in section 5.3. How I used agreement strength in practice is described in detail in section 5.4.4.

5.2.2.2 Negative controls to estimate false links

To estimate the minimum proportion of false links in each trial arm, it is possible to look at participants who died or were part of a birth cohort too old to be covered by the NPD data collection. No link was expected from these participants, but they were sent out for linkage as *negative controls* alongside eligible participants (Paixao et al., 2019). For example, most participants in the NEP-1 and NEP-2 trial were negative controls because they were part of a birth cohort too old to be captured in the NPD. Negative controls who did link were

intended to generate a minimum estimate of false links at the first linkage step and help find identifier-agreement-levels that are particularly likely to give rise to a false link (see section 5.4.4.2 on u-probabilities).

5.2.2.3 Comparison of linked and unlinked records

To identify subgroups of records that were more prone to linkage error and to estimate how representative the linked sample was of the originally randomised participants, characteristics of linked and unlinked participants were compared between trial arms.

5.3 Identification data used to link trial and NPD data

5.3.1 Participant-pupil identification data

To link trial and NPD data, personal identifiers that are shared between the trial and NPD data were used:

- Sex
- First names (and other first names if applicable)
- Surname (and other surnames if applicable)
- Date of birth
- Location history (all available addresses)

Trial identifiers were collected during the trial periods 1982-2002 and updated at each follow-up, whereas the NPD identifiers were returned by schools to the DfE at each termly census (Chapter 3). Data completeness for the trial identification data is shown in **Fig. 5.8**. Information on the missingness of identification data in the NPD is not available. However, it seems reasonable to assume that it has low missingness because schools capture and return identification data to the DfE every term, and both the DfE and FFT conduct regular cleaning and validation checks in consultation with schools and in relation to public league tables.

5.3.2 Digital recording of trial identifiers

5.3.2.1 Data sources

The identifying records from all trials were retrieved from the UCL archives in their original paper-based form. In total, there were 19 archive boxes, three notebooks, and four folders on a total of 4,051 participants[**]. About 80% of all records were handwritten; the rest were printed.

Fig. 5.2: Overview of physical records of personal identifiers processed in this PhD study

[**] Including participants from Scottish trials that were digitised with the intention to be linked to Scottish education data in the initial plan for this PhD study.

5.3.2.2 Data entry process of personal identification data

Within the UCL Data Safe Haven, I created a database using the software RedCap (UCL, 2021). RedCap enables data entry through customised digital forms. **Fig. 5.3** through **Fig. 5.7** show the forms which I refined throughout a pilot data entry phase. I applied automatic logic checks to avoid errors during data entry. After I entered 300 participants, I created a standard operating procedure manual for data entry, and two data entry assistants were hired. The data entry assistants underwent information governance training before gaining access to the data. A data entry diary was kept and shared between the three of us to record encountered difficulties which were discussed in a weekly meeting. The database was designed in such a way that a unique ID was generated for each new participant who was entered (**Fig. 5.3**). Records for each participant had to be accessed and updated multiple times, depending on the length of the trial and intensity of follow-up (**Fig. 5.4**) because identifying data for one participant was often fragmented across several separate physical locations (boxes, notebooks, folders) (**Fig. 5.6**). The total number of record-level updates was 10,639, with an average of 3.5 updates per participant. The trials did not always use the same participant identifiers for each follow-up study, but records could be retrieved by searching for attributes such as name, date of birth, or street. In most cases, postcodes were not recorded and had to be derived from street and city, using postcode directories. This required consideration for the year at which a participant lived at that address because postcodes can change over time in the UK. Where possible, twins and triplets were linked by saving sibling IDs into a participant's RedCap record (**Fig. 5.3**). This facilitated later logic checks in which it was expected that twins and triplets had the same addresses and date of birth (+/- one day) information.

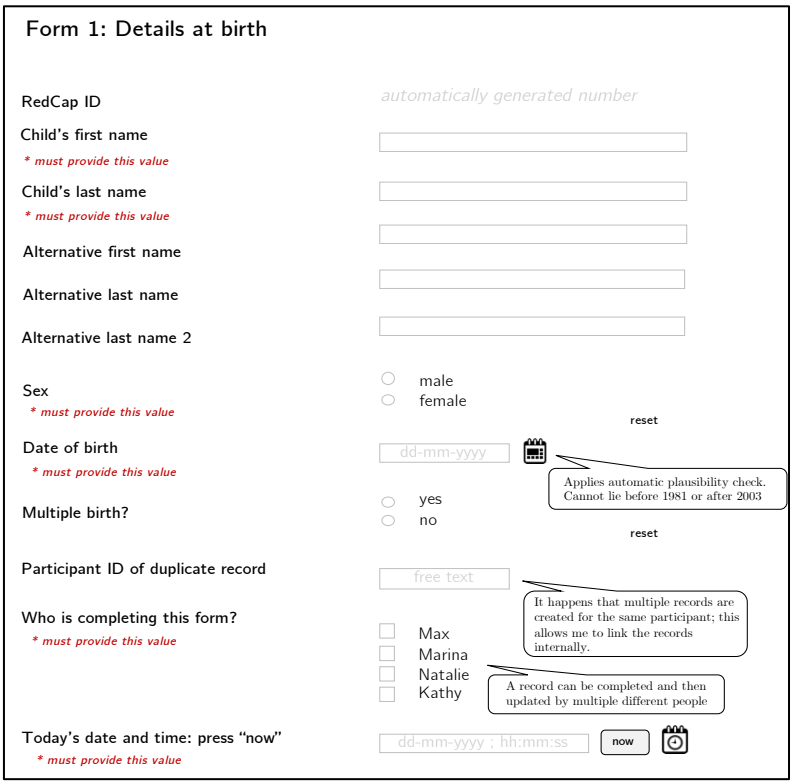


Fig. 5.3: Data extraction form for digitising identifiers (1/5)

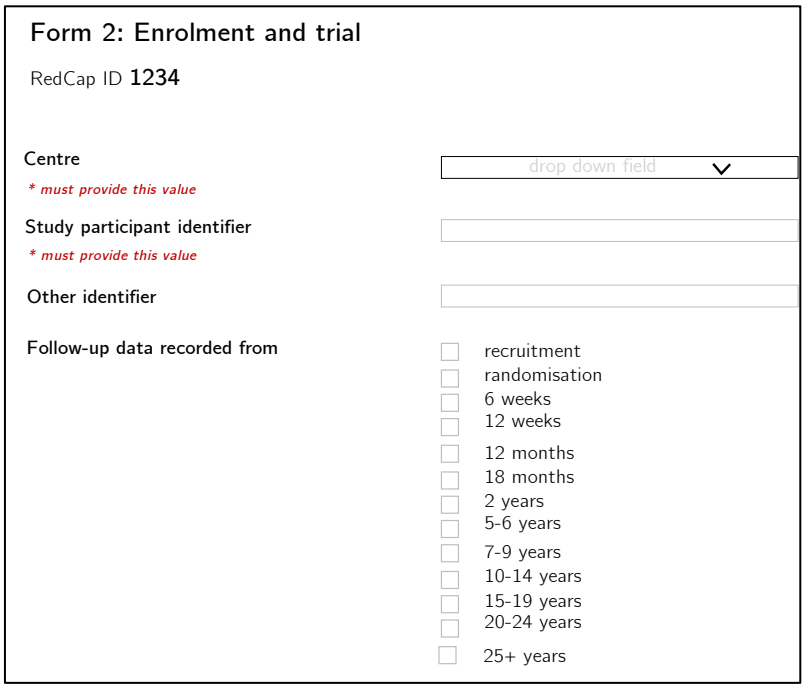


Fig. 5.4: Data extraction form for digitising identifiers (2/5)

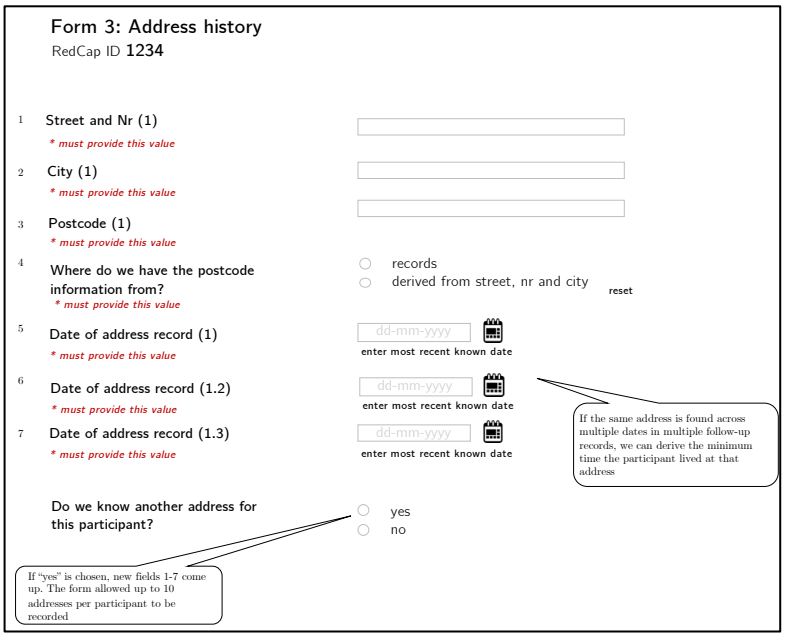


Fig. 5.5: Data extraction form for digitising identifiers (3/5)

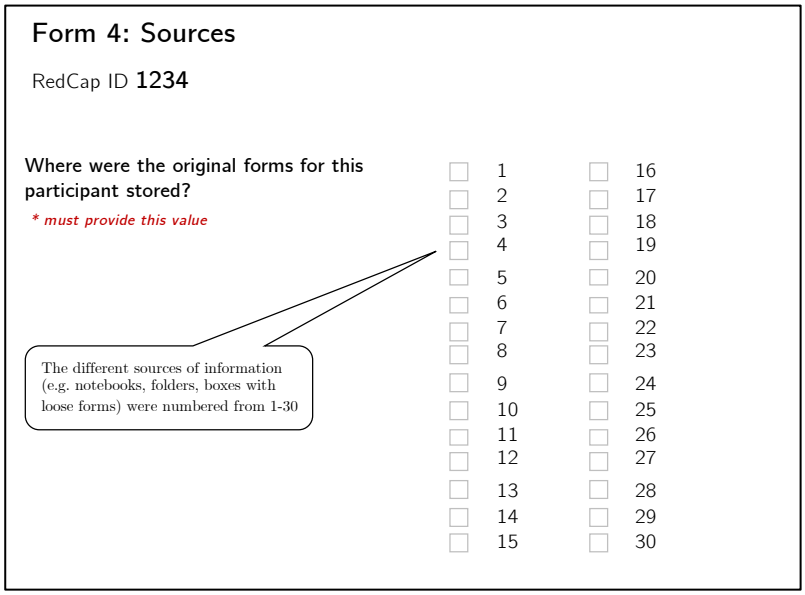


Fig. 5.6: Data extraction form for digitising identifiers (4/5)

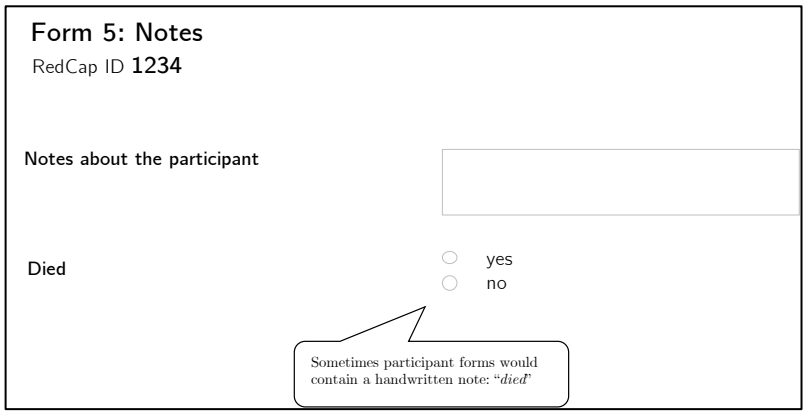


Fig. 5.7: Data extraction form for digitising identifiers (5/5)

5.3.2.3 Checks to detect and correct data quality issues in the identifier data

Prior to the digital recording of identifiers, I had created a database called "expected participants" that contained some basic information on all randomised participants: participant ID number(s), trial name, infant sex, date of birth, and recruitment centre based on publications and clinical trial data I had received from the trial investigators. The progress of identifier data entry was regularly cross-checked against the number of participants expected for each trial. The identifier data was then imported to the statistical software Stata for cleaning and validation. **Table 5.2** on the next page shows the validation actions I applied to ensure that the identification data was valid and plausible. Examples of validation actions include checking that all entries had plausible dates of birth, at least one postcode, first name, and last name, that twins and triplets had the same date of birth (± one day), and that infant sex and infant name combinations were plausible. Where problems were identified, I compared digitised data with paper records to correct mistakes; if uncertainty persisted, I set the field to missing.

Table 5.2: Validation actions for digitisation of identification data

Item	Action
Implausible address date	Cross-check with paper files – if nothing found set to missing.
Missing postcode	Check postcode directory for street and street number and derived postcode closest to trial time.
Missing date of address	If it's an IRON, NUCLEO or PALM study participant and it's the only address, replace date with date of birth (they did not update addresses). If not, leave missing.
Implausible RedCap data birthday	Check all for plausibility (around date of recruitment) range. Replace with clinical birthday if missing
Implausible clinical data birthday	Check if in plausible range, if not set to missing. Replace with redcap birthday if missing.
Implausible infant sex / name combinations	Use infant sex to check first names and identify potentially incorrect entries. Cross-check with paper files.
Common spelling mistakes in cities	Check cities against common city typos. Spell out abbreviations.
Common spelling mistakes in counties	Check counties against common county typos. Spell out abbreviations.
Exact duplicates	Drop all but one.
All personal identifier information missing	Cross-check with paper files.
Twins or triplets	Flag as multiples and use to complete missing identifiers.
Same address, different person	Cross-check again with paper-based files.
One person, multiple trials	Create one record per trial.
Multiples with different address	Cross-check again with paper-based files.
Multiples with different date of birth	Cross-check again with paper-based files.
Study number and trial do not match	Cross-check again with paper-based files.
Special character typos	Clean places and names from special characters and double spaces
A participant is in the digital attribute dataset but not in the identifier database	Check whether there is a problem with study number harmonisation. Check with paper-based files.
A participant is in the identifier database but not in the digital attribute dataset	Check whether problem with study number harmonisation. Check with paper-based files.
Redcap flag field incomplete/unverified	Inspect case by case, go back to paper files.
Age at last address	Boxplot, identify outliers and go back to paper files.

5.3.2.4 Completeness of participant identification data

Fig. 5.8 contains information on the missingness of identification variables by trial and trial arm. Overall completeness of personal identifiers was high. However, in most cases, it was unclear over which period a participant lived at a certain address. This would have been an important additional piece of information for linkage to pupil records because pupil records only contain identification data during school-age – and addresses recorded at birth and during school age might well be different. The reason for missing time stamps for the addresses in the RCT data are twofold. First, addresses were updated in the RCT data only when a child participated in a follow-up and only when there were changes compared to previously collected data. For example, if a participant moved in 2003 and the follow-up happened in 2005, the new address and the date of that address would be recorded in the trial forms as 2005, not 2003. If a participant did not move, no information was collected. Move dates were also not collected retrospectively. Second, some notebooks and forms did not include any date information nor information on the specific follow-up, meaning dates could not be derived. Hence, address data is only available with indirect time stamps, making it impossible to tell whether an address is *missing* for a given point in time during follow-up. The NEP-1 and NEP-2 trials were most affected by missing identifiers overall. The IRON trial was particularly affected by missing address dates as it had no planned follow-ups beyond the primary outcome, and many of the participant documents did not have a date field, so the missing address date was substituted as the date of recruitment in these records.

		Number of participants	First name	Last name	At least 1 postcode	Date of birth	Date of postcode
NEP-PD	Modified	*113*	0%	0%	0%	0%	0%
	Standard	*116*	0%	0%	0%	0%	0.9%
NETSGA	Modified	*152*	0.7%	0%	0%	0%	0%
	Standard	*147*	0.7%	0%	0%	0%	0%
LCPUFAP	Modified	*96*	2.1%	0%	0%	0%	0%
	Standard	*100*	1.0%	0%	0%	0%	1.0%
LCPUFAT	Modified	*155*	0%	0%	0%	0%	0%
	Standard	*154*	0.6%	0%	0%	0%	0%
IRON	Modified	*162*	0%	0%	0%	0%	8.0%
	Standard	*165*	0%	0%	0%	0%	11.5%
PALM	Modified	*103*	1.0%	0%	0%	0%	0%
	Standard	*100*	0%	0%	0%	0%	0%
NEP-1	Modified	*219*	0.5%	0.5%	0%	3.2%	0%
	Standard	*204*	0%	0%	0%	4.9%	1.0%
NEP-2	Modified	*178*	0%	0%	0%	2.2%	0%
	Standard	*191*	0%	0%	0%	2.7%	0%
NUCLEO	Modified	*99*	0%	0%	0%	0%	0%
	Standard	*97*	0%	0%	0%	0%	0%

Missingness indicator:

0% 10% 20%

Fig. 5.8: Proportion of missing identification data by trial and trial arm.

5.3.3 Level of agreement in trial-NPD identification data

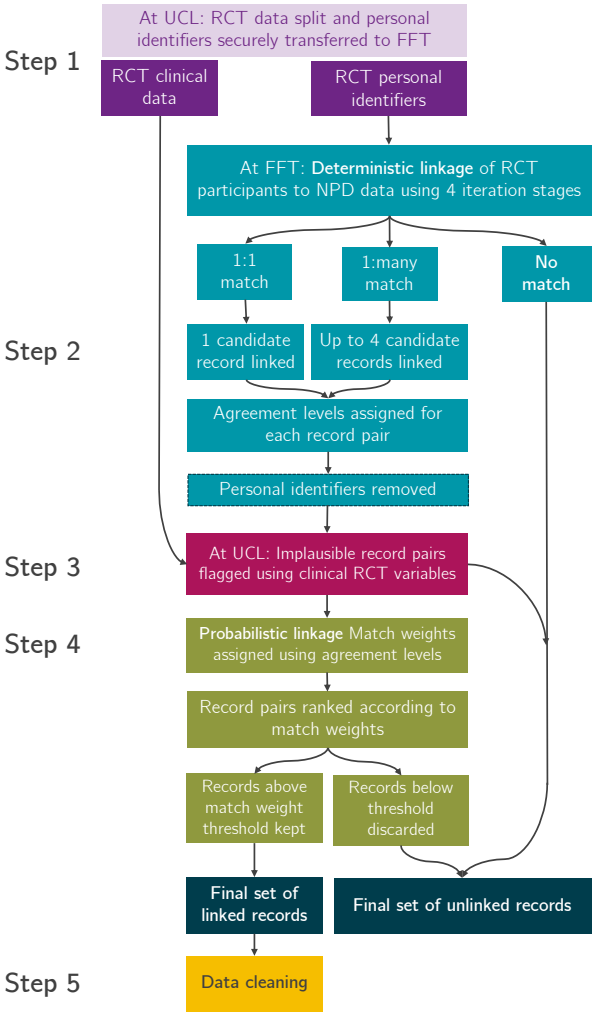
Participant identification data was sent to the FFT. I arranged with the FFT that they supply me with information on the strength of agreement between participant and pupil identifiers for all candidate links. This was to make the quality of linkage more transparent, so I could make decisions about the best possible pupil match for each participant without needing access to the identifiers myself. For each given pupil-participant candidate pair, there is only one possible level of agreement strength *per personal identifier*. **Table 5.3** gives an overview of the different agreement levels that were possible for each identifier. Note that more than one level within each identifier can denote exact agreement, but some carry more information about the link than others. For instance, a link between a participant record with two first names that match exactly to a pupil record with two first names is more certain than a link between a participant record that only contains one first name matching exactly to a pupil record with only one first name.

Table 5.3: Agreement strength scenarios between linkage variables: hierarchy of agreement in descending levels of certainty (each record pair can only have one possible level of agreement strength per identifier variable)

Identifier variable	Description
First name	First name and other first name both match exactly
	First name matches other name in both directions
	First name matches exactly
	First name matches other name
	Other name matches exactly
	First name truncated at any hyphen matches
	First name matches via common name alternatives lookup
	Pattern match function
	Pattern match function - AND first character of first name matches
	First name/surname match in both directions
	No link
Surname	Surname matches exactly (including alternative surnames)
	Surname truncated at any hyphen matches
	Pattern match function
	Pattern match function - AND first character of surname matches
	First name/surname match in both directions
	No link
Date of Birth	Date of birth matches exactly (day, month and year)
	Day on source matches month on match, and vice versa; year matches (i.e., transposed date)
	Day and month match (i.e., wrong year)
	Day and year match (i.e., wrong month)
	Month and year match (i.e., wrong day)
	Either participant or pupil DOB is 1st January; year matches
	Either participant or pupil DOB is 1st September; year matches
	No link
Location	Postcode matches exactly
	Local Authority matches
	Neighbouring/nearby Local Authority matches
	No link

5.4 Step by step overview: how RCT and NPD data were linked in this thesis

The linkage iterated through a series of steps, first allowing a broad definition of what might constitute a match and then using agreement levels between linkage variables to select the most plausible link of multiple possible links. The linkage strategy is illustrated in **Fig. 5.9**, with resulting linkage rates shown in **Fig. 5.15.** The individual steps are discussed in detail below.



FFT *Fischer-Family Trust*; NPD *National pupil data base*; RCT *randomised controlled trial* ; UCL *University College London*

Fig. 5.9: Overview of linkage strategy

5.4.1 Step 1: Preparing data for linkage

The trial data described in Chapter 2 was kept in two separate datasets: a clinical dataset and a personal identifier dataset, which did not include any clinical data. A unique participant identification number was shared between both trial datasets. Section 5.3.2 above describes in detail how the personal identifier dataset was prepared to allow linkage between trial data and school data.

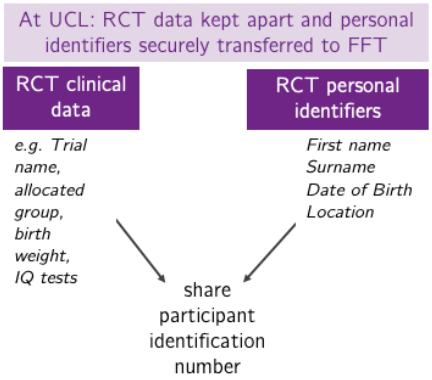


Fig. 5.10: Step 1

5.4.2 Step 2: Deterministic linkage at the Fischer Family Trust

Deterministic linkage is a method that relies on rules of agreement between linkage variables (e.g., 'agrees exactly on name and postcode' independent of the likelihood of that exact agreement). It typically iterates through a series of linkage stages, from requiring exact matches to incorporating more approximate matches.

The initial linkage was deterministic and not conducted by me at UCL but by the Fischer Family Trust (FFT), a private organisation authorised by the Department for Education to process and share NPD data with third parties. I securely shared participant identification data (no clinical data) for all randomised participants with the FFT. FFT then linked the identification data to pupil records in the NPD in several stages.

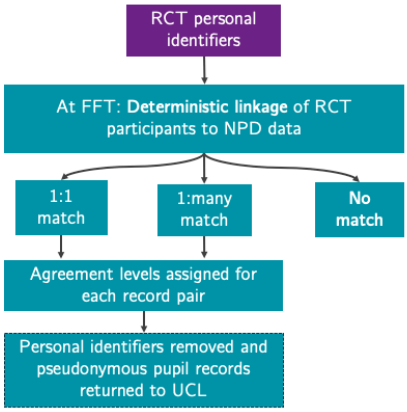


Fig. 5.11: Step 2

The initial stage only considered exact links for the supplied personal identifiers (1:1 links). Following this, 'fuzzy' linkage was conducted, yielding multiple

possible links for each participant (1:many links). Fuzzy linkage took into account more approximate agreement levels between identifiers, including neighbouring postcodes and authorities, as well as name checks against previous surnames and preferred first names. If a participant had multiple addresses in their identification data, all were given equal weight. The FFT matching routines found the strongest link to any record in the history of the matched pupil. This was then followed by manual checking of the results. Overall, each participant linked to up to four different pupils with various agreement patterns. An indicator of agreement strength for each identifier was provided for each candidate record pair (**Table 5.3**, page 95 above). FFT then removed personal identifiers leaving only the NPD candidate link and the unique participant number for each participant. These were then securely transferred into the UCL data safe haven (see **Table 5.4** for an overview of supplied files).

Table 5.4: Overview of NPD Files supplied by the Fischer Family Trust

Filename		Rows and Columns	Description
1	**AllData**	1,766,399 observations 7 variables: TableID, PupilReferene, Dataset, acyear, laestab, NPDfieldreference, indicator value	Covers academic years 2000-01 until 2015/16.
2	**AllDataSubj**	2,784,513 observations 9 variables: TableID, PupilReferene, Leapcode, Qan, Dataset, acyear, laestab, NPDfieldreference, indicator value	Covers academic years 2003/04 until 2015/16 and information on KS2Exam and KS4Exam
3	**LinkTable**	3,544 observations 11 variables: TableID, studyid1, PupilReference, MatchScore, Assigned, FNScore, SNScore, DBScore, LocationScore, CommonNameScore, MatchProbability	Brings together participant ID and pupil ID. It assigns the different scores for the linkage variables

5.4.3 Step 3: Flagging implausible record pairs in-house

In step 3, I checked the plausibility of all participant-pupil links. This was necessary because not every participant was expected to link to a pupil in the NPD (e.g., due to death or non-overlapping period of trial / NPD data collection). Implausible links, therefore, approximate false links. To facilitate the plausibility checks, I used auxiliary variables from the RCT clinical data, such as information on death, to the NPD data using the shared participant identification number. **Table 5.5** shows which variables were used to perform the plausibility checks and from which source they were obtained:

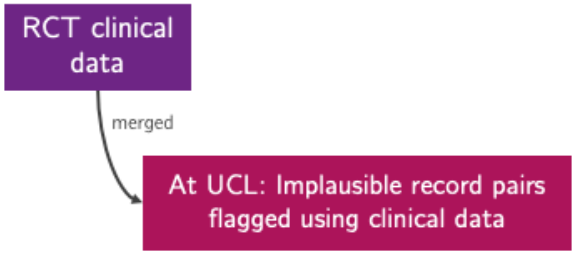


Fig. 5.12: Step 3

Table 5.5: Auxiliary variables used to perform plausibility checks.

Variables	RCT clinical data	NPD
Participant ID number	✓	✓
Pupil ID number		✓
Trial name	✓	
Death (if applicable)	✓	
Recruitment centre	✓	
Gender	✓	✓
Age during academic year	✓	✓
Date of study entry	✓	
Attendance at previous follow-ups	✓	
Agreement level indicator		✓
Sibling status	✓	
Sibling participant ID number	✓	
Took part in multiple trials	✓	
First linked academic year		✓
Last linked academic year		✓

5.4.3.1 Known fact of death

The background rate for child death is very low in the UK but generally higher in infancy and for children born small or preterm. This means I could expect most participant deaths to be known and recorded in the trial data. As expected, mortality was higher in trials that were conducted among preterm-born babies and babies born small-for-gestational-age. There were 100 known cases of participant death (**Fig. 2.2**). If a participant who was known to be dead matched to a pupil record, this record pair was flagged as a false link.

5.4.3.2 Implausible age

I flagged 271 record pairs as false links where pupils in the earliest available academic year were older than 17 years, and the agreement strength for the date of birth identifier indicated 'no link'. This is because the age from the trial data is highly likely to be correct as it corresponds to the recruitment periods and was confirmed repeatedly during the follow-up period. Seventeen years was chosen as a cut-off because this age is above the age expected for a pupil in KS4 (pupils in KS4 are usually 14-16 years old). Implausible age concerned mostly participants from NEP-1 and NEP-2 trials with birth years 1982 to 1985, for which only a small subgroup was expected to link to any NPD module.

5.4.3.3 Twins and triplets

The risk of linkage error is particularly high for those who share the same birthday, address, and last name: twins and triplets. According to the RCT data, there were 366 children randomised together with their twin or triplet siblings in the same trial. 10.5% of all twins and triplets had more than one candidate pupil record per participant record (compared to 8.2% of singleton participants). If twins and triplets of different sexes were linked to the same candidate pupil record, I compared the infant sex recorded in the trial data to the infant sex recorded in the NPD data and flagged those who did not agree on sex as false links.

5.4.3.4 Children who took part in multiple trials

Children who participated in more than one trial had more than one participant identification number (1 per trial) but were expected to only have one correct pupil link. A total of 169 children participated in both the nutrient-enriched post-discharge formula trial for preterm infants and the LCPUFA formula for preterm infants trial. All links of these children were manually reviewed to ensure that the best matching pupil record they linked to was the same across all participant IDs.

5.4.3.5 Negative controls

There were a total of 462 negative control participants who were not expected to link to the NPD, either based on their birthdates (born before 1984) or because RCT records indicated that they had passed away. Participants who did link unexpectedly would have allowed me to identify factors associated with false-positive links and give me a minimum estimate of false-positive links produced through the deterministic linkage step. However, only a small number (n=24) of negative control participants had a pupil record, and none of their NPD modules contained any data. Indeed, identification data agreement patterns indicated that linked participant-pupil records among the negative controls were likely to be true links because all agreed on date of birth exactly, and the most common pattern was an exact match on all identifiers. The fact that these unexpected links were likely to be true links, indicates that the NPD data held at the FFT goes slightly further back in time than expected (even if pupil data was not recorded). None of the participants who died had linked pupil records. Therefore, the negative controls suggest that the deterministic linkage step at FFT was likely to introduce few, if any, false-positive links.

5.4.4 Step 4: Probabilistic linkage and de-duplication

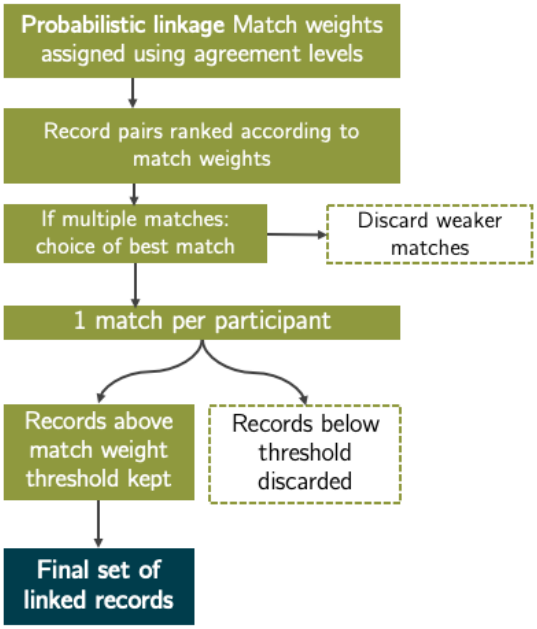


Fig. 5.13: Step 4

In step 4, I used probabilistic linkage to determine the one best match for each participant (if any). In contrast to deterministic linkage, probabilistic linkage makes use of the likelihood that identifiers agree under certain conditions. This likelihood is used to derive an overall match weight for each record pair. First, two conditional probabilities are calculated for each agreement level (overview of agreement levels in **Table 5.3** above):

> *M-probability:* Probability of two records achieving a specific agreement level given they belong to the same child
> *U-probability:* Probability of two records achieving a specific agreement level by chance (i.e., the records do not belong to the same child)

5.4.4.1 Calculating m-probabilities

The M-probability can be estimated using the error rate in a specific identifier. For instance, if month of birth was not matching in 3% of record pairs truly belonging to the same child, the m-probability for month of birth would be 0.97.

5.4.4.2 Calculating u-probabilities

The u-probability represents the probability of agreement by chance. For example, the probability that any two records agree by chance on month of birth is approximately $1/12$ (12 different calendar months).

5.4.4.3 Approximating m- and u-probabilities

I did not have information on the pairs that truly belonged or did not belong to the same child, so I used pairs with exact agreement on each identifier to approximate true links and priors from a previous study. This previous study linked UK infants from a trial to administrative hospital data but had gold-standard data and was therefore able to estimate false links (Harron, 2014). **Table 5.6** on the next page shows the m- and u-probabilities that were used for the identifiers in this thesis.

Table 5.6: Agreement strength levels of linkage variables with corresponding m- and u-probabilities

Identifier	Agreement strength level	M-probability	U-probability
Infant sex	Agrees	0.9596	0.5200
	Disagrees	0.0249	0.4700
	Missing in at least one record	0.0155	0.0100
First name	Both first names match	0.0657	0.0036
	First name matches	0.8706	0.0099
	Others match	0.0130	0.0020
	First name truncated matches	0.0040	0.0095
	First name matches alias	0.0233	0.0070
	Pattern match1	0.0091	0.0180
	Pattern match2	0.0041	0.0180
	Matches surname and vice versa	0.0010	0.0020
	No link	0.0092	0.9300
Surname	Surname matches exactly	0.9387	0.0200
	Surname at hyphen matches	0.0125	0.0050
	Pattern match function 1	0.0190	0.0200
	Pattern match function 2	0.0170	0.0200
	Matches first name and vice versa	0.0100	0.0010
	No link	0.0028	0.9340
Date of birth	DOB matches	0.9830	0.0200
	Transposed date	0.0045	0.0500
	Wrong year	0.0001	0.0500
	Wrong month	0.0038	0.0900
	Wrong day	0.0085	0.1000
	No link	0.0001	0.6900
Location	Postcode exact match	0.5200	0.1500
	Local authority matches	0.2900	0.1000
	Neighbouring authority matches	0.1400	0.0500
	No link	0.0500	0.7000

Note that, within each identifier, m-probabilities add up to 1 and u-probabilities add up to 1.

5.4.4.4 Calculating match weights

After calculating the m-and u-probabilities, match weights were assigned to the record pairs according to the likelihood that pairs would agree or disagree on certain linkage variables. For each linkage variable (infant sex, first name, last name, date of birth, location) the variable weight is calculated using: $log_2(m/u)$. The final match weight for a participant-pupil pair was then calculated using the sum of all weights across all identifiers in that pair: $\sum log_2(m_i/u_i)$. This resulted in large positive weights for record pairs that agree on multiple (and rarer) identifiers and negative weights for record pairs with disagreements.

The most common agreement pattern (32.8%) was an exact match on first name, surname, date of birth, local authority and infant sex and corresponded to a match weight of 20.05597. Candidate pairs were then ranked by match weight, with the lower-ranked pairs having a lower probability of being a true match and the higher ranked pairs having a higher probability of being a true match.

5.4.4.5 De-duplication: choosing the best match for each participant

Some participants still matched to more than one candidate pupil record. When the highest match weight for a pair was 10% above that of the second-best match (arbitrarily selected threshold), I automatically kept the best match and discarded the other(s). When the difference was lower, I manually reviewed the pair. This was the case for 8 pairs (3 unique participants), of which 2 were twins or triplets. In the manual review, I decided to discard all pairs where there was no exact link between the first name in the participant and the pupil record, given that this is the variable that would differ between twins/triplets.

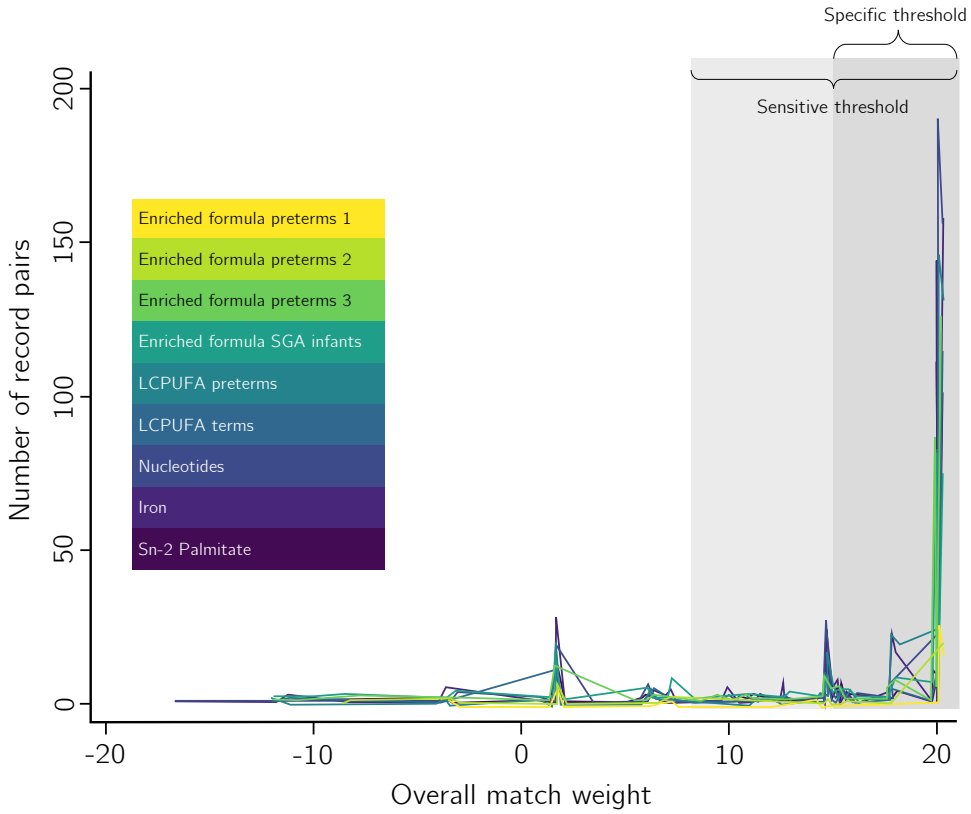


Fig. 5.14: Match weight distribution, sensitive (9+) and specific (15+) thresholds

5.4.4.6 Choice of thresholds: sensitivity analyses

The remaining de-duplicated candidate pairs were plotted on a histogram (**Fig. 5.14**). We would expect sufficiently strong pairs to group to the right of the graph and weak links to group to the left of the graph.

To identify further pairs where no record was likely to be a true match, I chose two cut-off thresholds based on visual inspection of the match weight plot: one sensitive threshold (increasing likelihood of true positive links but also increasing likelihood of false-positives) and one specific threshold (decreasing likelihood of false-positive links but also decreasing likelihood of true positives). Links below these thresholds were considered to be too tenuous to be included in any further analysis.

The advantage of maximising my analysis sample for sensitivity would be to have a larger sample size (by 61 participants). This would, however, make it more likely that false-positive links are included in the final analysis sample.

Under the assumption that false-positive links are equally likely in modified formula and standard formula groups, this could introduce more noise and dilute any true differences in academic performance between modified formula and standard formula groups. Conversely, having a highly specific analysis sample would mean that false-positive links are minimised, but that some true positives are likely to be excluded, statistical power is reduced, and selection bias is possibly introduced.

To decide whether the analysis sample should be based on the sensitive or the specific threshold, I compared the observed characteristics of the sensitive and specific samples to the originally randomised sample like they would be used in the analysis: with one sample partly contained within the other (9+ and 15+, rather than 9-14.99 and 15+) (**Table 5.7** on the following page). I also wanted to see whether school results (i.e., outcomes) differed for participants defined by the two thresholds. Finally, I compared the characteristics of participants without (plausible) links to the original sample. Because the NEP-1, NEP-2 and NUCLEO trial were not included in the analysis sample, they were excluded from this sensitivity analysis.

Table 5.7: Comparison of characteristics in the randomised sample and the analysis samples defined by the sensitive (9+) and specific (15+) match weight threshold (excludes negative control trials NEP-1, NEP-2 and NUCLEO)

	Rand. sample	Match wt. 9+	Match wt. 15+
	(n=1563)	(n=1431)	(n=1370)
In modified formula group	780 (49.9%)	709 (49.6%)	678 (49.5%)
Baseline characteristics			
Male	793 (50.8%)	733 (51.0%)	705 (52.3%)
Birth weight, grams	2755 (1007)	2753 (1006)	2752 (1012)
Gestational age, wks.	37.3 (4.4)	37.2 (4.4)	37.2 (4.5)
Mum has degree	103 (7.8%)	87 (7.2%)	83 (7.1%)
Maternal age, years	27.3 (5.2)	27.4 (5.2)	27.4 (5.2)
Mum smoked while pregnant	522 (34.4%)	471 (33.9%)	452 (34.0%)
Previous cognitive outcomes			
BS MDI	91.2 (13.8)	91.2 (13.8)	91.5 (13.9)
IQ	106.0 (13.9)	106.3 (14.0)	106.7 (13.9)
Academic outcomes			
GCSE Maths (0-8 point scale)	-	4.5 (1.8)	4.5 (1.8)
GCSE English (0-8 point scale)	-	4.8 (1.6)	4.8 (1.5)
KS2 Maths (0-100 point scale)	-	58.6 (23.2)	58.8 (23.1)
KS2 English (0-50 point scale)	-	27.1 (9.8)	27.1 (9.8)
Ever qualified for special educational needs support	-	698 (48.7%)	667 (48.7%)

Data are mean (SD) or n (%); BS MDI *Bayley Scale of Infant Development Mental Development Index*; GCSE *General Certificate of Secondary Education*; KS2 *Key Stage 2*; SD *Standard Deviation*; wks. *weeks*; wt. *weight*; School grades: higher = better performance

Overall, there was little difference in terms of key characteristics between the participants who fell in the sensitive or the specific sample compared to the randomised sample. However, the sensitive sample was closest to the original sample. I assumed that randomisation was successful and observed characteristics also reflect unobserved characteristics, I decided to choose the sensitive sample to minimise selection bias. **Fig. 5.15** on page 110 shows the resulting linkage rates.

5.4.5 Step 5: Data cleaning

In the last step, I cleaned the linked data for analysis. The full Stata code is published online: https://github.com/MaxVerfuerden/PhD. Briefly, this included deriving and extracting outcome and covariate information from the different NPD files (**Table 5.4**), harmonising different scales of the same exam scores over time and creating a single dataset with one row per participant.

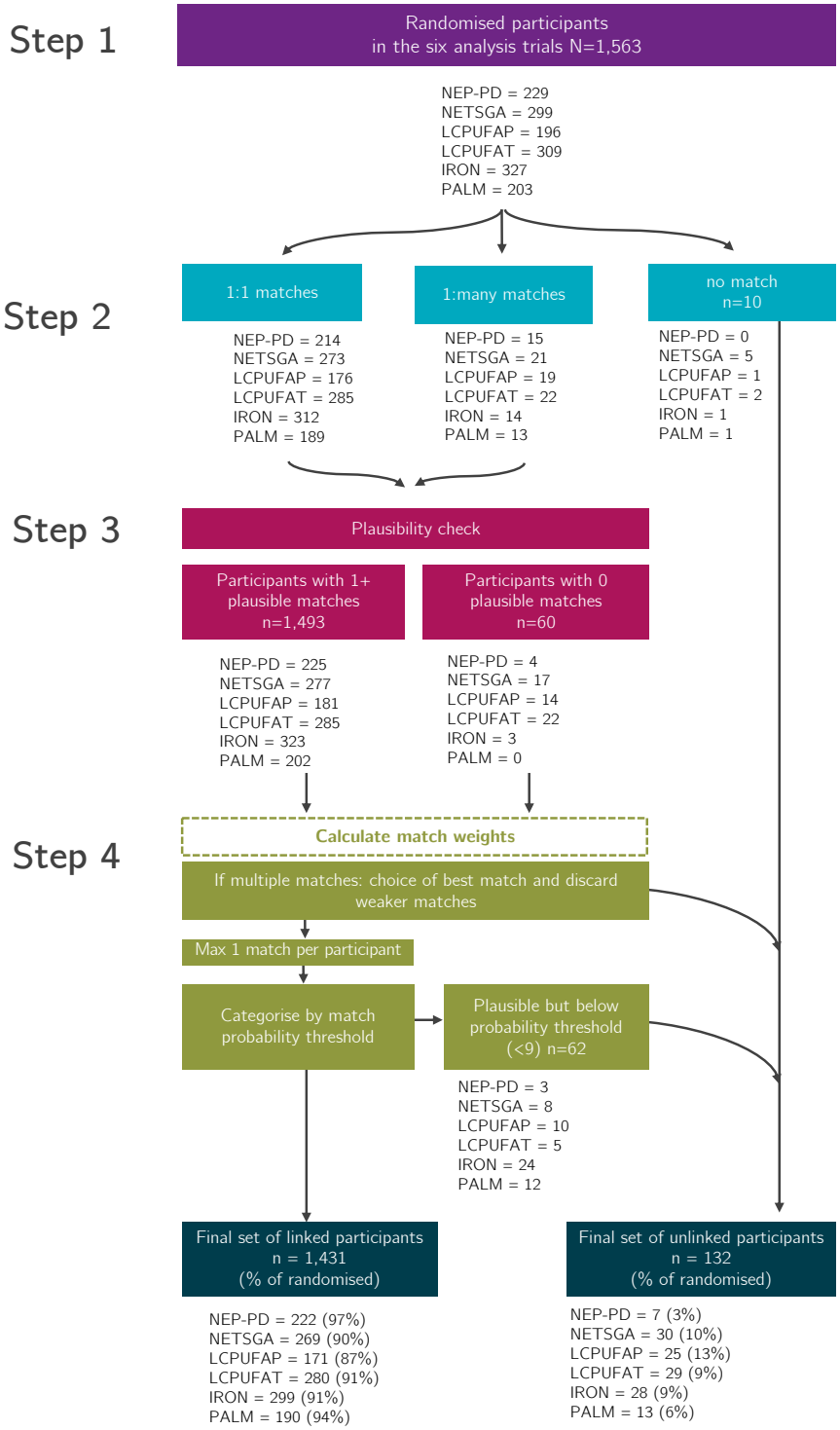


Fig. 5.15: Linkage rates at each step in the six trials eligible for analysis

5.5 Description of the linked sample

The linked trial-education data combines birth characteristics and information collected during trial follow-up with de-identified information about a participant's school trajectory (**Fig. 5.16**).

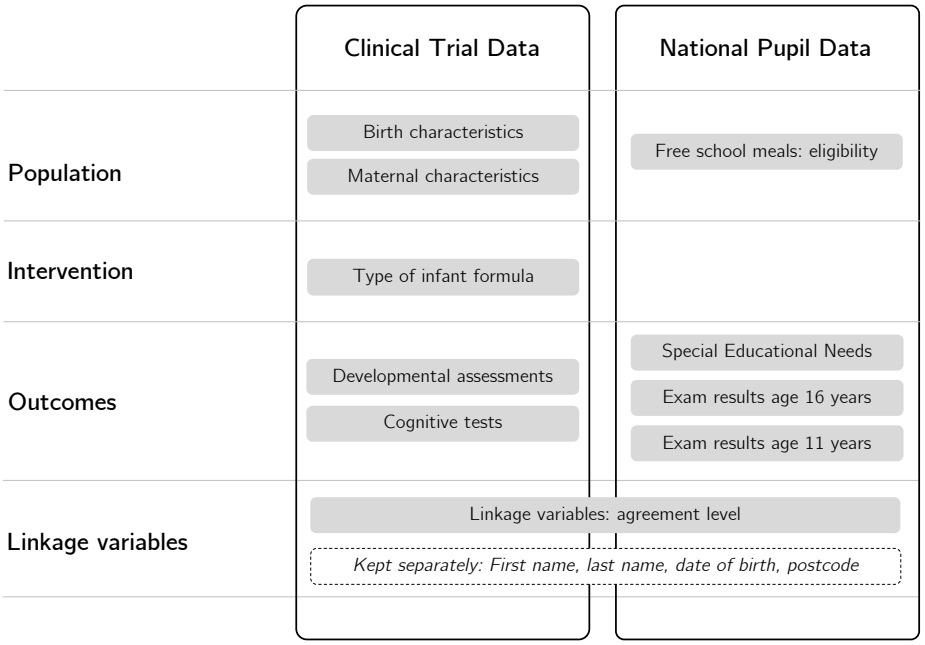


Fig. 5.16: Linked trial-NPD data: variable sources (identification data not available to me once linked data was received)

5.5.1 Overall linkage success rate

The number of participants from eligible trials[††] considered successfully linked (i.e., plausible and match weight 9+) to the NPD data was 1,431 (91.6% of all randomised participants).

[††] All trials except NEP-1, NEP-2, and NUCLEO

5.5.2 Linkage success rates by trial, formula group and key stage

Table 5.8 gives the linkage rates by trial, formula group, and key stage.
The rate of missed matches did not differ appreciably between modified and standard formula groups. Not all participants with a pupil record also had records for the relevant NPD key stages at which high-stake school exams were recorded (KS2 and KS4, i.e., GCSEs). The corresponding statistical power for the primary analysis is calculated and discussed on page 166 (Chapter 7).

Table 5.8: Linkage rates by trial and key stage in the analysis trials

	Randomised	NPD		KS2 module (age 7-11y)		KS4 module (age 14-16y)	
	N	n	%	n	%	n	%
Total	**1563**	**1431**	**92%**	**1306**	**83%**	**1342**	**86%**
NEP-PD							
All	*229*	*222*	*97%*	192	*84%*	*211*	*93%*
High energy	113	108	96%	88	78%	101	89%
Standard	116	114	98%	104	90%	110	95%
NETSGA							
All	*299*	*269*	*90%*	*254*	*85%*	*258*	*86%*
High energy	152	135	89%	125	82%	130	86%
Standard	147	134	91%	129	88%	128	87%
LCPUFAP							
All	*196*	*171*	*87%*	*129*	*66%*	*157*	*80%*
LCPUFA	96	82	85%	62	65%	74	77%
Unfortified	100	89	89%	67	67%	83	79%
LCPUFAT							
All	*309*	*280*	*91%*	*269*	*87%*	*263*	*85%*
LCPUFA	155	137	88%	131	85%	127	82%
Unfortified	154	143	93%	138	90%	136	88%
IRON							
All	*327*	*299*	*91%*	*283*	*87%*	*281*	*86%*
High-iron	162	149	92%	136	84%	138	85%
Standard iron	165	150	91%	147	89%	143	87%
PALM							
All	*203*	*190*	*94%*	*179*	*88%*	*172*	*85%*
Sn2-palmitate	103	98	95%	94	91%	91	88%
Standard pal.	100	92	92%	85	85%	81	81%

5.5.3 Linked sample and missed links vs original sample

Table 5.9 below shows that there were no major sociodemographic differences in terms of allocation group, infant sex, birth weight, gestational age, maternal age, maternal smoking during pregnancy, and maternal education at birth between the originally randomised group of participants and the linked analysis sample within each trial. Also, the missed links were not statistically significantly different from the original sample in those key characteristics. The linked sample characteristics for each trial further divided by formula group are given in **Table 8.1**, on page 187.

Table 5.9: Comparison of observed baseline characteristics in original samples vs in NPD-linked samples and missed links, by trial (Nr.1 out of 3 tables)

Variable	NEP-PD					NETSGA				
	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*
N	229	222		7		299	269		30	
Modified formula	113/229 (49%)	108/222 (49%)	*0.931*	*		152/299 (51%)	135/269 (50%)	*0.929*	17/30 (57%)	*0.734*
Male	110/229 (48%)	107/222 (48%)	*0.984*	*		142/299 (47%)	131/269 (49%)	*0.865*	11/30 (37%)	*0.480*
Birth weight, grams	1369 (297)	1375 (298)	*0.831*	1332 (183)	*0.744*	2567 (291)	2568 (290)	*0.967*	2556 (309)	*0.845*
Gestational age weeks	30.8 (2.5)	30.8 (2.4)	*1.000*	30.2 (4.0)	*0.540*	39.2 (1.3)	39.2 (1.3)	*1.000*	38.8 (1.1)	*0.105*
Mother's age, years	28.4 (5.7)	28.3 (5.7)	*0.852*	31.7 (5.0)	*0.132*	26.6 (5.1)	26.7 (5.0)	*0.814*	25.6 (5.9)	*0.314*
Mother smoked during pregnancy	78/219 (34%)	75/213 (35%)	*0.952*	*		130/276 (47%)	116/247 (47%)	*0.985*	14/29 (48%)	*0.943*
Mother has degree	20/223 (9%)	19/217 (9%)	*0.943*	*		15/297 (5%)	11/268 (4%)	*0.609*	*	

*Data are mean (SD) or n (%), two-sided p-value: students t-test for means and chi square test for proportions in comparison with original sample; *suppressed due to small cell sizes (n<5)*

Table 5.10: Comparison of observed baseline characteristics in original samples vs in NPD-linked samples and missed links, by trial (2/3)

Variable	LCPUFAP					LCPUFAT				
	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*
N	196	171		25		309	280		29	
Modified formula	96/196 (49%)	85/171 (50%)	*0.935*	14/25 (56%)	*0.707*	155/309 (50%)	137/280 (49%)	*0.862*	18/29 (62%)	*0.499*
Male	95/196 (48%)	85/171 (50%)	*0.927*	10/25 (40%)	*0.705*	165/309 (53%)	150/280 (54%)	*0.981*	15/29 (52%)	*0.924*
Birth weight, grams	1342 (281)	1348 (270)	*0.836*	1299 (347)	*0.484*	3594 (437)	3592 (438)	*0.956*	3614 (434)	*0.814*
Gestational age weeks	30.3 (2.4)	30.3 (2.3)	*1.000*	30.0 (2.7)	*0.562*	40.0 (1.3)	40.0 (1.3)	*1.000*	40.1 (1.3)	*0.692*
Mother's age, years	26.4 (5.2)	26.4 (5.2)	*1.000*	26.2 (5.3)	*0.743*	27.3 (5.2)	27.4 (5.2)	*0.816*	25.6 (4.6)	*0.090*
Mother smoked during pregnancy	81/196 (41%)	71/171 (42%)	*0.981*	10/25 (40%)	*1.000*	74/301 (25%)	67/273 (25%)	*0.993*	7/28 (25%)	*0.970*
Mother has degree	10/109 (9%)	8/101 (8%)	*0.766*	*		19/304 (6%)	17/276 (6%)	*0.966*	*	

*Data are mean (SD) or n (%), two-sided p-value: students t-test for means and chi-square test for proportions in comparison with original sample; *suppressed due to small cell sizes (n<5)*

Table 5.11: Comparison of observed baseline characteristics in original samples vs in NPD-linked samples and missed links, by trial (3/3)

Variable	IRON					PALM				
	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*	Original sample	Linked sample	*p-value linked vs original*	Missed links	*p-value missed vs original*
N	327	299		28		203	190		13	
Modified formula group	161/327 (49%)	149/299 (50%)	*0.931*	12/28 (43%)	*0.698*	103/203 (51%)	98/190 (52%)	*0.925*	5/13 (39%)	*0.607*
Male	163/327 (50%)	152/299 (51%)	*0.887*	11/28 (39%)	*0.598*	118/203 (58%)	108/190 (57%)	*0.933*	*	
Birth weight, grams	3479 (473)	3476 (466)	*0.936*	3513 (551)	*0.719*	3528 (440)	3532 (443)	*0.929*	3476 (415)	*0.679*
Gestational age weeks	39.8 (1.4)	39.8 (1.4)	*1.000*	40.1 (1.3)	*0.275*	40.0 (1.3)	40.0 (1.3)	*1.000*	39.8 (1.4)	*0.593*
Mother's age, years	27.6 (4.9)	27.7 (4.8)	*0.797*	26.8 (5.4)	*0.411*	27.2 (5.4)	27.3 (5.3)	*0.853*	26.1 (7.1)	*0.486*
Mother smoked during pregnancy	95/322 (30%)	83/295 (28%)	*0.727*	12/27 (44%)	*0.260*	64/203 (32%)	59/190 (31%)	*0.942*	5/13 (39%)	*0.715*
Mother has degree	36/323 (11%)	30/296 (10%)	*0.715*	6/27 (22%)	*0.147*	*	*		*	

*Footnotes: Data are mean (SD) or n (%), two-sided p-value: students t-test for means and chi-square test for proportions in comparison with original sample; *suppressed due to small cell sizes (n<5)*

5.6 Strengths and limitations

The different linkage stages were each subject to a different set of limitations and assumptions. From a linkage accuracy point of view, one might criticise the lack of transparency generated by delegating the first step of the linkage to an external party (the FFT), which returned only a de-identified list of candidate links. This approach was necessary, however, to preserve participant confidentiality and is common data linkage practice. Fortunately, the FFT agreed to supply several possible links per participant and provide information on the strength of linkage agreement. This enabled me to use additional steps to refine and evaluate the linkage. None of the negative control participants were returned with implausible NPD links, which provides some evidence for high linkage quality at this step.

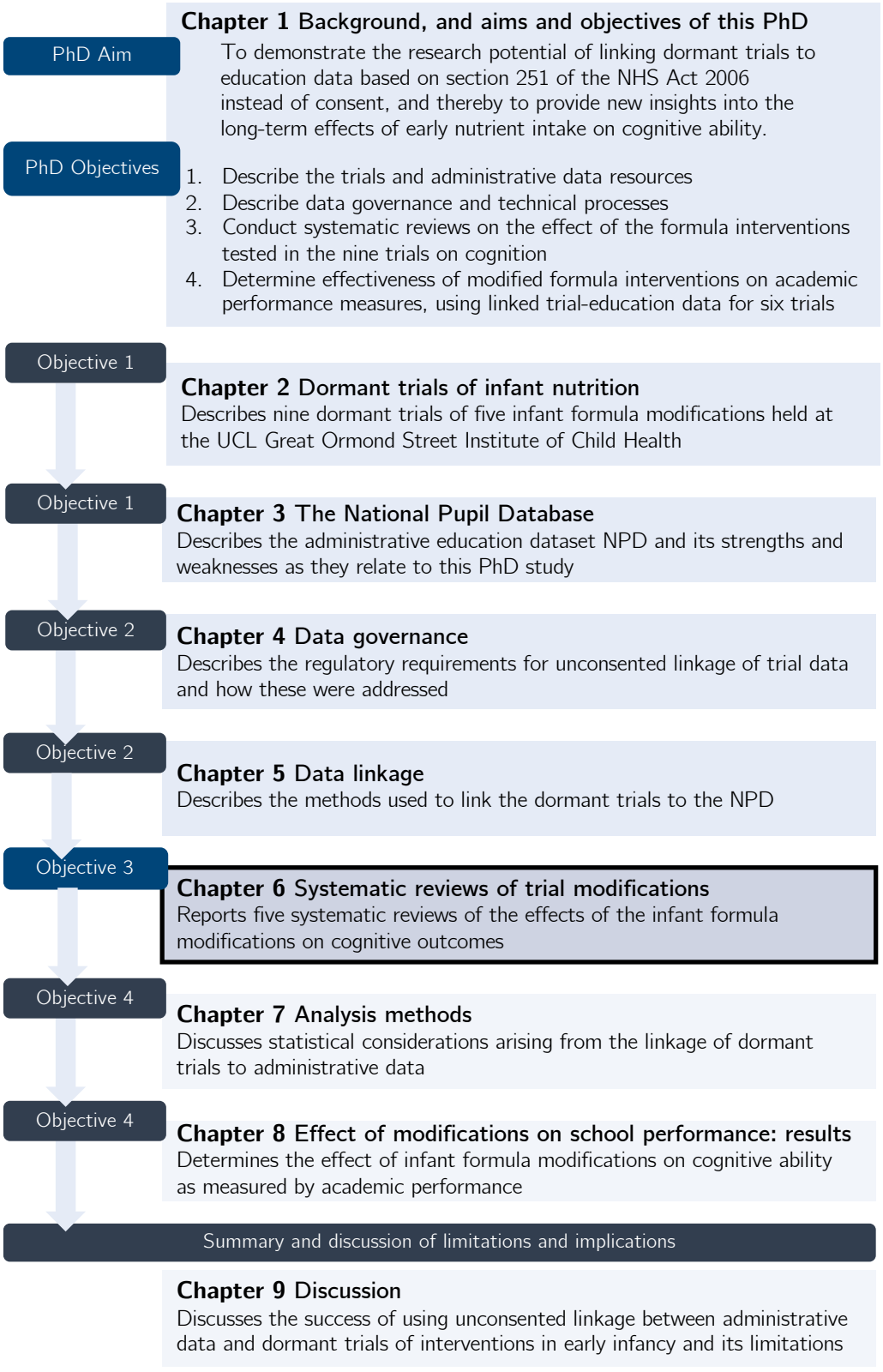
The most complex step was the derivation of match weights. Due to the lack of a gold-standard dataset, this step rested on several unverifiable assumptions. Instead, this thesis made use of the multiple possible candidate matches per participant together with indicators of linkage strength supplied by the FFT and prior probabilities from external datasets to calculate probabilistic match weights and identify characteristics of plausible and implausible links. While this method does not allow the direct evaluation of linkage error, it allowed informed approximations of linkage error, while minimising personal data disclosure risk. In terms of their observed characteristics, linked participants were representative of the randomised samples, leaving me to assume that linkage did not introduce substantial selection bias.

Overall, the different linkage steps presented in this chapter emphasise the need to comprehensively evaluate linkage quality, taking into account the specific characteristics of the trial datasets involved. As different stages of linkage can be conducted by different institutions, careful coordination between data linkers and data users is required to balance transparency and participant confidentiality.

5.7 Key points from Chapter 5

- The digitisation of paper-based participant identification data can involve a significant amount of time and personnel investment, especially when identifiers are handwritten.
- The linkage process was conducted in several steps and involved an external organisation (FFT) in preserving participant confidentiality. FFT supplied me with up to 4 pupil matches per participant as well as information on identifier agreement strength for each participant-pupil pair. Without having access to the identifiers themselves, I completed the linkage using a combination of deterministic and probabilistic methods, making use of clinical information, and information on agreement strength between participant and pupil identifiers.
- Within the analysis trials, 92% of all randomised participants were successfully linked to the NPD; and 86% of all randomised participants from the analysis trials were successfully linked to the NPD module that contains the primary outcome.
- The linkage process described in this chapter can provide a generalisable guide for the linkage of dormant trials to similar datasets.

The next chapter, Chapter 6, addresses objective 3: "*to conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition*".


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 6 What is already known about the cognitive effects of the trial interventions? Five systematic reviews

A section of this chapter resulted in a peer-reviewed paper, published here: https://doi.org/10.1371/journal.pone.0241800

Verfuerden ML, Dib S, Jerrim J, Fewtrell M, Gilbert RE (2020) Effect of long-chain polyunsaturated fatty acids in infant formula on long-term cognitive function in childhood: A systematic review and meta-analysis of randomised controlled trials. PLOS ONE 15(11)

6.1 Chapter structure and content

In order to understand the broader evidence context around the formula modifications presented in Chapter 2, this chapter systematically reviews and critically assesses previously published and unpublished literature on cognitive effects of infant formula. The chapter is divided into the following sections: five systematic reviews of randomised controlled trials (one for each modification: nutrient-enriched formula post-discharge for preterm infants or from birth in term SGA infants, and supplementation of infant formula with either LCPUFA, iron, nucleotides, or sn-2 palmitate) and a discussion of common themes. I argue that, so far, there is no conclusive evidence that the discussed infant formula modifications benefit cognition and that the interpretation of the available evidence on long-term effects is particularly uncertain due to high participant attrition over time.

6.2 Review 1: Effect of nutrient-enriched formula for preterm infants and term SGA infants on cognitive outcomes

6.2.1 Dormant trials this review aims to provide context for:

- NEP-1 trial
- NEP-2 trial
- NEP-PD trial
- NETSGA trial

6.2.2 Background

Babies born preterm and babies born small-for-gestational-age (SGA) (which can be preterm or term) are smaller and lighter and have fewer nutrient reserves than healthy term-born babies. Nutrient-enriched formula has been hypothesised to benefit both groups. For preterm babies (SGA or appropriate for gestational age, AGA), the justification for enrichment is in meeting their increased nutrient requirements, while term SGA babies are given nutrient-enriched formula to support catch-up growth. Previous studies (Hack et al., 1991, Cooke and Foulder-Hughes, 2003, Leppänen et al., 2014) suggested that preterm and term SGA infants who still have growth deficits by the end of their first year of life are more likely to have poor cognitive and educational outcomes in childhood compared to children without growth deficit.

Nutrient enrichment of infant formula for preterm infants and term SGA infants has been associated with improved short-term growth and improved developmental outcomes, particularly in boys (Lucas et al., 1998, Kumar et al., 2017, Isaacs et al., 2009). However, some studies have reported that benefits to cognition observed in term SGA babies might occur at the expense of increased metabolic risks (Castanys-Muñoz et al., 2017). In addition, two recent Cochrane reviews have compared the effect of enriched versus standard formula in preterm infants and found no conclusive evidence of cognitive benefit (Young et al., 2016, Walsh et al., 2019). None of these reviews have taken into account cognitive outcomes after 18 months of age, however. This systematic review and meta-

analysis examines published and unpublished data. The aim was to see whether there is evidence that feeding nutrient-enriched formula after birth reduces the risk of poor cognitive outcomes for preterm and term SGA infants compared to feeding standard term formula. This review includes interventions during both the pre-and post-discharge period. While it is possible that intervening during these different time periods could potentially have different effects on outcome since the brain is growing most rapidly in the period prior to term, subgroup analyses were beyond the scope of this review. Similarly, previous studies suggested a possible difference in effectiveness in boys vs girls. Assessing sex-specific effects was also beyond the scope of this review.

6.2.3 Methods

Objective	To determine the effects of nutrient-enriched infant formula vs standard formula on cognitive ability.
Population	Preterm infants, defined as being born $<$ 37 weeks gestation or infants born with low birth weight ($<$ 2.5 kg) or small-for-gestational-age (SGA) defined as birth weight $<$ 10th centile for their gestational age.
Types of studies	Randomised controlled trials in humans with at least six weeks of follow-up.
Types of interventions	Post-discharge enriched formula: energy content ≥72 and ≤75 kcal/100 ml, protein content >1.7g/100ml OR Preterm In-hospital enriched formula: energy content >75 kcal/100 ml, protein content >2/100ml compared to standard term formula: energy content <72 kcal/100 ml, protein content ≤1.7g/100ml
Types of outcomes	Bayley Scores of Mental Development, IQ Scores or similar standardized measures of cognitive ability or measures of academic achievement.
Search strategy	I searched MEDLINE and EMBASE up to Sept 2020, without language restrictions (search terms specified in appendix p 42). Data were entered (with a 10% random sample double entered by my colleague Kathy Kennedy) into a RedCap form modelled on the Cochrane Risk of Bias Tool 2. I included all studies that compared enriched infant formula with standard formula.

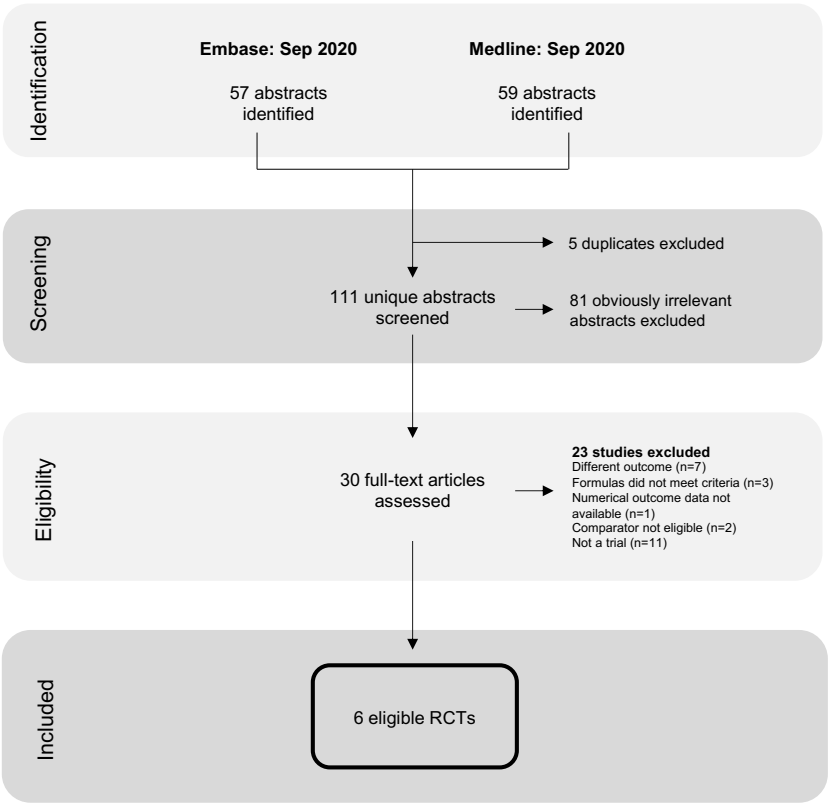


Fig 6.1: Study selection: nutrient-enriched infant formula

6.2.4 Results

6.2.4.1 Included studies

The search identified 57 abstracts on Medline and 59 articles in Embase. After screening abstracts and excluding duplicates, I reviewed 30 full-text publications for detailed assessment. Six trials (Lucas et al., 1998, Cooke et al., 2001, Lucas et al., 2001, Morley et al., 2004, Jeon et al., 2011, Ruys et al., 2018) met the review criteria (**Fig 6.1**, **Table 6.1**). The trials were in preterm infants (both AGA and SGA) and term SGA infants. The included trials were conducted in the UK (Lucas et al., 1998, Cooke et al., 2001, Lucas et al., 2001, Morley et al., 2004), South Korea (Jeon et al., 2011), and the Netherlands (Ruys et al., 2018).

The six included RCTs assessed cognitive ability of 704[‡‡] eligible participants, of which 354 (50%) received the enriched formula preparation. All trials assessed participants' development with the Bayley Scales of Infant Development Mental Development Index (BSID MDI); three studies additionally assessed Wechsler IQ score (Lucas et al., 1998, Fewtrell et al., 2001, Ruys et al., 2018).

[‡‡] One trial (Cooke 2001) did not report how many participants were cognitively assessed. For this trial the randomised participants are counted which is likely to be an overestimate.

Table 6.1: Characteristics of trials included in review for nutrient-enriched formula

Publication and trial centres	Place and years of randomisation	Population	Duration	Cognitive outcomes	Followed-up/ Randomised
(Lucas et al., 1989)	*Norwich, Sheffield* 1982-1984	<1850g bw, <33 wga (preterm)	From birth for 4 weeks	BSID MDI II at 18m CA	M: 168/211 S: 166/213
Lucas et al. (1998)				IQ at 7.5-8 years	M: 385/462 S: 380/464
Cooke et al. (2001)	*Newcastle upon Tyne* unclear	≤1750g bw, ≤34 wga (preterm)	Term for 6 months CA (post-discharge formula)	BSID MDI II at 18m CA	M: ?/30 S:?/31
Lucas et al. (2001)	*Cambridge, Leicester, Nottingham* 1993-1995	<1750g bw, <37 wga (preterm)	Term until age 9 months CA (post-discharge formula)	BSID MDI II at 18m CA	M: 105/ 113 S: 102/ 116
Morley et al. (2004)	*Cambridge, Leicester, Nottingham* 1993-1995	bw < 10th centile ≥37 wga (term SGA)	Birth until age 9 months	BSID MDI II at 18m CA	M: 113/ 152 S: 122/ 147
Unpublished				IQ at 16 years	M: 14/ 152 S: 14/ 147
Jeon et al. (2011)	*Seoul* Unclear	<1500g bw, <33 wga (preterm)	Term for 6 months (post-discharge formula)	BSID MDI II at 18m CA	M: 19/ 34 S: 21/ 35
(Ruys et al., 2018)	*Amsterdam* 2003-2006	<1500g bw, <33 wga (preterm)	Term to age 6 months CA (post-discharge formula)	BSID MDI II at 24m CA	M: 45/ 54 S: 35/ 48
				IQ at 8 years	M: 29/ 54 S: 14/ 48

M=modified formula group, S=standard formula group, m=months, wga= weeks gestational age, bw=birth weight; BSID MDI: Bayley Score of Mental Development Index; blue highlight: dataset discussed in detail in chapter 2 and available for linkage with routine education data

6.2.4.2 Quality of evidence

In all studies except for Lucas et al. (2001) and Morley et al. (2004) attrition of participants (as measured at the latest available cognitive outcome), was high (**Fig. 6.2**); in Cooke and Foulder-Hughes (2003) there was no information about the number of participants followed-up. For Jeon et al. (2011) allocation and randomisation circumstances were unclear except that the authors state: "*Because the number of enrolled infants was not stratified according to hospital,*

inter-hospital variations might have confounded the outcome". Furthermore, substantial between-group differences in baseline variables existed, and there were several possible outcome scales ("raw score" and "index score") without indication which one was the prespecified primary outcome scale. In Ruys et al. (2018), allocation concealment was unclear; attrition was substantially related to randomised group, and there were several possible cognitive outcome scales (according to corrected age and chronological age).

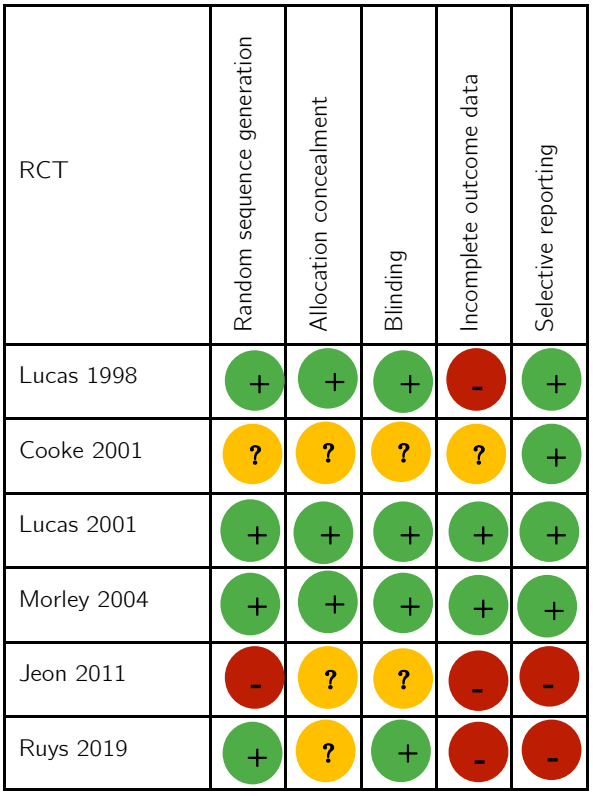


Fig. 6.2: Risk of bias assessment: nutrient-enriched infant formula

6.2.4.3 Effect on cognitive ability

Fig. 6.3 shows the pooled standardised mean difference in Bayley Mental Development scores from six RCTs of preterm and SGA term babies. There was no evidence of benefit, with the mean difference close to 0 and the 95% Confidence interval consistent with both benefits and harms (Standardised mean

difference: -0.08 SD, 95% CI: -0.38, 0.21). Heterogeneity was high (I^2=84%)[§§]. Three RCTs reported measures of Wechsler IQ score at age 7-15 years, showing no difference in standardised mean differences (0.05, 95% CI: -0.08, 0.19), and low heterogeneity (I^2=0%), **Fig. 6.4**.

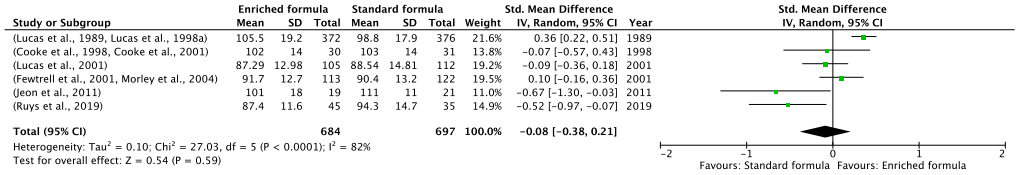


Fig. 6.3: Enriched vs standard formula: BSID MDI at 18 months

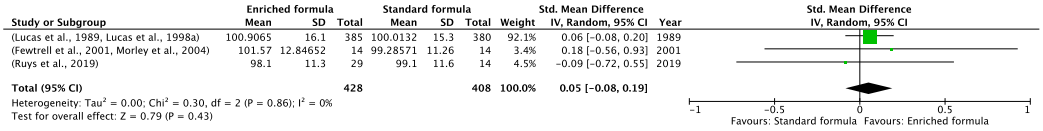


Fig. 6.4: Enriched vs standard formula: IQ Score at 7-15 years

6.2.5 Conclusion

Data from 6 RCTs with a total of 704 participants found no evidence of benefit or harm of nutrient-enriched infant formula for premature infants on cognitive ability, however quality of evidence was low. It is important to note, that this finding does not exclude a role for this intervention in treating or preventing the significant undernutrition often seen in preterm infants at the time of hospital discharge and beyond. Future research should focus on sex-specific effects and determine whether effects are modified by the timing of the intervention (pre- vs post-discharge period) by using individual patient meta-analyses.

[§§] When Jeon 2011 (where risk of bias was judged to be high) is excluded it does not change this conclusion

6.3 Review 2: Effect of LCPUFA-fortified formulas on cognitive outcomes

6.3.1 Dormant trials this review aims to provide context for:

- LCPUFAP trial
- LCPUFAT trial

6.3.2 Background

Lack of preformed long-chain polyunsaturated fatty acids (LCPUFA) in infant formula has been hypothesised as contributing to the finding of reduced IQ scores in formula-fed compared with breastfed infants (Lucas et al., 1992, Morrow-Tlucak et al., 1988, Anderson et al., 1999, Kramer et al., 2008). The EU Commission recently mandated the addition of one type of LCPUFA, *DHA*, to all infant and follow-on formulas (EU Commission, 2016). However, research on cognitive benefits to date has been inconclusive (Jasani et al., 2017, Moon et al., 2017), and mandatory supplementation may result in price rises across the market (Hughes et al., 2017). The present review combines published and unpublished trial data, acquired through contacting trial authors, to compare the cognitive effects of LCPUFA-fortified versus unfortified infant formula in children born at term and preterm.

6.3.3 Methods

Objective	To compare the long-term cognitive effects of LCPUFA-fortified versus unfortified infant formula in children born at term and preterm.
Population	Babies born at term or preterm
Types of studies	Randomised controlled trials in humans with at least 1 month of follow-up.
Types of interventions	Infant formula fortified with LCPUFA (DHA alone or DHA together with AA, at any dose) vs unfortified control formula
Types of outcomes	Bayley Scores of Infant Development, IQ Scores or similar standardized measures of cognitive ability, and measures of academic achievement.
Search strategy	I searched Medline, Embase, proceedings from major scientific meetings of child nutrition and the Cochrane Central Register of Controlled Trials up to September 2020, without date or language restrictions. I reviewed the reference lists of the included studies and traced subsequent publications. I first identified RCT participant cohorts based on any infant formula supplementation with LCPUFA, independent of whether cognitive outcomes were reported. I then contacted a total of 18 trialists, ethics committees or industry representatives to identify potential unpublished data, clarify study details and to ask whether they knew of any other eligible trials that had measured cognitive outcomes. Data were entered (with a 10% random sample double entered by my colleague Sarah Dib) into a RedCap form modelled on the Cochrane Risk of Bias Tool 2.
Data analysis and risk of bias assessment	I performed separate analyses for term and preterm-born participants because healthy term infants can synthesise LCPUFA from fatty acid precursors, whereas preterm babies are born with fewer LCPUFA reserves accumulated in utero and are less able to synthesise LCPUFA than term-born babies. All analyses were performed in RevMan v5.4 and included participants with the relevant outcome in the groups to which they were randomised. Risk of bias was assessed using the Cochrane Risk of Bias Tool 2.

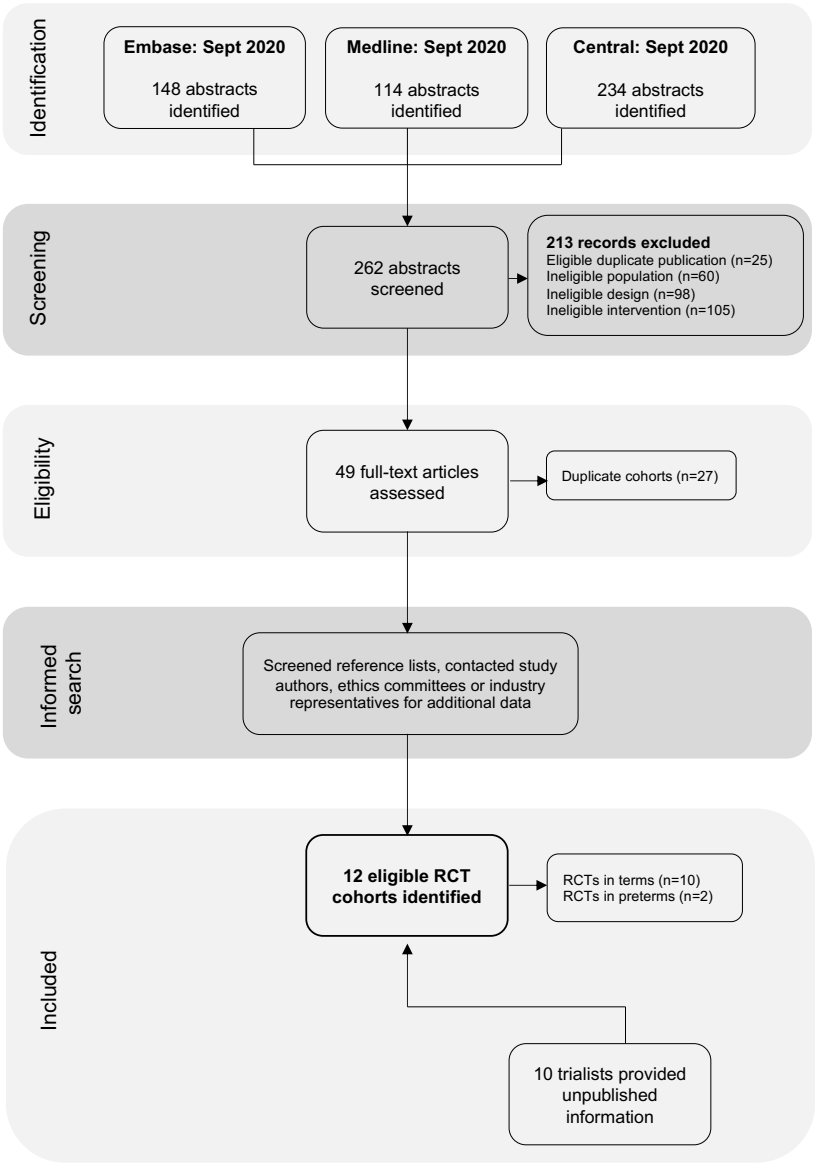


Fig. 6.5: Study selection: LCPUFA and cognitive ability

6.3.4 Results

6.3.4.1 Included studies

12 trial cohorts were included, 10 of which randomised infants born at term and two infants born preterm (**Fig. 6.5**). I obtained previously unpublished outcome data for two RCTs: 1) a two-centre trial of term babies in England (Lucas et al., unpublished), where investigators provided unpublished follow-up data on IQ

assessments using the Wechsler Preschool and Primary Scales of Intelligence-Revised (WPPSI-R) at age 4.5 years and the Wechsler Abbreviated Scales of Intelligence (WASI) at age 16 years.; 2), a two-centre trial in England of babies born preterm, where investigators provided unpublished data from their IQ assessments using the WASI at age 16 years (Fewtrell et al., 2002). I also received partly published outcome data from 3 trials: IQ using the WASI at 9 years from a Dutch trial of term babies (de Jong et al., 2015), which was previously published with an interaction term only, and IQ using the WASI at age 16 years from the Kansas centre of the US based DIAMOND trial conducted in term infants, which previously was available only as a figure (Colombo et al., 2017); as well as IQ assessments adjusted for maternal education at ten years from a two-centre study in Scottish preterm children (Isaacs et al., 2011). **Table 6.2** shows the characteristics of all included trial cohorts. The total number of children randomised was not available for one RCT (Willatts et al., 2013), despite contacting several trial investigators and the relevant ethics committee. All studies randomised participants to infant formula fortified with LCPUFA (DHA with or without AA) or to unfortified infant formula. DHA was sourced from egg, fish, fungi, algae, or starflower oil and made up between 0.12 and 0.96% of total fat content. Ratios of DHA to AA ranged from 1:0.8 to 1:3.6. The LCPUFA content in Ben et al. (2004b) was not clear. Duration of the intervention ranged from 2 to 12 months in term infants and 3 weeks to 9 months in infants born preterm. Children in one preterm trial (Isaacs et al., 2011) could receive some breastmilk during the first months, all others were exclusively formula-fed.

Three RCTs had more than one randomised intervention group (Colombo et al., 2017, Auestad et al., 2003, Birch et al., 2007). To minimise heterogeneity, I included only the intervention group in my analysis that was most similar in DHA dose and source to the other included RCTs. Colombo et al. (2017) randomised babies in two centres and then conducted different cognitive assessments at different ages, with follow-up stratified by centre. As these were not at risk of double-counting participants, I regarded these as independent and included both reported cognitive assessments in my analysis.

Table 6.2: Characteristics of trials included in review for LCPUFA fortified formula (continues on next page)

Latest publication and trial centres	Place of randomisation	Population	Modification	Outcomes	Follow-up/ randomised[1]
Term studies:		wga			
(Scott et al., 1998) Auestad et al. (2003)	Portland, Kansas, Seattle	>36	DHA 0.12 + AA 0.43 for 12m	BSID MDI age 1y PPVT age 3.3y SB IQ 3.3y	91/91* 72/91
Auestad et al. (2001)	Missouri, Arkansas, Pennsylvania Arizona	37–42	DHA (0.13%) AA (0.45%) for 12m	BSID MDI age 1y	57/159
Makrides et al. (1995)	Adelaide	37–42	DHA (0.35%) for 7m	BSID MDI age 1y	28/32
Makrides et al. (1999)	Adelaide	37–42	DHA (0.34%) AA (0.34%) for 4m	BSID MDI age 2y	38/56
Ben et al. (2004b)	Nanjing	37–40	unclear for 6m	BSID MDI age 6m	41/121
Willatts et al. (2013)	Milan, Birmingham, Dundee, Leuven, 2 unknown	37-42	DHA 0.30[§]+ AA 0.44[§] for 4m	WPPSI-R age 6y	147/ n/a**
(Lucas et al., 1999) Unpublished Unpublished	Nottingham, Leicester	>36	DHA 0.32 + AA 0.30 for 6m	BSID MDI age 1y WPPSI-R age 4.5y WASI age 16y	234/309 184/309 41/309
(Birch et al., 2000) Birch et al. (2007)	Dallas	37-40	DHA 0.36 + AA 0.72 for 3.9m	BSID MDI age 1y WPPSI-R age 4y	39/53 36/53
(Bouwstra et al., 2005) de Jong et al. (2015)	Groningen	37-42	DHA 0.30 + AA 0.45 for 2m	BSID MDI age 1y WASI age 9y	290/315 214/315
Colombo et al. (2017) + unpublished data (Drover et al., 2011) (Drover et al., 2012)	Kansas Dallas	37-42	DHA 0.32 + AA 0.64 for 4m	WPPSI-R age 6y (numbers unpublished) BSID MDI age 1y PPVT age 3.3y	42/92 57/80 30/80

Preterm studies:		bw, wga			
Fewtrell et al. (2002)	Nottingham, Leicester	<1750g , <37	DHA 0.32 + AA 0.64 for 0.69m	BSID MDI age 1y	158/196
Unpublished				WASI age 16y	17/196
(Fewtrell et al., 2004)	Glasgow	≤2000g , <35	DHA 0.17 + AA 0.31 for 9m	BSID MDI age 1y	199/238
Isaacs et al. (2011)				WASI age 10y	107/238

Footnotes:
bw: birth weight, wga: weeks gestational age, m: months; PPVT: Peabody Picture Vocabulary Test, SB IQ Stanford-Binet IQ, BBCS-R: Bracken Basic Concept Scale-Revised, WASI: Wechsler Abbreviated Scale of Intelligence, WPPSI-R: Wechsler Preschool and Primary Scale of Intelligence-Revised; 1: Only dose/source of interest, some studies had more than one randomised dose/source group (only one per trial is included here);* unclear number initially randomised but overall number % lost was 23% §Two different concentrations were published: DHA= 0.30 (Agostoni et al., 1997) or 0.21 (Willatts et al., 2013) – AA= 0.44 (Agostoni et al., 1997) or 0.35 (Willatts et al., 2013)) ‡data was published in graphical form or differently modelled; n/a: not available; ** the 4 centres that were followed up randomised 237 infants between them but it is unknown how many were randomised in the remaining two centres and why they were not followed-up. It follows that the follow-up rate was < 62% blue highlight: dataset discussed in detail in chapter 2 and available for linkage with routine education data.

6.3.4.2 Quality of evidence

Fig. 6.6 presents the risk of bias assessment for the included trials. Overall quality of evidence was low. Completeness of follow-up for cognitive assessment was low, ranging from 9% to 93% of children initially randomised (median 63%). All trials were also included in the 2017 Cochrane review by Jasani et al. (2017), who rated the risk of bias lower but did not include a potential conflict of interest analysis (Roseman et al., 2012). While funding by industry was not classified as conflict of interest, several authors had financial interest through patents or were employed by industry (**Table 6.6** on page 155). Jasani et al. (2017) analysed the multi-centre trial by Willatts et al. (1998) and Agostoni et al. (1995) as two separate single-centre RCTs, while both studies have been part of the same multi-centre trial (Willatts et al., 2013) and therefore double-counted some participants. The primary outcome in Colombo et al. (2017) was not clear, and reported outcome data was likely to have been selected from multiple outcome measurements. Furthermore, there might be reason to question the authenticity of (Ben et al., 2004b). The study sample (including the number allocated to each group) seems identical to (Ben et al., 2004a), which investigated galacto-

oligosaccharides. Both publications have the same authors and were conducted at the same time in the same study location.

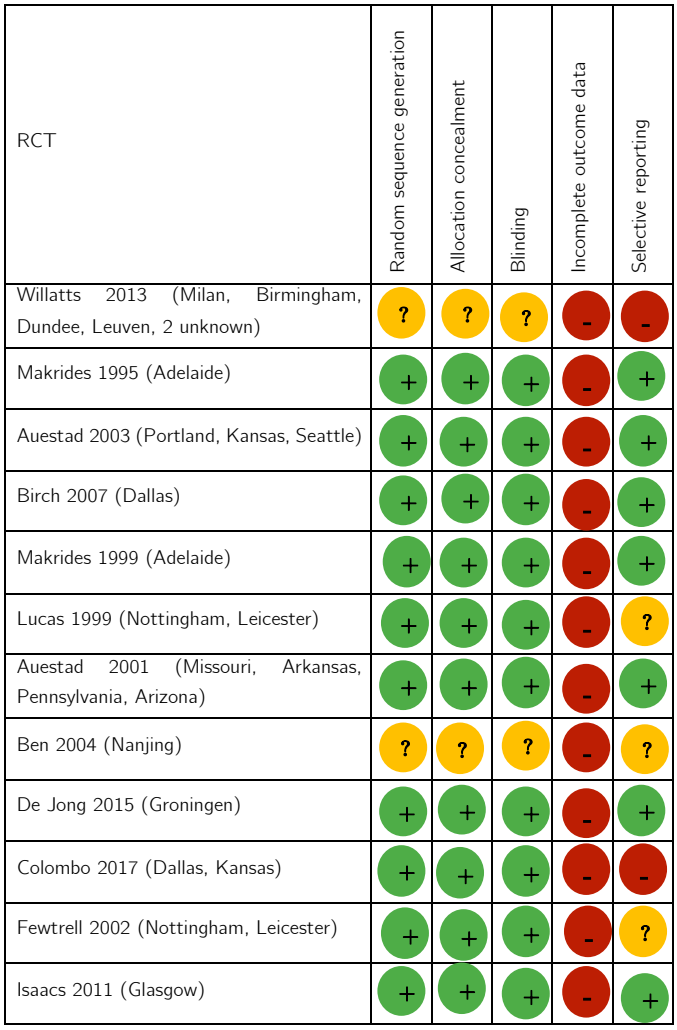


Fig. 6.6: Risk of bias assessment: LCPUFA fortified formula

6.3.4.3 Effect on cognitive ability

Participants randomised to formula with LCPUFA did not differ from infants randomised to unfortified formula in any of the cognitive outcomes in term (**Fig. 6.7**, **Fig. 6.9**, **Fig. 6.10, Fig. 6.11** and **Fig. 6.12**) and in preterm infants in a random effects meta-analysis (**Fig. 6.8**). There was also no evidence of benefit when the latest available cognitive measures from each trial cohort were pooled

together: data from 9 RCTs showed a -0.16 standard deviation reduction in the mean difference, 95% confidence interval: -0.32, 0.00 (**Fig. 6.13**).

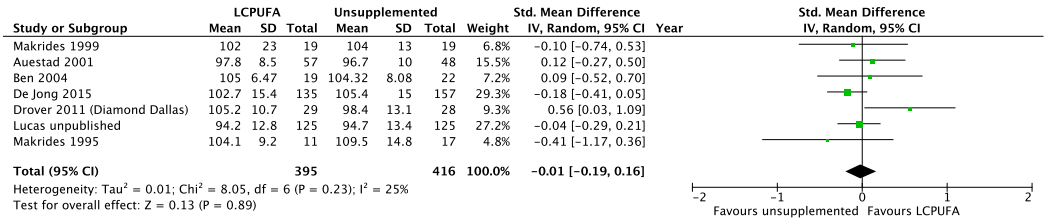


Fig. 6.7: LCPUFA vs standard: BSID MDI at 1-2 years (term infants)

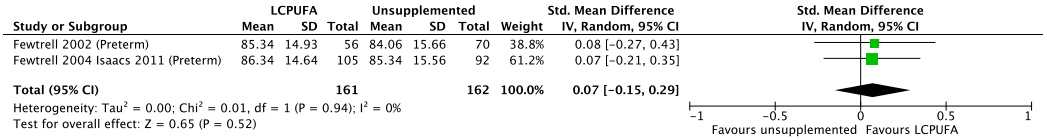


Fig. 6.8: LCPUFA vs standard: BSID MDI at 1-2 years (preterm infants)

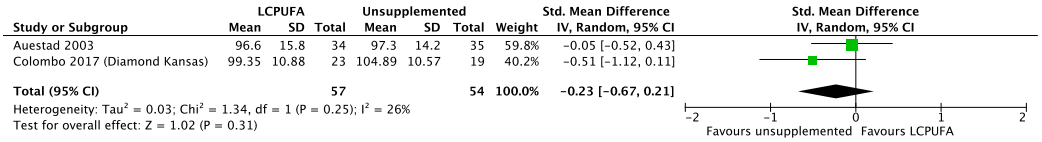


Fig. 6.9: LCPUFA vs standard formula: PPVT at 2-4 years (term infants)

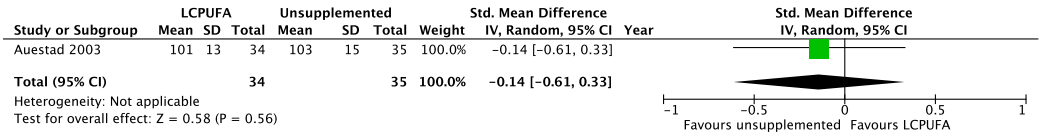


Fig. 6.10: LCPUFA vs standard: Stanford Binet IQ at 3.3 y (terms)

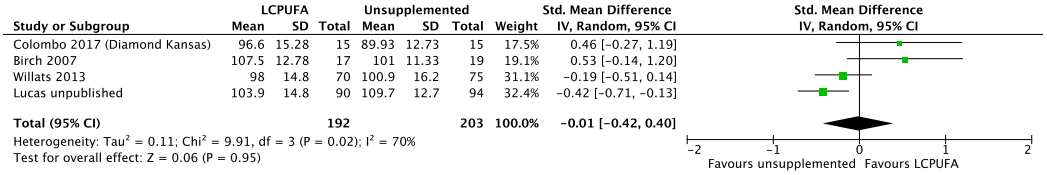


Fig. 6.11: LCPUFA vs standard: WPPSI IQ at 4-9 y (terms)

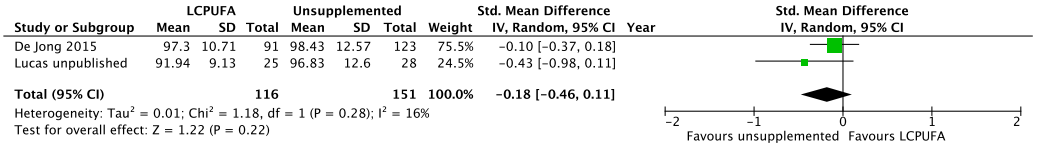


Fig. 6.12: LCPUFA vs standard: WASI IQ at 6-16 y (terms)

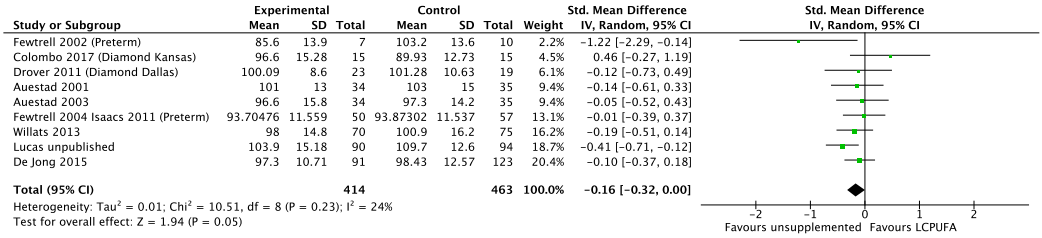


Fig. 6.13: LCPUFA vs standard: latest available cognitive outcome

6.3.5 Conclusion

Data from 12 trial cohorts with a total of 1,888 participants suggests that LCPUFA-fortified infant formula does not improve cognition among children born at term or preterm. Effect estimates were uncertain, and 95% confidence intervals included potential for meaningful benefit but also harm.

6.4 Review 3: Effect of infant formulas with added nucleotides on cognitive outcomes

6.4.1 Dormant trials this review aims to provide context for:

- NUCLEO trial

6.4.2 Background

Nucleotides are the building blocks of RNA and DNA and they play a role in energy transfer processes, the metabolism of macronutrients as well as in increasing the bioavailability of iron (Yu, 1998, EFSA Panel on Dietetic Products and Allergies, 2014). Nucleotides are dispensable nutrients that can be synthesised de-novo in the human body, but they are also present in human breast milk (Gil and Sanchez-Medina, 1982, Liao et al., 2011). Under Directive 2006/141/EC nucleotides may be added to infant formula on a voluntary basis but are not mandatory.

Currently there is doubt about whether nucleotides in human milk fulfil a specific function for babies or whether they are by-products of human milk formation (EFSA Panel on Dietetic Products and Allergies, 2014). Previous studies have investigated effects on immune function but found inconsistent results (Pickering et al., 1998, Yau et al., 2003, Schaller et al., 2004, Carver et al., 1991, Hawkes et al., 2005). They also assessed effects on stool quality and growth, and found no evidence of benefit on either outcome (Singhal et al., 2008, Brunser et al., 1994, Singhal et al., 2010a, Schaller et al., 2004, Yau et al., 2003, Pickering et al., 1998); as well as occipitofrontal head circumference (OFC) where positive effects were reported in 2/3 RCTs (Hawkes et al., 2005, Cosgrove et al., 1996, Singhal et al., 2010a). No review has previously assessed the effects of nucleotide-fortification of infant formula on more direct measures of cognitive ability. In this section, I systematically assessed the available RCT evidence on cognitive effects of nucleotide-fortified versus unfortified infant formula in healthy babies.

6.4.3 Methods

Objective	The overall aim was to review the scientific literature on effects of infant and follow-on formula with added nucleotides on cognitive ability or educational outcomes.
Population	Healthy babies
Types of studies	Randomised controlled trials in humans with at least six weeks of follow-up.
Types of interventions	Nucleotide-fortified infant or follow-on formula, with levels $>$ 30 mg/ Litre, commenced within the first year after birth.
Types of outcomes	Bayley Scores of Infant Development, IQ Scores or similar standardized measures of cognitive ability, or academic performance.
Search strategy	I searched MEDLINE and EMBASE up to Sept 2020, limited to randomised controlled trials in humans without language restrictions, using the terms specified in appendix p 42. Data were entered (with a 10% random sample double entered by my colleague Kathy Kennedy) into a RedCap form modelled on the Cochrane Risk of Bias Tool 2. I included all studies that started the intervention in the first month of life and compared enriched formula with standard formula and reported cognitive assessments or academic achievement.
Data analysis and risk of bias assessment	All analyses were performed in RevMan v5.4 and included participants with the relevant outcome in the groups to which they were randomised. Risk of bias was assessed using the Cochrane Risk of Bias Tool 2.

6.4.4 Results

The search yielded 16 unique abstracts but no eligible trials. None of the studies identified in the search reported direct cognitive outcome measurements.

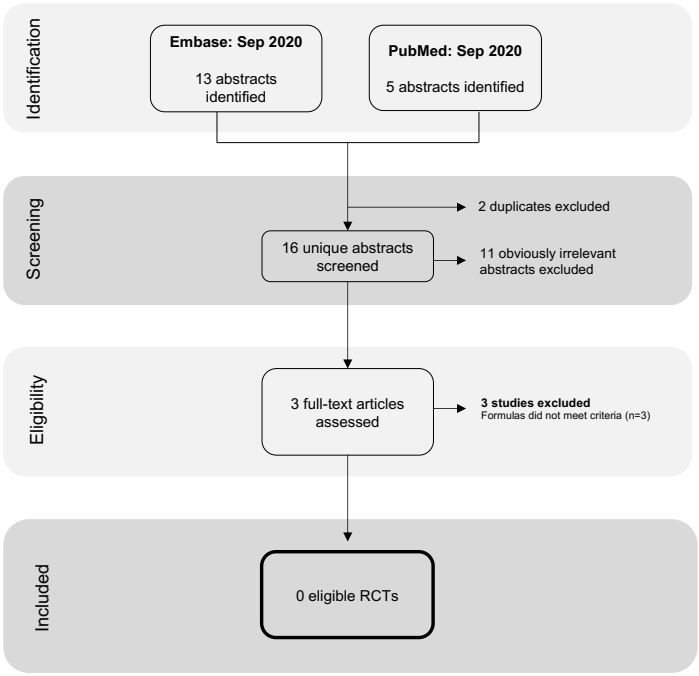


Fig. 6.14: Study selection: nucleotides and cognitive ability

6.4.5 Conclusion

There was no published research on the effect of nucleotide-fortified infant or follow-on formula on cognitive development.

6.5 Review 4: Effect of iron fortified infant and follow-on formula on cognitive ability

6.5.1 Dormant trials this review aims to provide context for:

- IRON trial

6.5.2 Background

Infants need iron for brain development, growth, and development of the central nervous system (Georgieff, 2008). Iron stores in exclusively breastfed term infants are thought to be sufficient until up to six months of age (Domellöf et al., 2014, Domellöf et al., 2002). The WHO considers iron deficiency and iron deficiency anaemia in children to be of major concern as it is associated with adverse effects on cognitive development (World Health Organization, 2008). To prevent iron-deficiency and iron-deficiency anaemia in formula-fed infants, iron-fortified formula is given until age six months. Follow-on formula can be introduced alongside complementary feeding from the age of six months for both breast and formula-fed infants (Jonsdottir et al., 2012). It is not clear whether iron-fortified formula is risk free in iron-replete infants (infants with sufficient iron stores). Walter et al. (1998) randomised 835 six-month-old infants to receive either six months of low iron follow-on formula (2.3 mg/L or 0.35 mg/100 kcal) or high-iron follow-on formula (12.7 mg/L or 1.95 mg/100 kcal). Consumption of low iron follow-on formula led to a significantly lower iron status than formula with higher iron content. However, there were no adverse effects on cognitive outcomes in the low-iron group. In follow-ups of the study, at ten years (Lozoff et al., 2012) and 16 years (Gahagan et al., 2019), cognitive scores were -contrary to expectation- lower in the group who had received the high-iron formula. However, effects were small, and 58% and 68% of participants were lost to follow-up, respectively, making the interpretation of the results uncertain.

In this section, I systematically assess the available evidence on cognitive effects of iron fortification vs low or no iron fortification in infants who did not have iron deficiency anaemia or whose iron status was unknown.

6.5.3 Methods

Objective	The overall aim was to review the scientific literature on effects of enteral*** iron fortification vs low or no iron content on cognitive ability or academic outcomes.
Population	Babies born at term, not having, or not tested for iron deficiency anaemia
Types of interventions	Non-intermittent enteral iron-supplementation (in infant formula, follow-on formula, or syrup), with levels $>$ 1 mg iron/ 100 ml as ferrous sulphate, commenced within the first year after birth for a minimum duration of 3 months vs unfortified or low-iron preparations.
Types of outcomes	Bayley Scores of Infant Development, IQ Scores or similar standardized measures of cognitive ability, or academic performance.
Search strategy	I searched MEDLINE and EMBASE up to Sept 2020, limited to randomised controlled trials in humans without language restrictions, using the terms specified in appendix p 42. Data were entered (with a 10% random sample double entered by my colleague Kathy Kennedy) into a RedCap form modelled on the Cochrane Risk of Bias Tool 2. I included all studies that started the intervention in the first year of life and reported cognitive assessments or academic achievement.
Data analysis and risk of bias assessment	All analyses were performed in RevMan v5.4 and included participants with the relevant outcome in the groups to which they were randomised. Risk of bias was assessed using the Cochrane Risk of Bias Tool 2.

*** via the gastrointestinal tract as opposed to e.g., tube feeding or intravenous administration

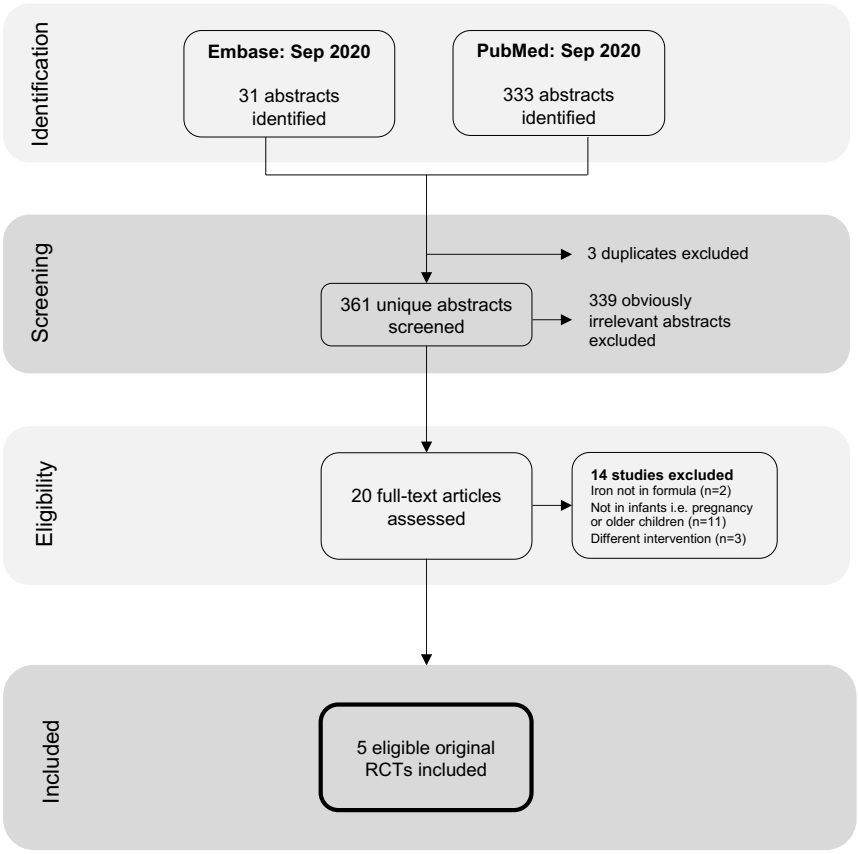


Fig. 6.15: Study selection: iron fortified formula

6.5.4 Results

The search identified 361 unique abstracts, of which I reviewed 20 studies for detailed assessment. Five trials (Moffatt et al., 1994, Walter et al., 1998, Lozoff et al., 2012, Morley et al., 1999, Friel et al., 2003, Lind et al., 2004, Gahagan et al., 2019) met the review criteria. Reasons for exclusion of the other RCTs are stated in **Fig. 6.15**. The included trials were conducted in Canada (Moffatt et al., 1994, Friel et al., 2003), Chile (Walter et al., 1998, Lozoff et al., 2012, Gahagan et al., 2019), the UK (Morley et al., 1999) and Indonesia (Lind et al., 2004) and all recruited healthy children. Among them, they assessed 1,762 participants, of which 993 (56%) received the high-iron preparation (**Table 6.3**).

One trial used iron-fortified infant formula (Moffatt et al., 1994), two used iron-fortified follow-on formula (Morley et al., 1999, Lozoff et al., 2012,

Walter et al., 1998) and two used syrups (Friel et al., 2003, Lind et al., 2004) to administer the intervention. In the control groups, iron dose ranged from 0 to 0.9 mg/L. In the intervention groups, iron dose ranged from 12 to 12.8 mg/L in the formula trials and 7.5 to 10 mg/day for the syrup trials. All trials assessed the Bayley Scales mental development index (BSID MDI) between age 12-18 months. One trial (Lozoff et al., 2012, Walter et al., 1998, Gahagan et al., 2019) also assessed later IQ using the Wechsler Intelligence Scale for Children (WISC) at age ten, and academic performance (computational Maths) at age 16 years.

Table 6.3: Characteristics of trials included in review for iron-fortified follow-on formula

Latest publication and trial centres	Place and years of randomisation	Population	Modification	Outcomes	Follow-up/ randomised[1]
Moffatt et al. (1994)	Winnipeg (Canada) 1988-1992	healthy, very low-income families, bottle feeding, iron status unknown	12.8 vs 1.1 mg iron/L in infant formula from age 2 months for 13 months	BSID MDI at 15 months	M: 77/113 S: 77/112
(Lozoff et al., 2012)	Santiago (Chile) 1982-1985	healthy, bw ≥3kg, singletons, bottle feeding by age 6 months, subgroup where iron status was unknown	12.7 vs 2.3 mg iron/L in follow-on formula from age 6 months to age 12 months	BSID MDI at 12 months	M: 430/576 S: 405/544
				Full Scale IQ at 10 years	M: 244/576 S: 229/544
Gahagan et al. (2019)				Academic performance at 16 years	M: 216/576 S: 189/544
Morley et al. (1999)	Leicester, Norwich, Nottingham (UK)	healthy term singletons, bw > 2500g	12 vs 0.9 mg iron/L in follow-on formula from age 9 months to 18 months	BSID MDI at 18 months	M: 133/162 S: 135/165
Friel et al. (2003)	St John's (Newfoundland) 1999-2000	healthy breastfed term infants	7.5mg/day vs placebo in syrup from age 1 to age for 6 months	BSID MDI at 13 months	M: 26/42 S: 20/35
Lind et al. (2004)	Purworejo (Indonesia) 1997-1999	healthy singletons	10 mg iron/day as ferrous sulphate vs placebo in syrup at age 6m to age 12 months	BSID MDI at 12 months	M: 163/170 S: 164/170

Footnotes: M=modified formula group, S=standard formula group, bw= birth weight, m=months, wga= weeks gestational age,; blue highlight: dataset discussed in detail in chapter 2 and available for linkage with routine education data blue highlight: dataset discussed in detail in chapter 2 and available for linkage with routine education data

I found no evidence of a significant benefit or harm of high-iron fortification on BSID MDI (the conclusion did not change when a fixed effect analysis was performed as a sensitivity analysis). There was no heterogeneity (I^2=0%, p=0.77), **Fig. 6.16**.

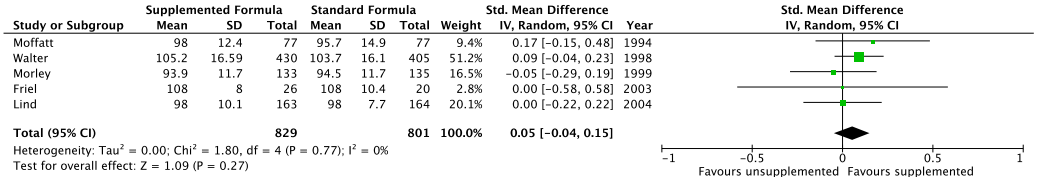


Fig. 6.16: High-iron vs low iron follow-on formula: Bayley Mental Development Index Scores

At age ten, participants in the high-iron group of Walter et al (Lozoff et al., 2012, Walter et al., 1998) had lower IQ scores compared to the low-iron group: iron group: 91.5 (SD 14.05) vs control: 93.3 (13.62), mean difference: -1.80 (p=0.217). This finding was replicated in a follow-up of the same study at age 16 years: Academic performance (computational Maths) was lower (MD-1.9, p=0.087) in the iron fortified group.

6.5.5 Quality of evidence

All studies except Morley et al. (1999) were judged to be at risk of bias (**Fig. 6.17**). Rates of follow-up were between 60% and 93% in the Bayley MDI scores and 42% and 36% for later scores. Incomplete outcome data was an issue for Moffatt et al. (1994), Gahagan et al. (2019) and Friel et al. (2003), with loss to follow-up >25%. In Gahagan et al. (2019) and Lind et al. (2004) allocation was not concealed, instead study personnel gave participating infants the next available formula number on a randomly generated list. In Friel et al. (2003) allocation concealment was unclear.

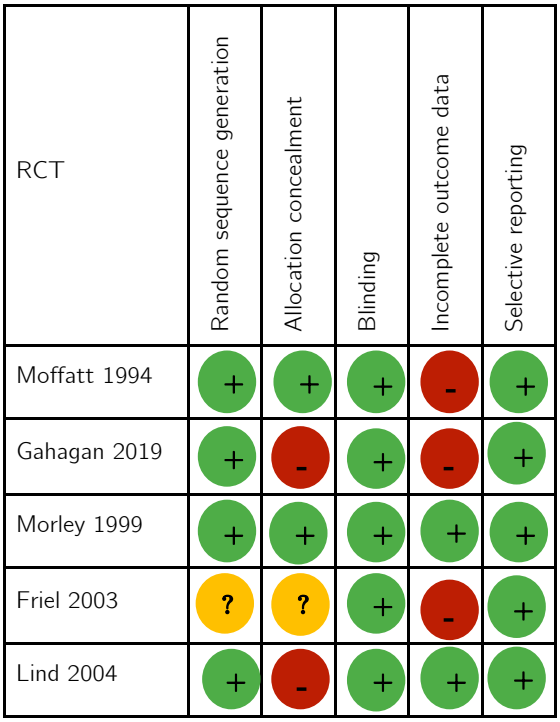


Fig. 6.17: Risk of bias assessment: iron fortified preparations

6.5.6 Conclusion

Iron fortification for babies without any symptoms or testing for iron deficiency anaemia showed no effect on Bayley MDI scores between ages 12-18 months. Follow-ups in childhood and adolescence showed that high-iron fortification of infant formula was associated with a lower IQ at age 10 years and lower arithmetic scores at age 16 years. However, interpretation of this data was uncertain due to high attrition rates. Data was too sparse to assess whether there was a difference based on the timing, duration or medium of fortification (infant formula vs follow-on formula vs syrup). More long-term studies with higher follow-up rates are needed to assess whether iron-fortification in infancy is risk-free in iron-replete infants or in infants whose iron status is unknown.

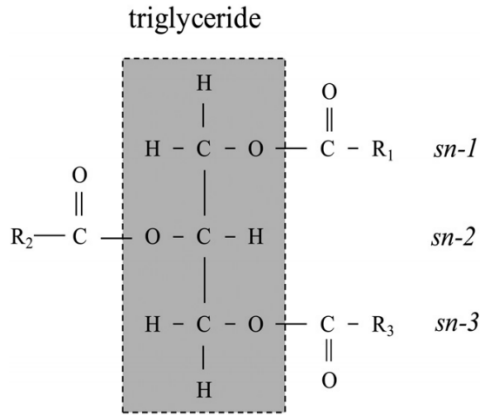
6.6 Review 5: Infant formula with high Sn-2 palmitate content and cognitive ability

6.6.1 Dormant trials this review aims to provide context for:

- PALM trial

6.6.2 Background

Palmitic acid is the most abundant saturated fatty acid in human breast milk (Innis, 2016). The arrangement of palmitic acid is unusual in human breast milk because most of it is bound to the sn-2 positions of triglycerides, compared to animal- and vegetable triglycerides, where fatty acids are usually attached to the sn-1 and sn-3 positions (**Fig. 6.18**). Some infant formulas contain sn-2 palmitic acid to facilitate the absorption of calcium, which is hypothesised to prevent constipation and increase bone strength, but effects on bone strength seem to be short-lasting (Zou et al., 2016, Miles and Calder, 2017, Bronsky et al., 2019). As the biological plausibility for cognitive effects in healthy infants who are unlikely to go through periods of negative energy balance is very weak, the trial that tested this modification (Kennedy et al., 1999) has been included in the dormant trials for linkage primarily to act as a negative control. However, evidence for an effect on cognition has not been reviewed before. Therefore, the aim of this section is to review evidence for potential cognitive effects of sn-2-palmitate used as a source of fat in infant formula.



From: Innis (2011) reprinted with permission from *Oxford University Press*.

Fig. 6.18: Structure of a triglyceride showing fatty acids esterified at the sn-1 (R1), sn-2 (R2), and sn-3 (R3) position

6.6.3 Methods

Objective	To assess whether the use of infant or follow-on formula fortified with palmitate mainly esterified at the sn-2 position has any impact on cognitive ability or academic outcomes when compared to formula fortified with palmitate mainly esterified at other positions.
Population	Preterm or term babies
Types of studies	Randomised controlled trials in humans with at least six weeks of follow-up.
Types of interventions	Infant formula with >40% of palmitic acids esterified at the sn-2 position vs standard infant formula with <20% of palmitic acids esterified at the sn-2 position
Types of outcomes	Bayley Scores of Infant Development, IQ Scores or similar standardized measures of cognitive ability, or academic performance.
Search strategy	I searched MEDLINE and EMBASE up to Sept 2020, limited to randomised controlled trials in humans without language restrictions, using the terms specified in appendix p 42. Data were entered (with a 10% random sample double entered by my colleague Kathy Kennedy) into a RedCap form modelled on the Cochrane Risk of Bias Tool 2. I included all studies that started the intervention in the first month of life and compared enriched formula with standard formula and reported cognitive assessments or academic achievement.
Data analysis and risk of bias assessment	All analyses were performed in RevMan v5.4 and included participants with the relevant outcome in the groups to which they were randomised. Risk of bias was assessed using the Cochrane Risk of Bias Tool 2.

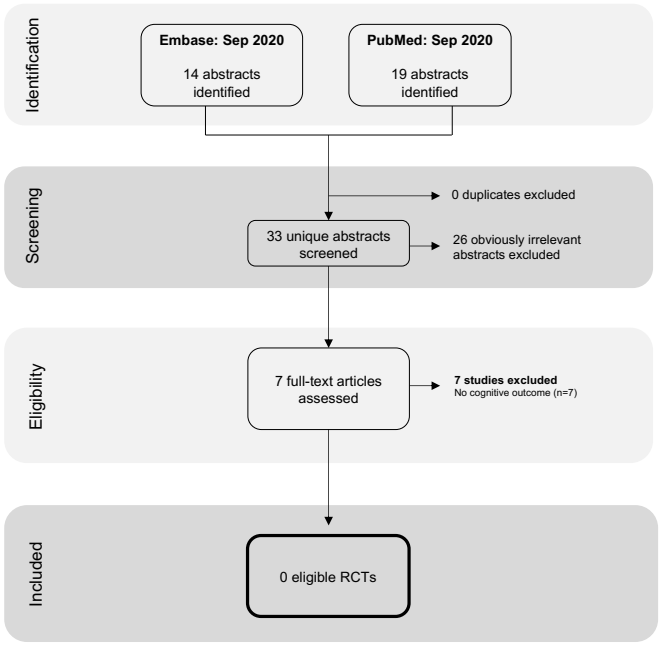


Fig. 6.19: Study selection: sn-2 palmitate

6.6.4 Results

The search yielded 33 unique abstracts, but no study assessed the effect of sn-2 palmitate enriched infant formula on cognitive ability (**Fig. 6.19**).

6.6.5 Conclusion

There is a lack of evidence on the cognitive effects of sn-2 palmitate in healthy infants. Given that there is no obvious mechanism through which sn-2 palmitate might affect cognitive ability, I assumed it is safe to use the PALM trial as a negative control trial in the analysis part of this thesis. More information on the concept of negative controls can be found in section 7.8 from page 178.

6.7 Summary

The above literature reviews show that there is either no or inconclusive evidence for cognitive benefits from infant formula that is nutrient-enriched, or fortified with LCPUFA, iron, nucleotides, or sn-2 palmitate. Evidence is of low quality due to unclear or absent reporting of key study design elements, short follow-ups, attrition in long-term follow-ups, and potential conflict of interest. These themes are discussed below (**Table 6.4**).

Table 6.4: Summary of evidence on cognitive effects

Formula modification	Effect direction	Quality of evidence
Nutrient enrichment	Inconclusive	Low
LCPUFA fortification	Inconclusive	Low
Nucleotide fortification	No data	n/a
Iron fortification	Inconclusive	Low
Sn-2 palmitate fortification	No data	n/a

6.7.1 Unclear or absent reporting of key study design elements

Most research was published more than two decades ago. This means that these publications are often not reported to the standards we now expect, such as consort flow diagrams, referring to pre-published analysis plans, and including clear descriptions of randomisation, allocation concealment, and blinding. This does not automatically mean that the results are also unreliable, but it severely limits the ability to grade the quality of the available evidence.

6.7.2 Insufficient follow-up time to measure differences in the outcome of interest

Results from early cognitive measures, such as Bayley scores, which are based on observing infant behaviour around the age of 1 or 2 years, are inherently dependent on the context in which measurement takes place. This can make them susceptible to bias. For instance, scores assigned by the outcome assessor may be influenced by the assessor's impression of the education or social standing of the accompanying parent's or even by the degree of cooperativeness that the

infant shows on that specific day. Early developmental measures are also not validated to predict later cognitive outcomes (Sun et al., 2015).

6.7.3 Participant drop-out is universal in long-term follow-ups

In contrast to early developmental outcomes, assessments at a later age, such as IQ scores, are likely to reflect underlying differences in cognitive ability more easily (Sun et al., 2015). However, high participant drop-out rates (attrition) mean that such outcomes are often not readily available or difficult to interpret. **Table 6.5** indicates that the loss of participants seems to be a universal problem across the studies investigated in this chapter. It also visualises the high degree of uncertainty and inconsistency around cognitive effects for each formula modification.

Table 6.5: Participant drop-out over time shaded by direction and statistical certainty of effect on cognitive ability

Intervention	Population	Latest trial publication	N	% followed-up at age			
				<2 years	2-4 years	5-10 years	11-18 years
Nutrient-enriched	Preterm and SGA	Lucas et al. (1998)	926	81%		83%	4%
		Cooke et al. (2001)	61	?			
		Lucas et al. (2001)	229	86%			
		Morley et al. (2004)	299	79%			
		Jeon et al. (2011)	60	67%			
		Ruys et al. (2018)	102	78%	42%		
LCPUFA	Term	Auestad et al. (2003)	131	?	53%		
		Willatts et al. (2013)	?		?	?	
		Lucas et al. (unpublished)	309	81%	60%		14%
		Birch et al. (2007)	53	70%	68%		
		de Jong et al. (2012)	314	93%		68%	
		Colombo et al. (2017)	91	33%			
		Colombo et al. (2017)	80	58%*	53%	38%	
	Preterm	Fewtrell et al. (2002)	196	64%			9%
		Isaacs et al. (2011)	238	83%			45%
Iron	All	Moffatt et al. (1994)	225	68%			
		Gahagan et al. (2019)	1120	75%		42%	36%
		Morley et al. (1999)	327	82%			
		Friel et al. (2003)	77	60%			
		Lind et al. (2004)	340	93%			
Sn-2 Palmitate	All	no studies	0				
Nucleotides	All	no studies	0				

Legend:

Effect estimate direction	Statistical certainty		
Favours unfortified	P<0.05	P>0.05 and P<0.20	P>0.20 and P<0.50
Unclear	P>0.50		
Favours fortified	P<0.05	P>0.05 and P<0.20	P>0.20 and P<0.50

Footnotes: *% Followed-up* was calculated as participants in the latest cognitive follow-up in each age period over participants randomised (N) to the same group. *"?"* is placed where either number randomised, or number followed up were unclear.

6.7.4 Potential for conflict of interest

There are many ways in which authors can have a conflict of interest, which might lead them to analyse, interpret, or publish results selectively. Examples are patents on formula ingredients, author-industry employment, or industry involvement in design, analysis, or publication. It is critical to emphasize, however, that industry involvement in the form of free supply of study formulas or financial support to the investigators is almost universal in nutrition studies as formulas are often not commercially available at the time. Hence, formula donation is not necessarily predictive of bias. I explored the potential for conflict of interest in detail for review 2 (LCPUFA) because it had the most publications available. **Table 6.6** highlights several potential conflicts of interest. Few were disclosed in the associated publications – this does not necessarily signal malintent but reflects journal policy and disclosure standards at the time. The investigation of potential conflict of interest involved searching for all authors of publications with cognitive outcomes using the Google Patent Register, the European Patent Register and the United States Patent and Trademark Office for authorship on patents that could pose financial incentives to report specific cognitive outcomes. It also involved searching author names on LinkedIn and similar platforms to identify author-industry employment and searching publications and research protocols for details of industry involvement in design, analysis, or publication.

Table 6.6: Conflict of interest analysis for LCPUFA studies discussed in this chapter (table spans 3 pages)

Trial	All co-authors on all publications of cognitive outcomes	Nature of potential conflict of interest	Patents for the use of DHA in infant formula	Years cognitive outcomes were published	Funder of original trial / provider of study formula	Funder(s) for cognitive follow-up studies
Term US: 3 centres (1992-93)	Auestad Nancy, Scott David T, Janowsky Jeri, Jacobsen Cynthia, Carroll Robin E, Montalto Michael B, Halter Robin, Qiu Wenzi, Jacobs Joan R, Connor William E, Connor Sonja L, Taylor J, Neuringer Martha, Fitzgerald-Gustafson KM, Hall Robert T	Several authors (Auestad N, Janowsky J, Halter R, Fitzgerald-Gustafson KM, Neuringer M, Montalto MB) held patents for DHA infant formulas at the time of publication, several authors were employed by Ross at the time of publication (Auestad N, Monalto MB, Halter RMA, Qiu W, Jacobs JR)	US20020045660-A1 (Infant formulas containing long-chain polyunsaturated fatty acids and uses thereof, 2001) US20030190363-A1 (Infant formulas containing long-chain polyunsaturated fatty acids and uses thereof, 2001)	1997, 1998, 2003	Ross (Abbott Laboratories)	Ross (Abbott Laboratories)
Term Europe: 6 centres (1992-93)	Willats Peter, Forsyth J Steward, DiModugno MK, Varma S, Colvin M, Casaer Paul, Agostoni Carlo, Bruzzese Maria Grazia, Trojan Sabina, Bellu Roberto, Riva Enrica, Bissenden J , Smith M, Elliot A, Eggermont Ephrem, McNaughton A, Boehm Günther	Boehm G (senior author) was employed by Danone (NUMICO) at the time the trial was conducted and held patents for DHA infant formulas at the time of publicaton, the trial was also designed by industry	WO2010110658-A1 + WO2010110649-A1 (The present invention concerns a kit of parts of infant milk formula comprising different amount of DHA for stimulating the development of brains, 2009) EP1656839-A1 (Nutrition containing lipid blend, 2004)	1995, 1997, 1998, 2003, 2013	Danone (Numico)	Danone (Numico)
Term ENG: 2 centres (1993-95)	Lucas Alan, Stafford Mai, Abbott Rebecca, Stephenson Terence, MacFayden Una, Elias-Jones Alun, Clements Helena	/		1999	Nestlé (Nestec Ltd.)	Medical Research Council (MRC) and EU

						framework 6 grant (EARNEST)
Term US: Dallas (1993-95)	Birch Eileen, Garfield Sharon, Castaneda Y, Hughbanks-Wheaton D, Hoffman Dennis R, Uauy Ricardo, Birch David G	Several authors (Birch E, Hoffman DR) held patents for DHA infant formulas at the time of publication	US20020045660-A1 (Infant formulas containing long-chain polyunsaturated fatty acids and uses thereof. Methods for providing nutrition and for enhancing neurological development of preterm infants are disclosed, 2001) US7413759-B2 (Method of enhancing cognitive ability in infant fed DHA containing baby-food compositions, 1998)	1998, 2000, 2007	Mead Johnson (Enfamil)	National Institutes of Health (NIH)
Term NL: Groningen (1997-99)	De Jong Corina, Kikkert Hedwig, Fidler Vaclav, Hadders-Algra Mijna, Bouwstra H, Dijck-Brouwer D, Wildeman JA, Tjoonk HM, van der Heide JC, Boersma ER, Muskiet FA, Boehm Günther	Boehm G was employed by Danone (NUMICO) at the time the trial was conducted and held patents for DHA infant formulas, Boehm G also reviewed drafts of follow-up publications	WO2010110649-A1 + WO2010027258-A1 (The present invention concerns a kit of parts of infant milk formula comprising different amount of DHA for stimulating the development of brains, 2009) EP1656839-A1 (Nutrition containing lipid blend, 2004)	2003, 2005, 2010, 2012	Danone (Numico)	EU framework 6 grant (EARNEST)
Term Dallas/ Kansas DIAMOND	Birch Eileen, Colombo John, Carlson, Susan E, Cheatham CL, Castaneda YS, Doty T, Diersen-Schade Deborah A, Drover James	Several authors (Birch E, Carlson SE, Fitzgerald-Gustafson KM, Hoffman DR, Diersen-Schade D) hold patents for DHA infant formulas, additionally	US-9375028-B2 (Compositions and methods for nutrient delivery, 2010) US20020045660-A1 (Infant	2010, 2011, 2017	Mead Johnson (Enfamil)	Mead Johnson Nutrition, NIH, Kansas Intellectual and

Study (2002-04)	R, Fu VL, Fitzgerald-Gustafson KM, Hoffman Dennis R, Kepler A, Kerling EH, Liao K, Lepping RJ, Minns L, Mundy D, Marunycz, McCandliss BD, Sittiprapaporn W, Shaddy DJ, Wheaton DK	Drs. Hoffman and Birch are employed by the Retina Foundation which is funded by Mead Johnson	formulas containing long-chain polyunsaturated fatty acids and uses thereof - Methods for providing nutrition and for enhancing neurological development of preterm infants are disclosed, 2000) US7413759-B2 (Method of enhancing cognitive ability in infant fed DHA containing baby-food compositions, 2004)			Developmental Disabilities Research Centre
Preterm ENG: 2 centres (1993-96)	Lucas Alan, Fewtrell Mary, Morley Ruth, Abbott R, Singhal A, Isaacs EB, Stephenson T, MacFayden U	/		2002	Danone (Numico)	EU framework 6 grant (EARNEST)
Preterm SCT: Glasgow (1995-97)	Fewtrell M, Abbott R, Kennedy K, Singhal A, Morley R, Caine Eleanor, Jamieson EC, Cockburn F, Lucas A, Weaver L, Ross S, Isaacs EB	Weaver is a member of the Infant and Toddler Forum, an educational charity funded by Danone		2004, 2011	Heinz	EU framework 6 grant (EARNEST)

6.7.5 Implications for my thesis

6.7.5.1 Lost equipoise makes existing (dormant) trials uniquely valuable

Earlier in this section, I discussed that most trials included in the reviews reported their results at a time when reporting standards were lower, which makes the grading of evidence quality more difficult. However, it is important to note that older trials are also uniquely valuable: most of these trials could not have been conducted today, as the tested interventions have become almost universal, making randomisation to unfortified formula unacceptable, unfeasible, and therefore untestable. This is the case even as long-term safety and efficacy of today's "standard formulas" remain largely unexplored, and several studies indicating potential adverse effects on cognitive ability have not been published and were therefore unable to contribute to the scientific discourse. The implication for this thesis is that linking dormant trials to administrative education data provides a rare chance to investigate formulas that cannot be tested anymore in a randomised trial today (e.g., term formula for preterm infants or formula not supplemented with LCPUFA). Finally, work for this chapter led to publication of previously unpublished evidence on cognitive effects.

6.7.5.2 Long-term follow-up is necessary to detect late-emerging effects

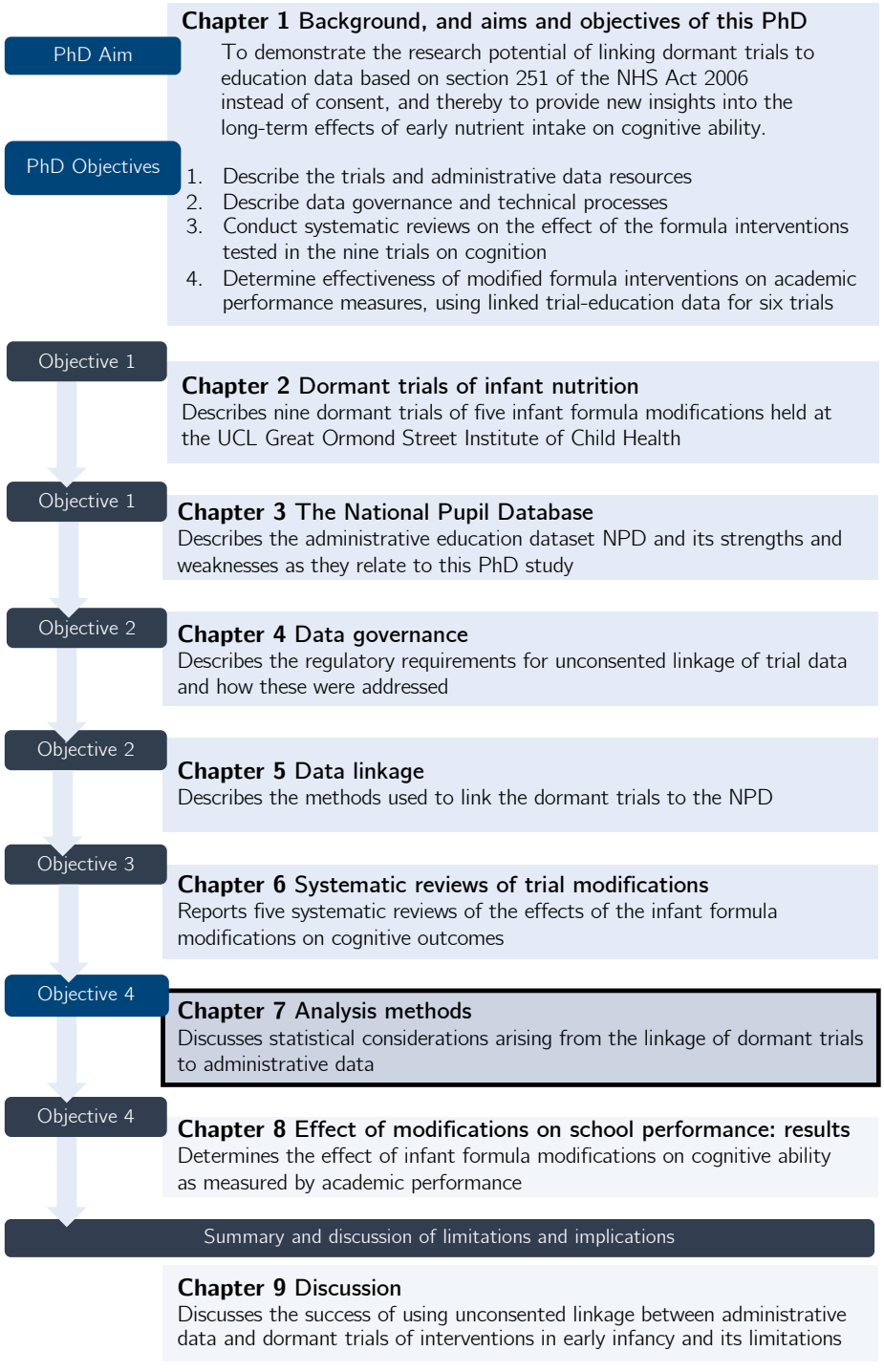
It would be worrying to discover that an infant formula modification might lead to harms that do not emerge until several years onwards. This was the case in the iron trial conducted in Chile, reported in review 4. Initially, there was no difference in mental development scores between the randomised groups. Then, at age 10 years and again at age 16 participants, who were randomised to high-iron formula, had significantly lower IQ scores and scored lower on verbal and arithmetic tests than those who were randomised to low-iron formula. While it is difficult to interpret the finding of this particular study in the context of high participant drop-out (**Table 6.5**), drop-out was not a concern in the study by Lewis et al. (1986) who also found emerging harms. Lewis and his colleagues found that baboons overfed in infancy kept a normal weight in childhood but

had a significantly higher fat mass compared to those fed regular formula in adulthood. In another study of enriched formula, this time in human infants, a nutrient-enriched diet in preterm infants seemed to have no effect on blood pressure at age 7 to 8 years but was associated with a significant increase in blood pressure at age 13 to 16 years when compared to infants fed banked breastmilk (Singhal et al., 2001). A final example of a late-emerging effect was found in an LCPUFA study reported in review 2 (Lucas et al., unpublished). At 18 months, infants randomised to LCPUFA fortified formula showed no difference in mental development compared to those randomised to unfortified formula (Lucas et al., 1999). At age four to six years, however, children who were part of the LCPUFA group had a significant six-point reduction in IQ compared with the unfortified group. Collectively, these emerging effects strongly signal that long term follow-up is necessary to establish the safety and effectiveness of nutritional modifications to infant formula.

6.8 Key points from Chapter 6

- This chapter confirmed that RCT evidence on cognitive effects of the infant formula modifications discussed in this thesis is very limited and of low quality due to insensitive early measures and high attrition rates in later follow-up studies. None of the investigated modifications were associated with consistent cognitive benefits or harms.
- Work for this chapter led to the publication of previously unpublished evidence on cognitive effects.

In the next chapter, I will answer part 1 of objective 4: *determine the effectiveness of modified infant formulas for improving academic performance measures, using linked trial-education data for six dormant infant formula trials,* by presenting and discussing the statistical methods used in the effectiveness analysis.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 7 Statistical analysis methods

Parts of this chapter resulted in a peer-reviewed paper, published here:
https://doi.org/10.1136/bmjopen-2019-035968
Verfürden M, Harron K, Jerrim J, et al Infant formula composition and educational performance: a protocol to extend follow-up for a set of randomised controlled trials using linked administrative education records BMJ Open 2020

7.1 Chapter structure and content

This chapter describes the statistical methods that I used to investigate the effectiveness of infant formula modifications for improving academic performance. I reiterate which trials were included in the analysis, state my analysis objectives and hypotheses, and describe the primary and secondary outcomes. I also present the power calculation, missing data strategy, and analysis strategy. The analysis plan was peer-reviewed and pre-registered (Verfürden et al., 2020).

7.2 Trials included in the academic performance analysis

As outlined in section 2.2 (page 36), the following trials were selected to determine formula modification effects on academic performance: NEP-PD, NETSGA, LCPUFAP, LCPUFAT, IRON, and PALM, whereas information from the other trials was primarily used to evaluate linkage methods and add additional information to the multiple imputation process.

As shown in Chapter 6, there is biological plausibility that modifications in the NEP-PD, NETSGA, LCPUFAP, LCPUFAT, and IRON trials affect cognitive ability. There is little to no biological plausibility that the formula modification in the PALM trial affects cognitive ability. I expected this modification to show no effect on cognitive ability, which is why this trial was used as a negative control trial.

7.3 Analysis objective and hypotheses

The primary objective of the academic performance analysis was to determine the long-term cognitive effect of:

a) post-discharge infant formula enriched with extra calories, protein and other nutrients compared to standard term formula in participants born preterm
b) term infant formula enriched with extra calories, protein and other nutrients compared to standard term formula in participants born term SGA
c) preterm infant formula with added LCPUFA (DHA and AA) compared to unsupplemented preterm formula in participants born preterm
d) term infant formula with added LCPUFA (DHA and AA) compared to unsupplemented term formula in participants born preterm
e) follow-on infant formula with high-iron content compared to low-iron follow-on formula in participants born at term
f) term infant formula with palmitate primarily esterified in the sn-2 position compared to term infant formula with palmitate in standard positions in participants born at term (negative control)

... as measured by academic performance.

Null hypothesis 1: That there is no difference in academic performance between participants randomised to the nutrient-enriched, LCPUFA, high-iron content and sn-2 palmitate formulas compared to those randomised to the respective standard formulas.

Null hypothesis 2: That participants randomised to the nutrient-enriched, LCPUFA, high-iron content and sn-2 palmitate formulas are equally likely compared to participants randomised to the respective standard formulas to qualify for special educational needs (SEN) support.

7.4 Outcomes

7.4.1 Primary outcome

The primary measure of academic performance was the mean difference in within-trial standardised GCSE Maths grades (see section 7.7.2 for details of internal standardisation) for each of the seven trials. In England, GCSE Maths exams are sat in year 11, the year the pupil turns 16 years old, and are compulsory and nationally administered. During the analysis period, GCSE exams were graded from A*=58 points to G=16, or ungraded U=0 points.

Table 7.1: GCSE grade structure during the analysis period (2001-2016)

Grade	Points equivalent	Remarks
A*	58	Highest pass
A	52	
B	46	
C	40	
D	34	
E	28	
F	22	
G	16	Lowest pass
U	0	Ungraded

Maths was chosen as a primary outcome over English (also compulsory and nationally administered) because exam results for Maths are considered to be less subjectively graded (Rhead and Black, 2018). The primary endpoint was chosen to be at age 16 years rather than age 11 as GCSEs are a more relevant predictor of future education and employment opportunities than KS2 exams. The rationale for using GCSE grades over traditional long-term measures of cognition such as IQ has been discussed at length in section 1.2.4, page 24 and section 3.4.1.1, page 56.

7.4.2 Secondary outcomes

As secondary outcomes, I investigated intervention effects of modified versus standard infant formula on:

- Mean difference in modified vs standard formula of GCSE English language exam as within-trial SD-scores.
- Mean difference in modified vs standard formula of Maths and English reading exams as within-trial SD-scores at age 11 years (KS2, final year of primary school).
- Odds ratio in modified vs standard formula of receiving five or more GCSE grades A* to C (including Maths and English).
- Odds ratio in modified vs standard formula of ever being eligible for SEN support.

All mentioned exams are compulsory and nationally administered. GCSE English language scores are graded like GCSE Mathematics scores. KS2 Maths exams are graded from 0 to 100, with 100 being the highest score. KS2 English reading exams at age 11 are graded from 0 to 50, with 50 being highest. Receiving five or more GCSEs with grades A* to C is a commonly reported measure of academic performance as this measure feeds into entry requirements for a large number of sixth form colleges and therefore determines future academic development.

If formula modifications support cognitive development, this could potentially reflect as a lower likelihood of special needs status such as dyslexia. I therefore also aimed to determine whether the modified infant formulas affect the probability of qualifying for SEN support for learning, behavioural, emotional, or specific impairments (e.g., speech and language, hearing, vision).

7.5 Post-linkage delta calculation

Fig. 7.1 displays the minimum detectable mean difference (delta) between modified and standard formula groups for each trial included in the analysis. The total number of participants linked to the primary outcome for each trial is shown on the x-axis, and the minimum detectable differences (δ) are shown on the y-axis. This estimate assumed 80% power (1-β) and a σ of 0.95. σ refers to the average standard deviation of the SD-scores within each trial (note σ is not 1 because covariate adjustment and multiple imputation led to improved statistical efficiency). Sample sizes for the primary outcome were briefly discussed in section 5.5.2, on page 112. Taking the NEP-PD trial as an example, a primary outcome was observed for a total of 211 participants. So, assuming 80% power (1-β) and a σ of 0.95, this trial could detect mean differences equal to and larger than 0.37 SD. Conversely, this implies that the NEP-PD trial is underpowered to detect mean differences that are lower than 0.37 SD.

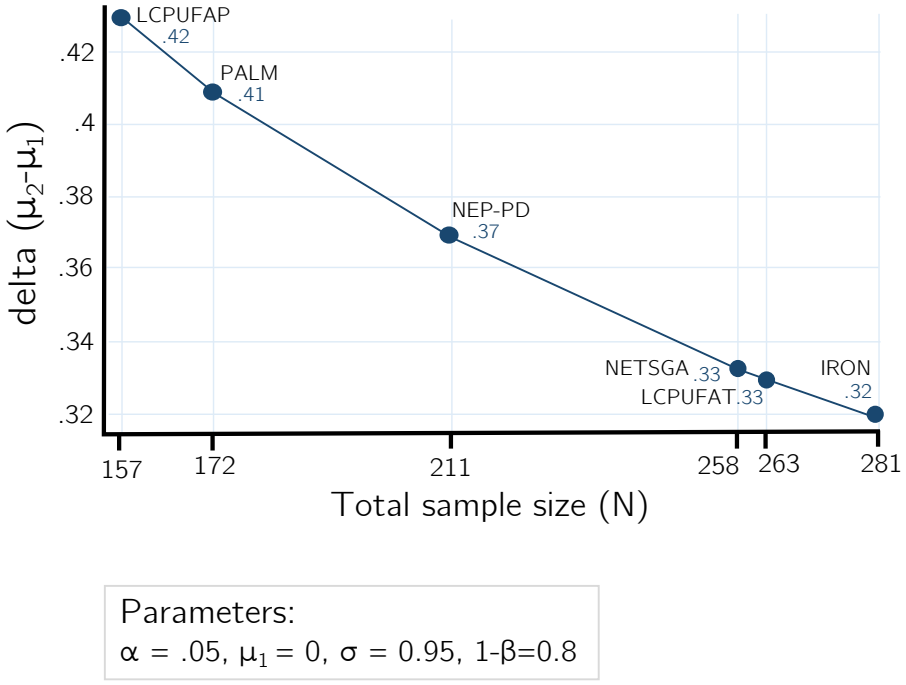


Fig. 7.1: Minimum detectable mean difference (delta) in standardised scores between modified and standard formula group for each dormant trial included in the analyses, relative to total size of linked sample, assuming 80% power and 0.95 σ

7.6 Missing data strategy

7.6.1 Theoretical basis

Missing data refers to all observations that were intended to be obtained but were not. This includes values on baseline covariates as well as outcomes that might not be obtained because of missed links or incomplete data. There are various strategies for handling missing data (**Table 7.2**). These should be selected considering the mechanisms of missingness:

- Missing completely at random (MCAR): If the probability of an observation being missing does not depend on any observed or unobserved measurements, then that observation is *missing completely at random* (MCAR).

- Missing at random (MAR): If the probability of an observation being missing given the observed data does not depend on unobserved data. For example, if I was just as likely to obtain pupil records for participants who wrote unsatisfactory grades compared to those who wrote satisfactory grades, given the information I already have about the participant (e.g., maternal education, birth weight etc.). This assumes that the value of the unobserved variables can be predicted from the observed variables, and therefore that response can be estimated without bias using the observed data exclusively.

- Missing not at random (MNAR): This situation arises when even accounting for all the available observed information, the reason for observations being missing still depends on the unseen observations themselves. For example, if I was less likely to obtain pupil records for participants who wrote unsatisfactory grades than for those who wrote satisfactory grades. This scenario implies that the values of the unobserved variables cannot be predicted without bias by the model.

It is impossible to test empirically whether the probability of observing a value for a variable depends on the value of that variable. However, I considered it plausible that the probability of missing academic performance data was independent of participants' academic performance. This assumption is based on the facts that schools sent exam data to the NPD independent of academic performance and blind to formula group allocation, and that the quality of match-variables was high and did not vary systematically by formula group (**Fig. 5.8**, page 94). Through both the RCT and the NPD data, I had access to a rich set of auxiliary variables, some of them associated with the probability of data being missing, and variables that were predictive of variables that were subject to missingness (in the observed data). For the purposes of my analyses, I therefore assumed that the data was missing at random (MAR).

Complete case analyses, missing indicator analyses, and single imputation analyses are likely to give rise to bias (**Table 7.2**). Bias in these methods of handling missing data arises either because they assume MCAR, overestimate precision, or do not make use of observed data from excluded observations. The exclusion of observed data has several adverse consequences: (1) statistical power is reduced, (2) data are not representative of the randomised participants, (3) it prevents adherence to the intention-to-treat analysis principle, which requires that all participants who were randomised be included in the analysis, irrespective of events that occurred subsequently (Newell, 1992, Lewis and Machin, 1993, Deeks et al., 2011). I therefore decided to use multiple imputation to handle missing data.

Table 7.2: Overview of strategies to handle missing data

	Description of method	Advantages (+) Disadvantages (–)
Complete case analysis	Analysing only the data of participants with complete values	+ Easy to understand and execute – Does not make use of observed datapoints from excluded observations so is not intention-to-treat compatible – Not comparable across analyses using different sets of variables – Standard error too large (low power) – Data may not be representative of randomised sample – Potentially biased results – Assumes MCAR
Missing indicator	Missing values are grouped into a missing group	– Standard error too small (overestimation of precision) – Biased results
Single imputation	Replacing the missing data point with a single value and conduct analysis as if all the data were observed.	– Ignores uncertainty produced by imputation process: standard error too small (overestimation of precision) – Potentially biased results – Ignores relationship between variables – Assumes MCAR
Multiple imputation	Generating multiple copies of the original dataset by replacing missing values using an appropriate prediction model, analyse them as complete datasets and combine the different estimates to produce a single point estimate and standard error.	+ Can handle missing data under multiple missingness mechanisms + Standard error considers uncertainty of imputation process but is more precise compared to complete case analysis since it makes use of all available data – Depending on dataset size it requires large computational power – Difficulty of correctly specifying imputation model

To explain the process of multiple imputation within the context of my PhD study, let us suppose that my data has variables X (birth characteristics) and Y (academic performance), with some Y values MAR given X. The idea behind imputation is to use data from participants where both Y and X are observed to learn about the relationship between Y and X. This relationship forms the basis for a conditional distribution. Then, values for the missing Y observations are randomly drawn multiple times from this conditional distribution, giving rise to K complete data sets. Each of these datasets is then analysed in the usual way and combined using specific rules (referred to as Rubin's rules) (Carpenter and Kenward, 2012). This process is illustrated in **Fig. 7.2** below.

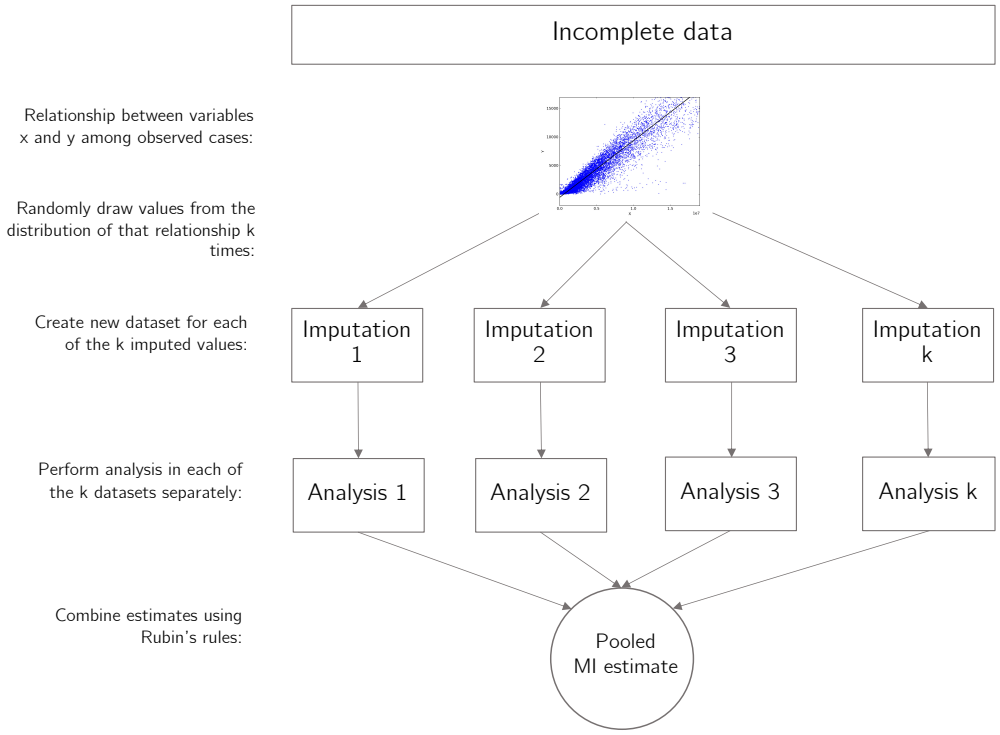


Fig. 7.2: Simplified illustration of multiple imputation process under missing at random (MAR) assumption

7.6.2 Multiple imputation process in this thesis

My imputation model combined observations from all trials to increase statistical power and make better use of all available information. I created K=15 imputations using Stata's *mi chained* command, setting the random seed at 300 (seed prespecified a-priori). The random seed determines the starting value of the random number generator in Stata so that the same set of random numbers will be generated each time the imputation is repeated. This was set to ensure the reproducibility of my model, which would otherwise generate results that vary slightly as values are drawn randomly from the conditional distribution. All analysis covariates (**Table 7.9**, page 177), outcomes, and variables assumed to be associated with the underlying values of the variables subject to missingness, as well as their identified predictors of missingness, were included in the imputation model, which is specified in **Table 7.3** below. The corresponding Stata code can be found here: https://github.com/MaxVerfuerden/PhD

Table 7.3: Multiple imputation by chained equations, model specifications

Analysis variables	Type of prediction	Auxiliary variables	Additional model specifications for this variable	% of missing observations in randomised participants, by analysis trial
Maths age 16	Predictive mean matching (Morris et al., 2014)	• Centre, • Trial, • Birth month, • Allocation group, • English age 16, • Maths age 10, • English age 10, • At least 5 GCSEs at grades A*-C, • Ever received Special educational needs support, • IQ score, • Bayley MDI, • Bayley PDI, • Maternal education, • Maternal age, • Smoking during pregnancy, • Birth weight • Gestational age • Multiple pregnancy • Number of previous follow-ups attended	Use only the distribution from the 8 closest matches (knn=8)	NEP-PD: 7.9% NETSGA: 13.7% LCPUFAP: 19.9% LCPUFAT: 14.9% IRON: 13.8% PALM: 15.3%
English age 16	Predictive mean matching		knn=8	NEP-PD: 13.2% NETSGA: 16.7% LCPUFAP: 25.5% LCPUFAT: 17.5% IRON: 48.5% PALM: 25.6%
Maths age 11	Predictive mean matching		knn=8	NEP-PD: 11.8% NETSGA: 15.0% LCPUFAP: 21.9% LCPUFAT: 12.9% IRON: 12.3% PALM: 12.3%
English age 11	Predictive mean matching		knn=8	NEP-PD: 13.6% NETSGA: 14.7% LCPUFAP: 25.5% LCPUFAT: 14.9% IRON: 14.1% PALM: 13.8%
At least 5 GCSEs at grades A*-C	Logistic regression			NEP-PD: 2.6% NETSGA: 10.0% LCPUFAP: 13.3% LCPUFAT: 10.0% IRON: 10.1% PALM: 8.9%
Ever eligible for SEN support	Logistic regression			NEP-PD: 2.6% NETSGA: 10.0% LCPUFAP: 12.8% LCPUFAT: 9.4% IRON: 8.3% PALM: 6.4%
IQ score	Linear regression		Impute only if IQ was measured in the trial	NEP-PD: 100% NETSGA: 90.6% LCPUFAP: 93.4% LCPUFAT: 40.5% IRON: 100% PALM: 100%
Bayley MDI	Linear regression		Impute only if Bayley score was	NEP-PD: 18.3% NETSGA: 27.4% LCPUFAP: 29.1%

			measured in the trial	LCPUFAT: 24.3% IRON: 19% PALM: 100%
Maternal education	Ordered logistic regression			NEP-PD: 2.2% NETSGA: 0.67% LCPUFAP: 44.4% LCPUFAT: 1.6% IRON: 0.92% PALM: 71.9%
Maternal age	Predictive mean matching			NEP-PD: 0% NETSGA: 0% LCPUFAP: 14.8% LCPUFAT: 0.65% IRON: 1.23% PALM: 0%
Smoking during pregnancy	Logistic regression			NEP-PD: 4.0% NETSGA: 7.7% LCPUFAP: 0% LCPUFAT: 2.6% IRON: 1.23% PALM: 0%
Birth weight	Predictive mean matching		knn=8	NEP-PD: 0% NETSGA: 0% LCPUFAP: 0% LCPUFAT: 0% IRON: 1.23% PALM: 0%
Multiple pregnancy	Logistic regression			NEP-PD: 2.6% NETSGA: 10% LCPUFAP: 12.8% LCPUFAT: 0% IRON: 0% PALM: 0%
Gestational age	Predictive mean matching		knn=8	NEP-PD: 0% NETSGA: 0% LCPUFAP: 0% LCPUFAT: 0% IRON: 0.31% PALM: 0%

7.7 Analysis strategy

7.7.1 Intention to treat principle

The strategy adopted for all outcomes was intention-to-treat (ITT) analysis. This means that analyses were conducted on all randomised participants in the group to which they were allocated, irrespective of what happened afterwards. Only participants who died before school age were excluded from analyses.

7.7.2 Standardisation of academic performance measures

To compare academic performance to previously collected in-trial cognitive outcomes and track cognitive development over time, I converted all outcomes into standard deviation (SD)-scores (also known as z-scores). Differences between group mean SD-scores are referred to as standardised mean differences, effect sizes, or Cohen's 'd'. They are common outcome measures, both in clinical and education trials, and are frequently used to compare GCSE scores across years and groups. An SD-score indicates how far above or below the mean of some distribution an observation lies, with the distance being measured in standard deviations. As my main interest was in the relative position of participants compared to other participants in the same trial, I used the distribution of observed outcomes within each respective trial to calculate standard deviation scores. This is called internal standardisation: $SD\,score = \frac{x-\mu}{SD_{pooled}}$, where SD_{pooled} is the pooled standard deviation of the respective exam point scores from modified and standard formula groups within the trial, x is the observed exam point score and μ is the mean of the exam point scores within the trial. One standard deviation equates to 11-13 points (depending on the trial) on the 58-point score (**Table 7.1**) and could mean the difference between an A and a C grade, which would be an important difference for an individual pupil.

As SD-scores are computed using means, they are sensitive to outliers and skewed distributions. I therefore checked within each trial for outliers and the assumption of normality.

7.7.2.1 Issues with standardising outcomes in multiply imputed data

Special consideration is warranted when internally standardising outcomes in multiply imputed data for two reasons. First, the distribution of exam scores within each trial needs to consider the imputed outcome values for unlinked participants. SD-scores must therefore also incorporate the added uncertainty generated by the imputation process. Second, the way statistical programs handle imputed datasets can bias the results and their confidence intervals, depending on the amount of missing data in an outcome, leading to trial means that are not 0 and trial SDs that are not 1.

To address the first issue, I used Stata's *mi passive* command, which applies Rubin's rules (Rubin, 2004) and thereby avoids artificial inflation of precision. To illustrate the second issue, how to derive valid means and standard deviations, I will use a simplified example of a fictive trial with a total of five participants with exam grades with possible points 0-10. Three participants have data on exams observed, while two have missing grade data (**Table 7.4**). Five imputations were performed for each observation using auxiliary variables.

Table 7.4: Example trial with N=5 with missing data and 5 imputations for illustrating the steps needed to internally standardise multiply imputed values

Total trial N=5	Observed exam grade	Imputed exam grade 1	Imputed exam grade 2	Imputed exam grade 3	Imputed exam grade 4	Imputed exam grade 5
Person 1	*missing*	**4**	**0**	**8**	**5**	**4**
Person 2	*missing*	**7**	**9**	**10**	**9**	**5**
Person 3	4	**4**	**4**	**4**	**4**	**4**
Person 4	8	**8**	**8**	**8**	**8**	**8**
Person 5	7	**7**	**7**	**7**	**7**	**7**

In the naïve (i.e., biased) approach, SD-scores would be calculated directly using the *mi passive* command, taking the mean and the SD for each of the five datasets (columns). For instance, in the dataset representing imputation 1:
Mean = (4+7+4+8+7) ÷ 5 = **6** and sum of differences to the mean squared = 14, sum of differences to the mean squared divided by count = 2.8 (variance). SD (square root of the variance) = **1.67** (**Table 7.5**). The SD-score would then

be the exam grade minus **6** divided by **1.67**. So, in person 1, MI dataset 1 this would be: (4 – 6) ÷ 1.67 = -1.20 (**Table 7.6**).

Table 7.5: Example trial showing mean and SD of exam grades within each MI dataset

Total trial N=5	Imputed exam grade 1	Imputed exam grade 2	Imputed exam grade 3	Imputed exam grade 4	Imputed exam grade 5
Person 1	4	0	8	5	4
Person 2	7	9	10	9	5
Person 3	4	4	4	4	4
Person 4	8	8	8	8	8
Person 5	7	7	7	7	7
Mean exam grade (columns)	**6**	**5.6**	**7.4**	**6.6**	**5.6**
SD of exam grades (columns)	**1.67**	**3.26**	**1.96**	**1.85**	**1.62**

As a consequence, the mean SD-score within this trial would not be 0 and the SD of the SD-score not 1, as they should be.

Table 7.6: Example trial showing that calculating SD-scores within each of the imputed datasets and then combining them leads to bias in the internally standardised variable with the trial mean SD-score not being 0 and the SD of the SD-scores not 1

Total trial N=5	*mi passive* generated SD-score 1	*mi passive* generated SD-score 2	*mi passive* generated SD-score 3	*mi passive* generated SD-score 4	*mi passive* generated SD-score 5	Average SD-score for each participant
Person 1	-1.20	-1.72	0.31	-0.86	-0.99	**-0.892**
Person 2	0.60	1.04	1.33	1.3	-0.37	**0.780**
Person 3	-1.20	-0.49	-1.73	-1.41	-0.99	**-1.164**
Person 4	1.20	0.74	0.31	0.76	1.48	**0.898**
Person 5	0.60	0.43	-0.2	0.22	0.86	**0.382**
Trial mean of SD-scores						***0.0008***
Trial SD of SD-scores						***0.862***

To overcome this issue, I first used *mi passive* to calculate each individual's average of their imputed exam scores. I then used *mi passive* again to generate SD-scores based on the average exam score within each participant (**Table 7.7**), leading to the correct summary statistics for the internally standardised exam grades (**Table 7.8**).

Table 7.7: Example trial dataset showing average grade for each participant

Total trial N=5	Imputed exam grade 1	Imputed exam grade 2	Imputed exam grade 3	Imputed exam grade 4	Imputed exam grade 5	Participant average grade
Person 1	4	0	8	5	4	**4.2**
Person 2	7	9	10	9	5	**8**
Person 3	4	4	4	4	4	**4**
Person 4	8	8	8	8	8	**8**
Person 5	7	7	7	7	7	**7**
Trial overall mean of exam grades						***6.24***
Trial overall SD of exam grades						***1.78***

Table 7.8: Example trial dataset showing unbiased estimates of the internally standardised exam grade by basing the trial SD-score on the average of the imputed datasets

Total trial N=5	Imputed exam grade 1	Imputed exam grade 2	Imputed exam grade 3	Imputed exam grade 4	Imputed exam grade 5	Participant average internally standardised grade
Person 1	-1.15	-1.15	-1.15	-1.15	-1.15	**-1.15**
Person 2	0.99	0.99	0.99	0.99	0.99	**0.99**
Person 3	-1.26	-1.26	-1.26	-1.26	-1.26	**-1.26**
Person 4	0.99	0.99	0.99	0.99	0.99	**0.99**
Person 5	0.43	0.43	0.43	0.43	0.43	**0.43**
Trial overall mean of SD-scores						***0***
Trial overall SD of SD-scores						***1***

7.7.3 Analysis of the primary outcome

Multivariable linear regression was used to compare modified and standard infant formula groups within each trial, adjusted for a-priori determined covariates to increase statistical efficiency (**Table 7.9**). This model assumes that Maths scores are linearly related to the other predictors, that there is no multicollinearity between the predictors, residuals are normally distributed and independent, and that the variance is constant across the outcome and the predictors. These assumptions were tested for each of the models, revealing heteroscedasticity in some models. I applied robust standard errors throughout to account for heteroscedasticity in the residuals.

To increase the statistical efficiency of my analysis, I adjusted for covariates that have a strong association with the outcome and were measured at randomisation (i.e., are not potentially on the causal pathway between intervention and outcome). Based on previously published studies (Sammons et al., 2014, Botting et al., 1998, Abel et al., 2017), I expected infant sex, maternal smoking and maternal education, and infant birth weight and gestational age to be strongly associated with academic performance. Analyses involving multi-centre trials with separate randomisation schedules for each centre were also adjusted for centre according to the guideline on adjustment for baseline covariates in clinical trials laid out by the European Medicines Agency (2015).

Table 7.9: Variables a-priori expected to be associated with academic performance

	Infant sex	Centre	Gestational age	Birth weight	Maternal smoking	Maternal qualifications
	Binary	*Nominal*	*Discrete*	*Continuous*	*Binary*	*Ordinal*
Details	*1=male, 2=female*	Nottingham, Sheffield, Norwich, Cambridge, Leicester, Ipswich, Kings Lynn	*Measured in full weeks, Range: 25-43 weeks*	*Measured in grams, Range 630-5400 grams*	*1=yes, 2=no*	*1= 3 CSEs or below, 2= at least 3 GCSEs or any O levels or A levels, 3=degree or higher*

7.7.4 Analysis of secondary outcomes

Where secondary outcomes were exam scores (GCSE English SD-scores, KS2 Maths and English SD-scores), they were analysed in the same way as the primary outcome. For binary outcomes (receiving five or more GCSE grades A* to C, ever eligible for SEN support), I used multivariable logistic regression to produce odds ratios.

7.8 Detection of bias in the analysis through negative controls

I aimed to derive valid causal inference of any observed associations between randomisation to modified formula and academic performance. To probe the credibility of observed associations, I used a negative control trial. Negative controls work by creating a setting that is very likely to involve the same sources of bias that may have been present in the original association, but that cannot involve the hypothesized causal mechanism (Lipsitch et al., 2010). Based on lack of evidence from RCTs (Chapter 6) and biological plausibility, the modified and standard formula groups in the PALM trial were not expected to affect academic performance. I assumed that the conditions under which PALM participants were included in the modified formula group and linked to the NPD were the same as in the analysis trials. In other words, if any bias is responsible for the effect observed between modified formulas and academic performance in the five analysis trials, the same bias might be associated with the modification in the PALM trial, which is not plausibly associated with academic performance. In practice, this means that observing a cognitive effect in all six trials would point towards potential for bias. In contrast, cognitive effects in any of the five trials and no cognitive effect in the PALM trial would grant the observed associations more credibility.

7.9 Additional analyses

7.9.1 Sensitivity analyses of the primary outcome

To understand whether the results from the primary analyses were sensitive to small baseline imbalances, missing data, or selection of standardisation reference distribution, I performed several sensitivity analyses:

- Unadjusted multiple imputation (MI) analysis
- Adjusted complete case analysis
- Unadjusted complete case analysis
- External standardisation to the national grade average in the years 2008/09 to 2011/12. (n.b. GCSE Maths results are by default normally distributed as they are standardised within each subject by Ofqual, results discussed in appendix p 52)

Similar mean-differences in all of these analyses would suggest that the results are likely to be robust to observed baseline imbalances, missing data strategy, and choice of standardisation reference distribution.

7.9.2 Exploratory subgroup analyses

Previous studies suggested that the effect of infant formula modifications may depend on whether the infant was born a boy or a girl (Lucas and Sampson, 2006) and whether the mother smoked during pregnancy or not (de Jong et al., 2012). While none of the trials were powered to detect interactions, I decided to present these subgroup analyses to make them available for inclusion in future meta-analyses (appendix p 50).

7.9.3 Consistency of academic performance with previously collected in-trial cognitive outcomes

To graphically explore cognitive trajectories over time, I plotted all cognitive outcomes for each trial, using internal standardisation to bring them on the same

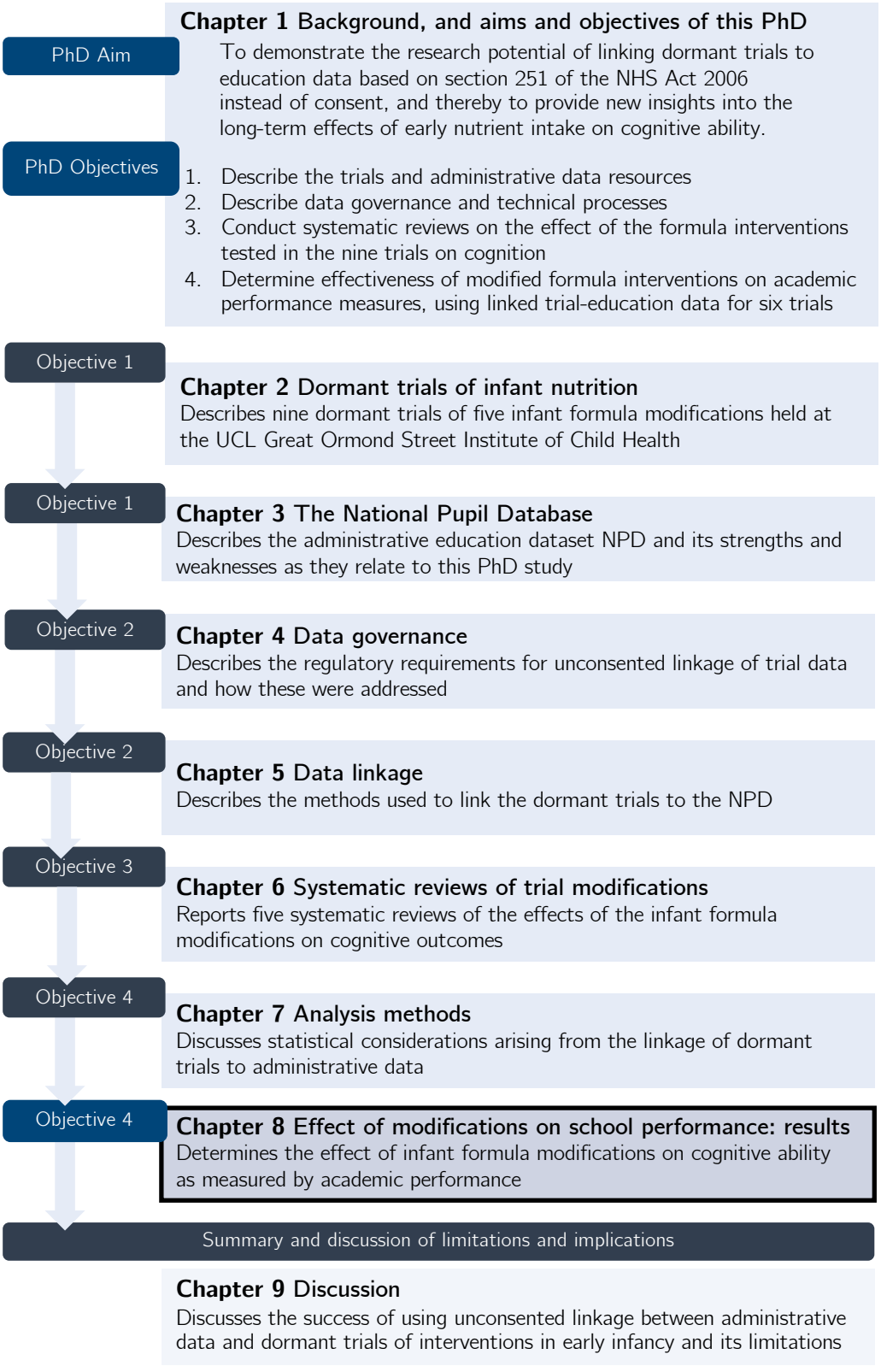
scale. This allowed me to visualise whether there might be specific trajectory effects, such as a widening or narrowing of mean differences between modified and standard formula groups over time, and to observe whether academic performance is in line with previous findings.

I also attempted to quantify how well early cognitive development predicts performance at school by determining the correlation between all measured cognitive outcomes.

7.10 Key points from Chapter 7

- This analysis plan was pre-registered, peer-reviewed and published.
- All analyses were conducted according to the intention-to-treat principle.
- The primary outcome was the within-trial standardised mean difference in GCSE Maths grades between modified and standard formula groups.
- Secondary outcomes were GCSE English within-trial SD-scores, KS2 Maths and English reading exam within-trial SD-scores, receiving five or more GCSE grades A* to C (including Maths and English) and ever being eligible for special educational needs support.
- Assuming 80% power this study was able to detect statistically significant mean differences (at the 5% level) in the primary outcome between 0.32 SD and 0.42 SD and above, depending on the trial.
- To address missing data in the outcomes and covariates, I used multiple imputation.
- To improve statistical efficiency, I adjusted analyses with a-priori determined covariates.

In the next chapter, I will conclude objective 4 by presenting and discussing the analysis results.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 8 Effectiveness of infant formula modifications for improving academic performance: results

Parts of this chapter were written up and submitted for publication.

All Stata code for the following analyses is published online here: https://github.com/MaxVerfuerden/PhD

8.1 Chapter structure and content

The work presented in the preceding chapters demonstrated that it is feasible to link dormant trial data (the infant nutrition RCTs) to administrative education data (the NPD) using archived names and addresses that were recorded during the trial but not updated after the trial was completed. Linkage resulted in a dataset with low attrition, representative of the original randomised trials in terms of the observed characteristics. This chapter uses this dataset and reports analyses of the effect of five infant formula modifications on academic outcomes according to the methods specified in Chapter 7. I show how consistent academic outcomes were with cognitive outcomes measured during the trials, and discuss the strength and limitations of my analyses.

8.2 Follow-up and characteristics of linked participants

8.2.1 Follow-up

Fig 8.1 illustrates the participant flow in the six analysis trials and shows the number of participants with previous cognitive data and the number who linked to the NPD by trial and trial arm. Of the 1,563 randomised participants in the six trials, 1,557 survived during the initial trial period. Of those, 92% (1,431/1,557) linked to the NPD, and 86% (1,342/1,557) had data on the primary outcome.

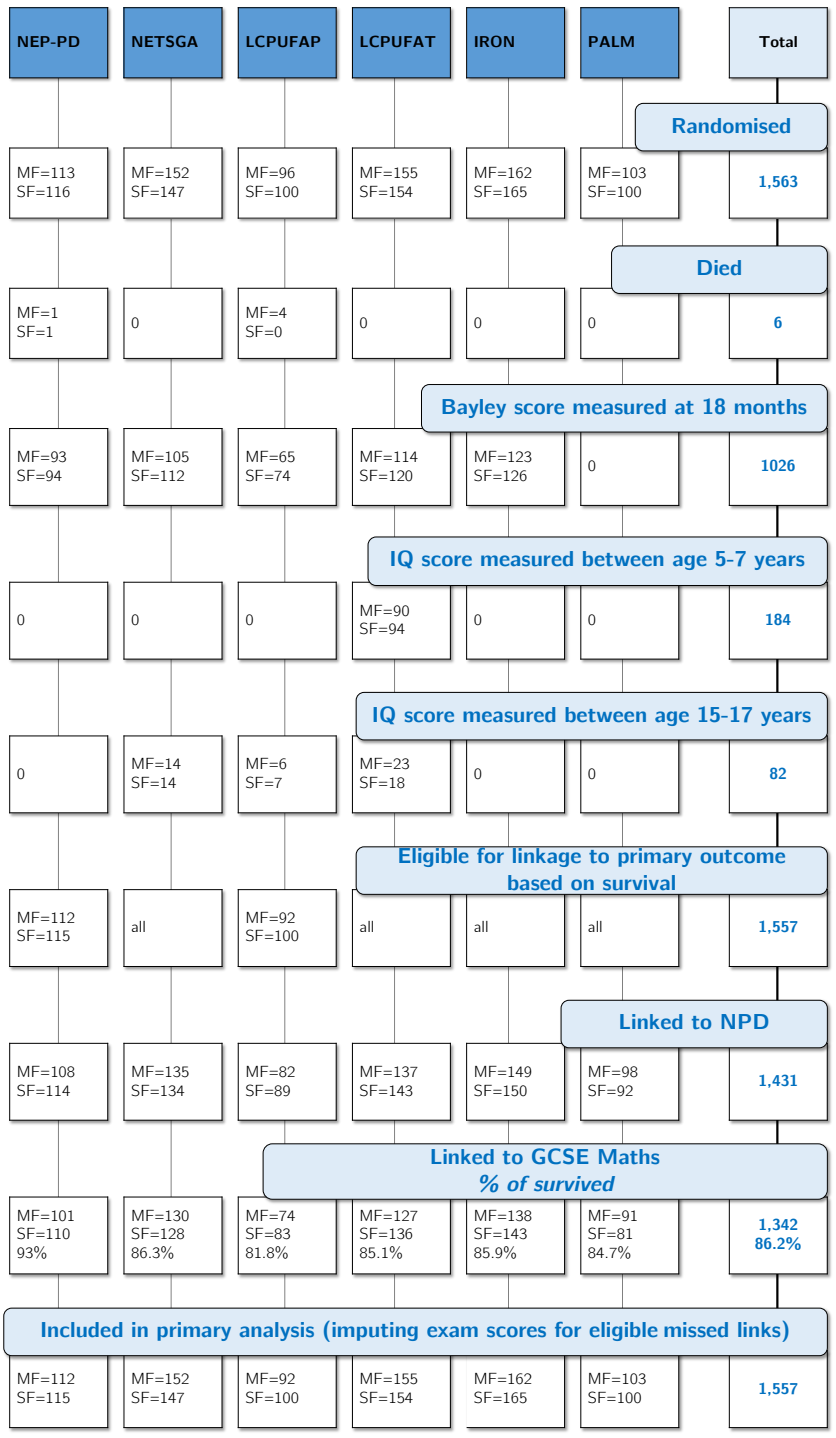


Fig 8.1: Flow diagram showing number of participants with cognitive data and number of participants who linked to the NPD by trial and trial arm (MF=modified formula, SF= standard formula).

8.2.2 Baseline characteristics

Baseline characteristics of those who survived and linked to an NPD record were mostly balanced across trial groups (**Table 8.1**). Infants in the intervention group of the NEP-PD trial were less likely to have a mother with a degree (6% vs 12%, p=0.131), and infants in the intervention group of the LCPUFAT trial were more likely to have a mother with a degree (8% vs 4%, p=0.169).

Table 8.1: Baseline participant characteristics in the population linked to the National Pupil Database

	NEP-PD		**NETSGA**		**LCPUFAP**		**LCPUFAT**		**IRON**		**PALM**	
	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard	Modified	Standard
Linked/randomised	108/113	114/116	135/152	134/147	82/96	89/100	137/155	143/154	149/162	150/165	98/103	92/100
Birth weight (grams)	1392 (775-2160)	1363 (630-2020)	2538 (1400-3160)	2597 (1770-3160)	1351 (640-1850)	1344 (740-1780)	3654 (2960-4900)	3532 (2680-4930)	3483 (2495-5046)	3468 (2466-4706)	3588 (2460-4730)	3472 (2520-5400)
Gestational age (weeks)	30.8 (26-36)	30.8 (25-36)	39.1 (37-42)	39.4 (37-42)	30.5 (25-36)	30.1 (25-36)	40.1 (37-42)	40.0 (37-42)	39.7 (36-43)	39.9 (35-43)	40.1 (37-42)	39.9 (37-42)
Mother's age (years)	28.0 (16-41)	28.5 (17-44)	26.9 (15-42)	26.6 (16-42)	26.2 (16-39)	26.6 (17-39)	27.8 (17-44)	27.0 (18-41)	27.8 (17-40)	27.6 (15-39)	27 (15-40)	28 (17-42)
Infant sex												
Male	51 (47%)	56 (49%)	68 (50%)	63 (47%)	38 (46%)	47 (53%)	73 (53%)	77 (54%)	76 (51%)	76 (51%)	62 (63%)	46 (50%)
Female	57 (53%)	58 (51%)	67 (50%)	71 (53%)	44 (53%)	42 (47%)	64 (47%)	66 (46%)	73 (49%)	74 (49%)	36 (37%)	46 (50%)
Mother smoked during pregnancy												
No	64 (62%)	74 (68%)	69 (55%)	62 (51%)	46 (56%)	54 (61%)	102 (76%)	104 (75%)	108 (73%)	104 (71%)	62 (63%)	69 (75%)
Yes	38 (38%)	35 (32%)	57 (45%)	59 (49%)	36 (44%)	35 (39%)	32 (24%)	35 (25%)	40 (27%)	43 (29%)	36 (37%)	23 (25%)
Missing	4	5	9	13	0	0	3	4	1	3	0	0
Mother has degree												
No	102 (94%)	96 (88%)	*	127 (95%)	*	*	124 (92%)	135 (96%)	130 (88%)	136 (92%)		
Yes	6 (6%)	13 (12%)	*	6 (5%)	*	*	11 (8%)	6 (4%)	18 (12%)	12 (8%)	*	*
Missing	0	5	0	1	29	41	2	2	1	2	71	73

Footnotes: Data are mean (min-max); n (%); * output suppressed due to small cell size (n<6)

8.3 Academic performance

8.3.1 GCSE Maths at age 16 years

8.3.1.1 Distribution of Maths grades by trial and group

The distribution of Maths grades by trial and group in **Fig. 8.2** below illustrates that – as expected – trials with participants born preterm or at term SGA had overall lower GCSE Maths grades (the first three trials on the left), compared to trials with participants born at term (the three trials on the right).

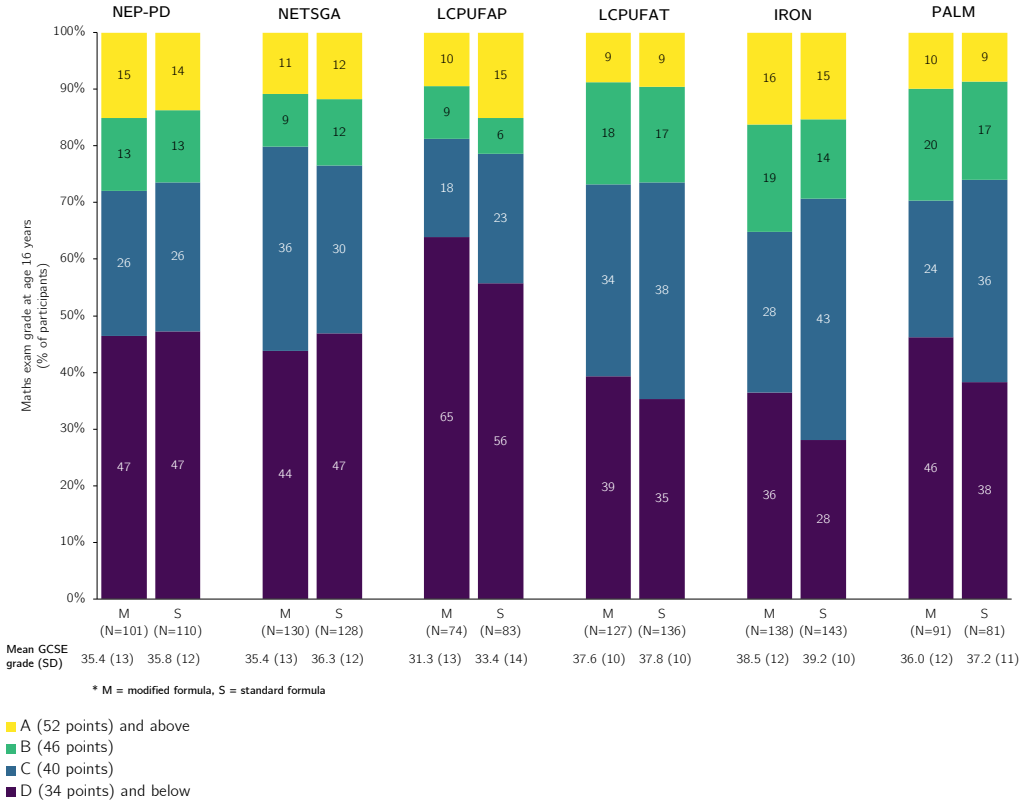


Fig. 8.2: Distribution of Maths grades at age 16 years (unadjusted complete case observations)

8.3.1.2 Mean differences between modified and standard formula groups: primary analysis and sensitivity analyses

All 1,557 surviving participants were included in the primary analysis through multiple imputation of outcomes and covariates. Within each trial, grades were distributed normally and without extreme outliers. **Table 8.2** to **Table 8.7** show the mean and standard deviation of within-trial standardised GCSE Maths grades in the modified and standard formula groups and their mean differences with 95% confidence intervals. They indicate that nutritional modification of the assessed infant and follow-on formulas did not affect Maths performance at age 16 years.

The upper 95% confidence intervals exclude benefits that are larger than 0.27 SD for nutrient-enriched post-discharge formula for preterm infants, larger than 0.12 SD for nutrient-enriched formula for SGA infants, 0.08 SD for LCPUFA fortified formula in preterm infants and healthy terms, 0.07 SD for iron fortified formula and 0.19 SD for formula with palmitate in the sn-2 position.

The lower confidence intervals include the possibility of a reduction of up to -0.22 SD for nutrient-enriched post-discharge formula for preterm infants, -0.33 SD for nutrient-enriched formula for SGA infants, -0.46 SD for LCPUFA fortified formula in preterm infants, -0.36 for LCPUFA fortified formula in terms, -0.31 SD for high-iron fortified formula, and -0.37 SD for the sn-2 palmitate formula. Thus, scope for benefit of the modified infant formulas on Maths performance is relatively small while scope for harm is comparatively large.

In all trials, results were robust to covariate adjustment, method of handling missing data and standardisation reference distribution as shown by the sensitivity analyses presented in the same tables.

Table 8.2: Primary analysis and sensitivity analyses of the primary outcome in the NEP-PD trial showing mean and standard deviation of within-trial standardised GCSE Maths grade in the nutrient-enriched and standard formula group and their mean differences with 95% confidence interval

NEP-PD	Nutrient-enriched formula group		Standard term-formula group			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_E=113, N_S=116)	0.01	0.95	-0.01	0.92	0.02	-0.22, 0.27
Sensitivity analyses						
MI unadjusted (N_E=113, N_S=116)	0.00	1.03	0.00	0.98	-0.01	-0.27, 0.25
Complete-case adjusted (N_E=98, N_S=100)	0.06	0.94	-0.03	0.89	0.09	-0.16, 0.35
Complete-case unadjusted (N_E=101, N_S=110)	-0.01	1.04	0.01	0.97	-0.02	-0.29, 0.26
MI adjusted national SD (N_E=113, N_S=116)	-0.35	1.06	-0.38	0.98	0.03	-0.24, 0.29

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; NEP-PD *nutrient-enriched preterm post-discharge trial*; N_E *Number of participants in the enriched formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

Table 8.3: Primary analysis and sensitivity analyses of the primary outcome in the NETSGA trial showing mean and standard deviation of within-trial standardised GCSE Maths grade in the nutrient-enriched and standard term formula group and their mean differences with 95% confidence interval

NETSGA	Nutrient-enriched formula group		Standard term formula group			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_E=152, N_S=147)	-0.05	0.97	0.05	0.99	-0.11	-0.33, 0.12
Sensitivity analyses						
MI unadjusted (N_E =152, N_S =147)	-0.02	1.00	0.02	1.00	-0.04	-0.27, 0.19
Complete-case adjusted (N_E =122, N_S =114)	-0.10	1.00	0.09	1.05	-0.19	-0.46, 0.08
Complete-case unadjusted (N_E =130, N_S =128)	-0.03	1.01	0.03	0.99	-0.07	-0.31, 0.18
MI adjusted national SD (N_E =152, N_S =147)	-0.40	0.93	-0.29	0.95	-0.10	-0.35, 0.14

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; NETSGA *nutrient-enriched term small-for-gestational-age trial*; N_E *Number of participants in the enriched formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

Table 8.4: Primary analysis and sensitivity analyses of the primary outcome in the LCPUFAP trial showing mean and standard deviation of within-trial standardised GCSE Maths grade in the LCPUFA preterm and standard preterm formula group and their mean differences with 95% confidence interval

LCPUFAP	LCPUFA formula (0.17% DHA and 0.31% AA / total fat)		Standard preterm formula group (no DHA or AA)			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_L=92, N_S=100)	-0.10	0.89	0.09	0.97	-0.19	-0.46, 0.08
Sensitivity analyses						
MI unadjusted (N_L=92, N_S=100)	-0.10	0.97	0.10	1.03	-0.20	-0.48, 0.09
Complete-case adjusted (N_L=48, N_S=44)	-0.02	0.92	0.20	1.00	-0.22	-0.62, 0.19
Complete-case unadjusted (N_L=74, N_S=83)	-0.09	0.97	-0.08	1.03	-0.17	-0.49, 0.14
MI adjusted national SD (N_L=92, N_S=100)	-0.69	0.93	-0.49	1.02	-0.20	-0.51, 0.12

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; LCPUFAP *long-chain polyunsaturated fatty acid supplemented formula for preterm babies trial*; N_L *Number of participants in the LCPUFA formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

Table 8.5: Primary analysis and sensitivity analyses of the primary outcome in the LCPUFAT trial showing mean and standard deviation of within-trial standardised GCSE Maths grade in the LCPUFA and standard formula group and their mean differences with 95% confidence interval

LCPUFAT	LCPUFA formula (0.32% DHA and 0.30% AA /total fat)		Standard term formula (no DHA or AA)			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_L=155, N_S=154)	-0.07	0.97	0.07	0.97	-0.14	-0.36, 0.08
Sensitivity analyses						
MI unadjusted (N_L=155, N_S=154)	-0.04	1.02	0.04	0.99	-0.07	-0.30, 0.15
Complete-case adjusted (N_L=122 N_S=130)	-0.02	0.96	0.07	0.98	-0.09	-0.33, 0.16
Complete-case unadjusted (N_L=127, N_S=136)	-0.02	1.01	0.02	0.99	-0.04	-0.28, 0.21
MI adjusted national SD (N_L=155, N_S=154)	-0.29	0.85	-0.17	0.84	-0.12	-0.34, 0.09

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; LCPUFAT *long-chain polyunsaturated fatty acid supplemented formula for term-babies trial*; N_L *Number of participants in the LCPUFA formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

Table 8.6: Primary analysis and sensitivity analyses of the primary outcome in the IRON trial showing mean and standard deviation of within-trial standardised GCSE Maths grade in the high-iron and low-iron formula group and their mean differences with 95% confidence interval

IRON	High-iron (12mg/dl) formula		Low iron (0.9mg/dl) formula			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_{HI}=162, N_{LI}=165)	-0.06	0.95	0.06	0.83	-0.12	-0.31, 0.07
Sensitivity analyses						
MI unadjusted (N_{HI}=162, N_{LI}=165)	-0.05	1.07	0.05	0.93	-0.10	-0.32, 0.12
Complete-case adjusted (N_{HI}=137, N_{LI}=140)	-0.05	0.98	0.04	0.86	-0.09	-0.31, 0.13
Complete-case unadjusted (N_{HI}=138, N_{LI}=143)	-0.02	1.06	0.02	0.94	-0.05	-0.28, 0.19
MI adjusted national SD (N_{HI}=162, N_{LI}=165)	-0.12	0.89	-0.01	0.78	-0.12	-0.32, 0.09

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; IRON *iron trial*; N_{HI} *Number of participants in the high-iron formula group*; N_{LI} *Number of participants in the low-iron formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

Table 8.7: Primary analysis and sensitivity analyses of the primary outcome in the negative control trial, PALM showing mean and standard deviation of within-trial standardised GCSE Maths grade in the sn-2 palmitate and standard formula group and their mean differences with 95% confidence interval

PALM	Sn-2 palmitate formula		Standard palmitate formula			
Primary outcome: within-trial standardised GCSE Maths grade	*Mean SD-score*	*SD*	*Mean SD-score*	*SD*	*Standardised mean difference*	*95% CI*
Primary analysis						
MI adjusted (N_P=103, N_S=100)	-0.04	1.03	0.05	0.96	-0.09	-0.37, 0.19
Sensitivity analyses						
MI unadjusted (N_P=103, N_S=100)	-0.07	1.04	0.07	0.96	-0.14	-0.41, 0.14
Complete-case adjusted (N_P=27, N_S=27)	0.21	1.05	0.36	0.81	-0.16	-0.71, 0.39
Complete-case unadjusted (N_P=91, N_S=81)	-0.05	1.03	0.05	0.96	-0.10	-0.40, 0.20
MI adjusted national SD (N_P=103, N_S=100)	-0.27	1.06	0.18	0.98	-0.09	-0.37, 0.19

Footnotes: GCSE *General Certificate of Secondary Education*; MI *multiple imputation of covariates and outcomes*; PALM *Sn-2 Palmitate trial*; N_P *Number of participants in the sn-2 palmitate formula group*; N_S *Number of participants in the standard palmitate formula group;* SD *standard deviation;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; National reference distribution: UK GCSE Maths grades and SD pooled for 2008/09 to 2011/12.

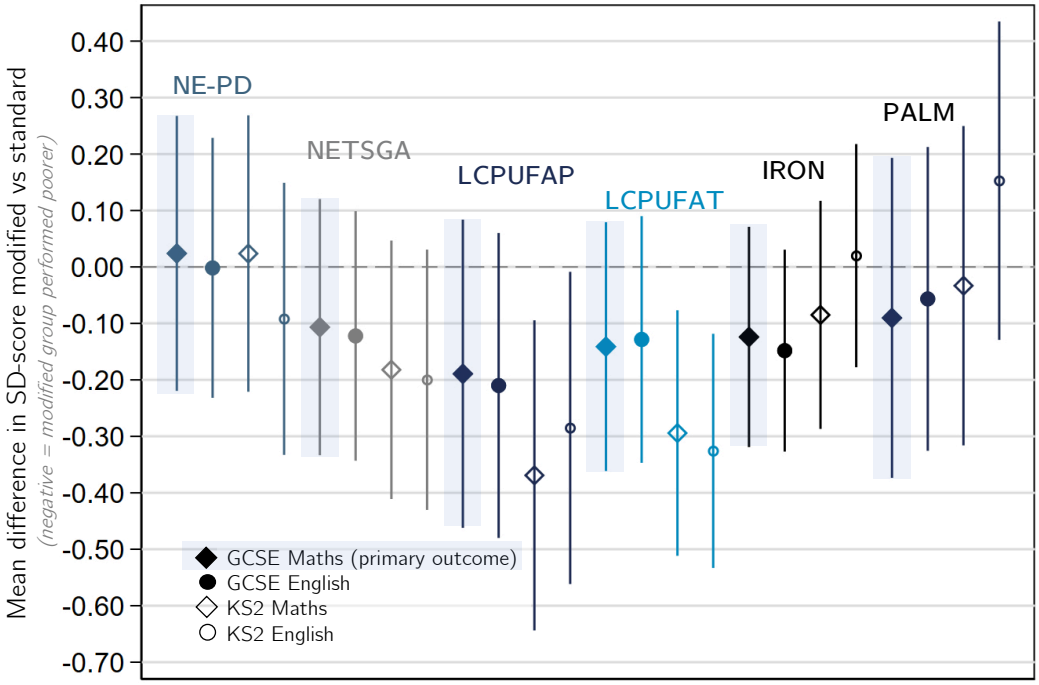


Fig. 8.3: Graph showing primary and secondary analysis results: mean differences in internally standardised grade scores in modified vs standard formula groups (negative mean difference = modified formula group performed less well compared to trial average score)

8.3.2 Secondary analyses

Secondary outcomes were generally consistent with the results for the primary outcome. Performance was significantly reduced at the 5% level among the modified formula groups for four outcomes in two RCTs: preterm and term infants randomised to LCPUFA had reduced Maths scores at 11y (preterm -0.37 SD; 95% Confidence Interval: -0.64 to -0.09; and term: -0.29 SD, -0.51 to -0.08) and reduced English scores at age 11 years (preterm -0.29 SD, -0.56 to -0.01; and term -0.33 SD, -0.53 to -0.11).

There was no difference between trial arms in the risk of qualifying for special educational needs support or attaining 5+ GCSE grades ≥C in any of the trials, see tables below.

Table 8.8: Secondary outcomes in the NEP-PD trial

NEP-PD N_I=112, N_C=115	Nutrient-enriched formula		Standard formula			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	0.00	0.87	0.00	0.90	0.00	-0.23, 0.23
KS2 Maths (age 11)	0.01	0.96	-0.01	0.92	0.02	-0.22, 0.27
KS2 English (age 11)	-0.05	0.91	0.05	0.93	-0.09	-0.33, 0.15
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					1.29	0.72, 2.32
5+ GCSE grades ≥C					1.27	0.70, 2.29

Footnotes: GCSE *General Certificate of Secondary Education*; NEP-PD *nutrient-enriched preterm post-discharge trial*; N_E *Number of participants in the enriched formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

Table 8.9: Secondary outcomes in the NETSGA trial

NETSGA N_I=152, N_C=147	Nutrient-enriched		Standard formula			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	-0.06	0.97	0.06	0.94	-0.12	-0.34, 0.10
KS2 Maths (age 11)	-0.09	1.01	0.09	0.98	-0.18	-0.41, 0.05
KS2 English (age 11)	-0.10	0.98	0.10	1.01	-0.20	-0.43, 0.03
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					1.49	0.90, 2.47
5+ GCSE grades ≥C					1.00	0.60, 1.71

Footnotes: GCSE *General Certificate of Secondary Education*; NETSGA *nutrient-enriched term small-for-gestational-age trial*; N_E *Number of participants in the enriched formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

Table 8.10: Secondary outcomes in the LCPUFAP trial

LCPUFAP N_I=92, N_C=100	LCPUFA formula (0.17% DHA and 0.31% AA/ total fat)		Standard preterm formula (no DHA or AA)			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	-0.11	0.88	0.10	0.93	-0.21	-0.48, 0.06
KS2 Maths (age 11)	-0.19	0.92	0.18	0.96	-0.37	-0.64, -0.09
KS2 English (age 11)	-0.15	0.96	0.14	0.91	-0.29	-0.56, -0.01
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					1.34	0.68, 2.64
5+ GCSE grades ≥C					0.65	0.32, 1.31

Footnotes: GCSE *General Certificate of Secondary Education;* LCPUFAP *long-chain polyunsaturated fatty acid supplemented formula for preterm babies trial*; N_L *Number of participants in the LCPUFA formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

Table 8.11: Secondary outcomes in the LCPUFAT trial

LCPUFAT N_I=155, N_C=154	LCPUFA formula (0.17% DHA and 0.31% AA/ total fat)		Standard preterm formula (no DHA or AA)			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	-0.06	0.97	0.06	0.94	-0.13	-0.35, 0.09
KS2 Maths (age 11)	-0.15	0.99	0.15	0.93	-0.29	-0.51, -0.08
KS2 English (age 11)	-0.16	0.93	0.16	0.93	-0.33	-0.53, -0.11
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					1.29	0.78, 2.14
5+ GCSE grades ≥C					0.69	0.41, 1.16

Footnotes:: GCSE *General Certificate of Secondary Education;* LCPUFAT *long-chain polyunsaturated fatty acid supplemented formula for term babies trial*; N_L *Number of participants in the LCPUFA formula group*; N_S *Number of participants in the standard formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

Table 8.12: Secondary outcomes in the IRON trial

IRON N_I=162, N_C=165	High-iron (12mg/dl) formula		Low iron (0.9mg/dl) formula			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	-0.08	0.81	0.07	0.85	-0.15	-0.33, 0.03
KS2 Maths (age 11)	-0.04	0.94	0.04	0.92	-0.08	-0.29, 0.12
KS2 English (age 11)	0.01	0.93	-0.01	0.90	0.02	-0.18, 0.22
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					1.32	0.80, 2.18
5+ GCSE grades ≥C					1.30	0.67, 2.52

Footnotes: GCSE *General Certificate of Secondary Education;* IRON *iron trial*; N_{HI} *Number of participants in the high-iron formula group*; N_{LI} *Number of participants in the low-iron formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

Table 8.13: Secondary outcomes in the negative control trial PALM

PALM N_I=103, N_C=100	Sn-2 palmitate formula		Standard palmitate formula			
Within-trial standardised grades:	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*95% CI*
GCSE English (age 16)	-0.03	0.95	0.03	0.93	-0.06	-0.33, 0.21
KS2 Maths (age 11)	-0.02	1.00	0.02	0.99	-0.03	-0.32, 0.25
KS2 English (age 11)	0.08	0.95	-0.08	1.03	0.15	-0.13, 0.43
Other secondary outcomes:					*Odds ratio*	*95% CI*
Ever qualified for special educational needs					0.81	0.42, 1.53
5+ GCSE grades ≥C					1.30	0.67, 2.52

Footnotes: GCSE *General Certificate of Secondary Education;* PALM *Sn-2 Palmitate trial*; N_P *Number of participants in the sn-2 palmitate formula group*; N_S *Number of participants in the standard palmitate formula group;* SD *standard deviation;* SMD *standardised mean difference;* adjusted for: infant sex, birth weight, gestational age, recruitment centre, maternal smoking during pregnancy, and maternal education at birth; covariates and outcomes imputed for missing participants who have not died.

8.4 Consistency of academic performance with earlier cognitive measures

8.4.1 Relative cognitive ability throughout childhood in modified formula vs standard formula groups

The results for academic performance were consistent with previously reported findings for cognitive ability at ages 18m (6 RCTs), 4-6y (2 RCTs), and 17y (2 RCTs) (**Fig. 8.4** to **Fig. 8.9**). There was no evidence of benefit of the modified infant formulas on cognitive ability. IQ at age 6 and 17 years was statistically significantly reduced in the LCPFUAT trial. All outcome measures presented in these figures are standardised to the trial population, use multiple imputation to deal with missing data, and adjust for baseline covariates associated with cognitive ability for statistical efficiency. It is important to note here that overlapping 95% confidence intervals do not necessarily indicate the absence of statistically significant differences between groups (Goldstein and Healy, 1995). The intervention group (black line, square markers) appears to perform worse than the standard group (blue line, round markers) in all trials except for the NEP-PD trial, where there was no consistent difference.

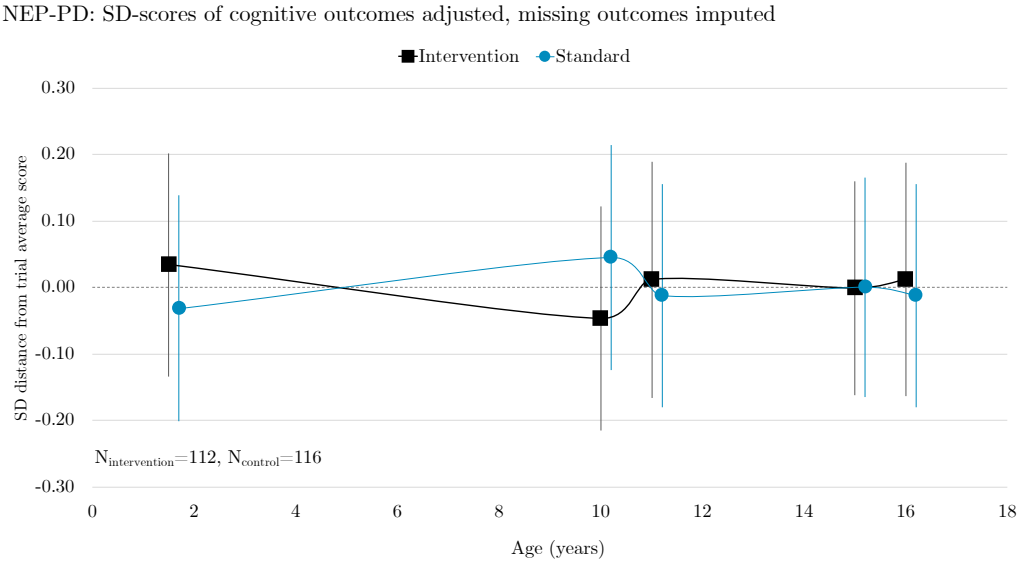


Fig. 8.4: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group in the NEP-PD trial

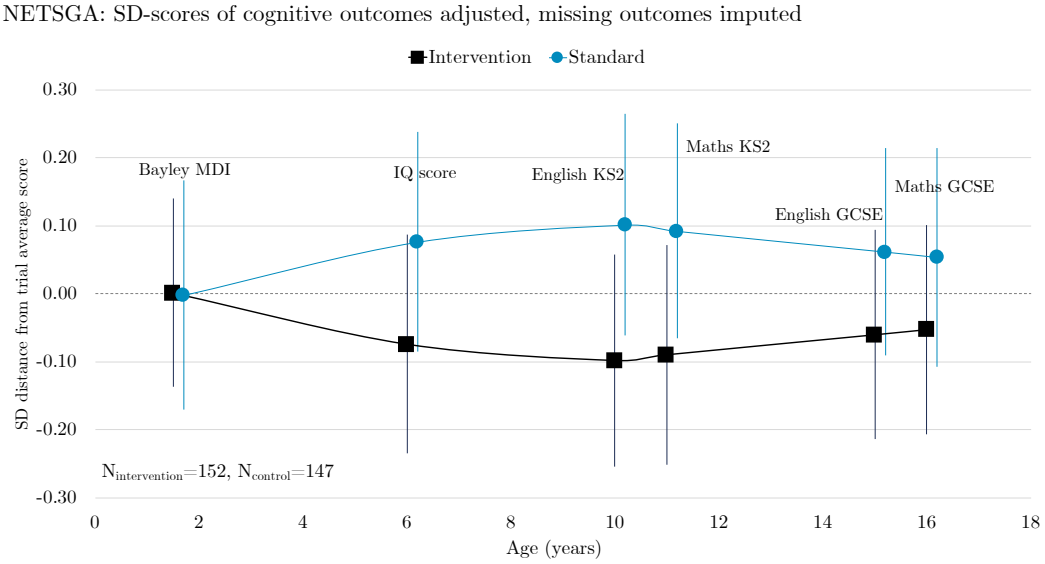


Fig. 8.5: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group in the NETSGA trial

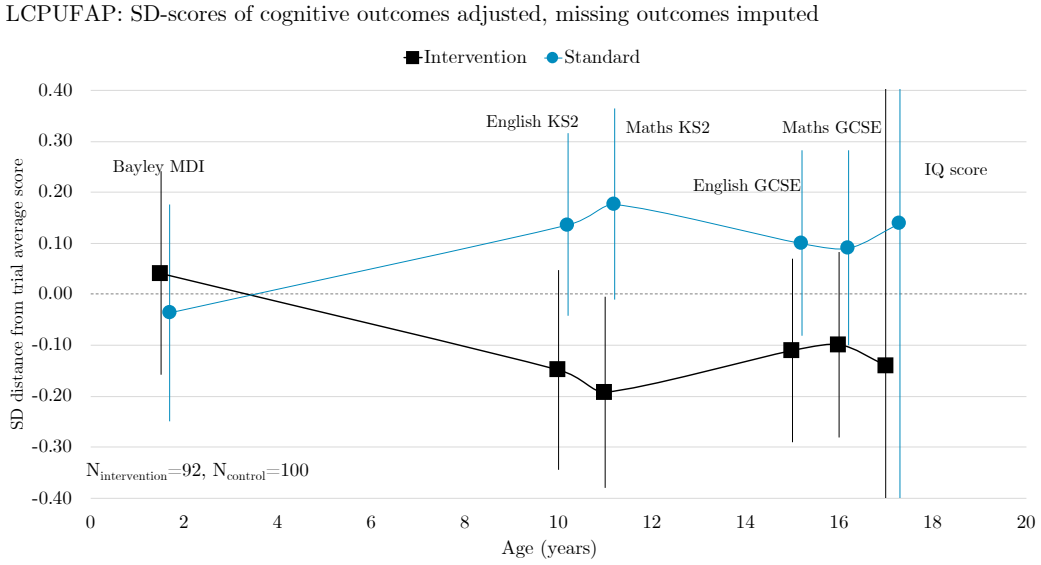


Fig. 8.6: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group group in the LCPUFAP trial

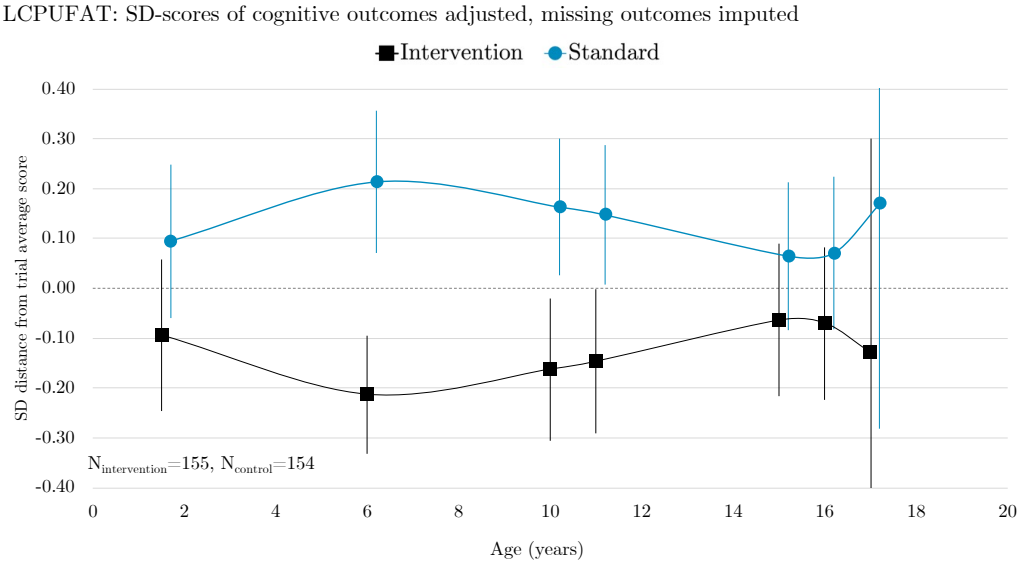


Fig. 8.7: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group in the LCPUFAT trial

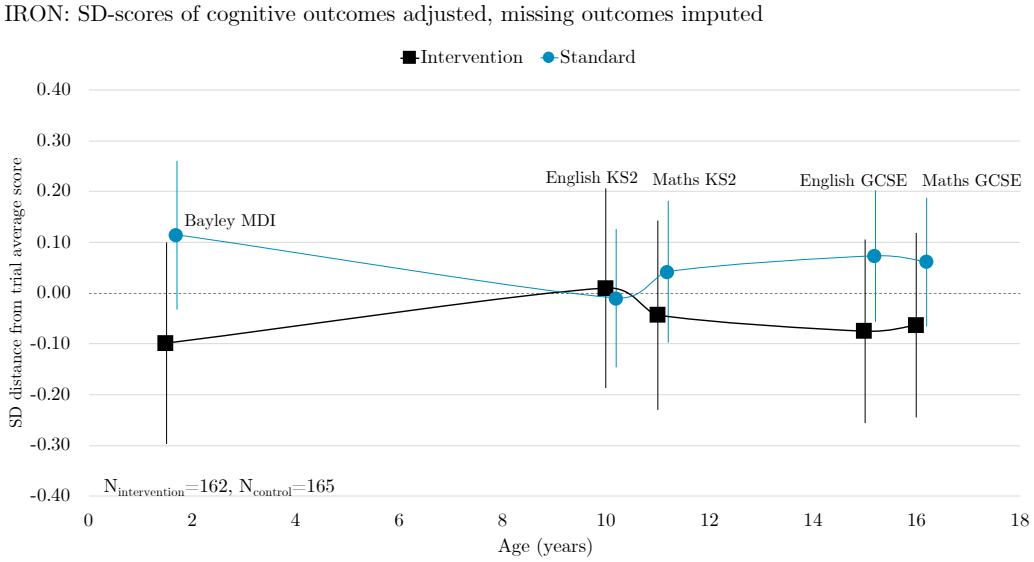


Fig. 8.8: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group group in the IRON trial

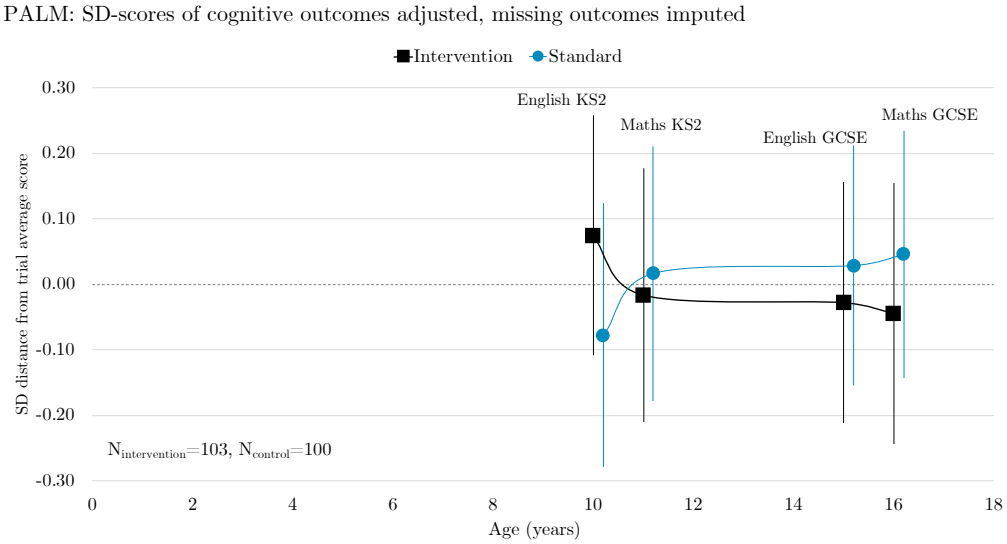


Fig. 8.9: Within-trial standardised mean scores for measure of cognitive ability at each follow up (95% Confidence Interval) in modified vs standard formula group in the PALM trial

8.4.1.1 Predictive ability of early cognitive tests

Table 8.14 below shows the correlation matrix for cognitive outcomes through childhood in all trials combined. All correlations were positive. It can be seen that the Bayley Mental Development Index only has a weak positive correlation with later cognitive outcomes, with correlations progressively weakening throughout childhood. Maths performance at age 16 years is slightly more strongly correlated with Maths performance at age 11 years than with English at age 16 years, providing some support to the idea that Maths and English measure overlapping but not identical aspects of cognitive ability. Furthermore, IQ score is more strongly correlated with Maths performance than with English performance at both ages. These observations were confirmed when adjusted for confounders (see next section) The highest correlation is between Maths at age 11 and Maths at age 16 (0.77), and the lowest between Bayley MDI at age two and IQ at age 17 (0.07).

Table 8.14: Pearson's correlation coefficient between cognitive measures

	IQ 4-6y	English 11y	Maths 11y	English 16y	Maths 16y	IQ 17y
Bayley MDI 2y	0.3711 *n=199*	0.3342 *n=946*	0.2671 *n=965*	0.3012 *n=849*	0.2672 *n=977*	0.0728 *n=53*
IQ 4-6y	1 *n=307*	0.5296 *n=193*	0.4684 *n=195*	0.5157 *n=195*	0.491 *n=198*	0.4669 *n=36*
English 11y		1 *n=1518*	0.699 *n=1510*	0.5945 *n=1151*	0.6062 *n=1314*	0.5664 *n=51*
Maths 11y			1 *n=1543*	0.5692 *n=1167*	0.7731 *n=1332*	0.7162 *n=52*
English 16y				1 *n=1221*	0.7495 *n=1208*	0.6231 *n=53*
Maths 16y					1 *n=1421*	0.7533 *n=54*
IQ 17y						1 *n=120*

Legend: 1 *Perfect positive* | 0.8 | 0.4 | 0 *No correlation* | -0.4 | -0.8 | -1 *Perfect negative*

Footnotes: Data from all trials combined

Adjusted regression coefficients estimating the predictive ability of early cognitive measures for later cognitive measures are given in **Table 8.15**. Earlier cognitive measures predicted later cognitive measures independent of infant sex, birth weight, gestational age, maternal education, maternal smoking during pregnancy, and trial. The predictions were statistically significant at the 5% level except for Bayley MDI on IQ at age 17 years. For a 1 SD increase in Bayley MDI, later cognitive ability measures are predicted to increase between 0.2 and 0.4 SDs. For a 1 SD increase in IQ at age 4-6 years, later cognitive ability measures are predicted to increase between 0.4 and 0.5 SD. For a 1 SD increase in English and Maths at age 11 years, later cognitive ability measures are predicted to increase between 0.5 and 0.7 SD. A 1 SD increase in English and Maths performance at age 16 years predicts an increase of IQ at age 17 years between 0.6 and 0.8 SD.

Table 8.15: Estimated regression coefficients of internally standardised cognitive scores (within-trial), 95% confidence intervals, standard errors (SE), p-values from Wald tests and observations (N) from linear regression models

	IQ 4-6y (outcome)	English 11y (outcome)	Maths 11y (outcome)	English 16y (outcome)	Maths 16y (outcome)	IQ 17y (outcome)
Bayley MDI 2y (predictor)	0.396 95%CI: 0.25, 0.54 SE: 0.0716 p-value: <0.001 N: 191	0.311 95%CI: 0.24, 0.38 SE: 0.035 p-value: <0.001 N: 865	0.256 95%CI: 0.18, 0.33 SE: 0.037 p-value: <0.001 N: 882	0.215 95%CI: 0.14, 0.29 SE: 0.039 p-value: <0.001 N: 770	0.207 95%CI: 0.14, 0.28 SE: 0.036 p-value: <0.001 N: 893	0.209 95%CI: - 0.20, 0.62 SE: 0.204 p-value: 0.311 N: 51
IQ 4-6y (predictor)		0.548 95%CI: 0.44, 0.66 SE: 0.056 p-value: <0.001 N: 185	0.460 95%CI: 0.34, 0.58 SE: 0.060 p-value: <0.001 N: 187	0.479 95%CI: 0.34, 0.58 SE: 0.070 p-value: <0.001 N: 187	0.439 95%CI: 0.33, 0.55 SE: 0.055 p-value: <0.001 N: 190	0.531 95%CI: 0.19, 0.87 SE: 0.164 p-value: 0.003 N: 35
English 11y (predictor)			0.687 95%CI: 0.64, 0.73 SE: 0.020 p-value: <0.001 N: 1,265	0.522 95%CI: 0.47, 0.58 SE: 0.028 p-value: <0.001 N: 955	0.560 95%CI: 0.51, 0.61 SE: 0.026 p-value: <0.001 N: 1,094	0.652 95%CI: 0.45, 0.85 SE: 0.098 p-value: <0.001 N: 49
Maths 11y (predictor)				0.520 95%CI: 0.47, 0.57 SE: 0.027 p-value: <0.001 N: 967	0.713 95%CI: 0.68, 0.75 SE: 0.020 p-value: <0.001 N: 1,108	0.718 95%CI: 0.55, 0.88 SE: 0.081 p-value: <0.001 N: 50
English 16y (predictor)					0.731 95%CI: 0.68, 0.78 SE: 0.026 p-value: <0.001 N: 1,002	0.597 95%CI: 0.37, 0.82 SE: 0.111 p-value: <0.001 N: 51
Maths 16y (predictor)						0.770 95%CI: 0.59, 0.95 SE: 0.087 p-value: <0.001 N: 52

Footnotes: adjusted for: infant sex, birth weight, gestational age, maternal education, maternal smoking during pregnancy, and trial. Colours indicate the strength of association: blue = strong positive, red=strong negative, green=medium positive, yellow=weak positive, orange=weak negative.

8.5 Discussion

8.5.1 Summary of findings

I found no beneficial effects of any of the four types of nutritionally modified infant formulas on academic performance. There was weak evidence that children, who as babies were randomised to LCPUFA-supplemented infant formula, performed worse in English and Maths at age 11 years. Sensitivity analyses suggested that findings were robust to covariate adjustment, method of handling missing data, and standardisation reference distribution. Academic performance was consistent with previously collected cognitive outcomes.

8.5.2 Findings from this chapter in relation to previous evidence

8.5.2.1 Nutrient-enriched post-discharge formula in preterm infants

Infants born preterm have limited nutrient reserves at birth and have different nutritional requirements compared to term infants, which are often not met during the in-hospital period. Nutrient-enriched post-discharge formulas are intended to meet these additional nutritional requirements after hospital discharge to support catch-up growth and development. The use of nutrient-enriched post-discharge formulas was hypothesised to translate into cognitive benefits, compared to those fed standard term formula post-discharge (Embleton et al., 2021). Chapter 6 presented evidence from RCTs indicating that feeding preterm infants with nutrient-enriched formula after hospital discharge neither benefitted short-term growth nor short-term development compared to standard term formula. This chapter addresses a gap in evidence on long-term cognitive outcomes of infants born preterm who were randomised to nutrient-enriched formula after hospital discharge compared to those randomised to standard term formula after hospital discharge. Results from this chapter show that post-discharge formulas did not have a measurable long-term benefit on cognitive ability. A potential mechanism for this lack of benefit is that the intervention took place during the period after hospital discharge when the brain is believed

to be less sensitive to nutrient supply compared to the period before term (in-hospital period for babies born preterm) (Colombo et al., 2019). Another, less likely, mechanism for the absence of benefit is that infants could have adjusted their volume of intake according to the energy density of the formula and therefore consumed a similar total amount of nutrients to infants fed the standard term formula. This could have been possible because the infants were fed in response to their hunger and satiation cues. While volume of formula intake was not measured in this trial, evidence from other trials using the same formula does not support this mechanism (Lucas et al., 2001).

8.5.2.2 Nutrient-enriched infant formula in term SGA infants

Term infants born small for gestational age are at increased risk for poor cognitive outcomes (Ido et al., 1995), but to date, there is no evidence that feeding nutrient-enriched formula after birth reduces this risk compared to feeding standard term formula. Previously published evidence, based on two RCTs (Singhal et al., 2010b) – one of which was the RCT analysed in this chapter (Fewtrell et al., 2001), found that feeding enriched formula to SGA term infants led to catch-up growth by increasing length and weight gain. While no benefits on short-term developmental measures were found, it was anticipated that cognitive benefits might be detected by more sensitive measures of cognitive development such as IQ assessment later in childhood. Yet, high rates of participant drop-out made IQ findings challenging to interpret. What is more, these trials found adverse effects on body fat and blood pressure in children previously randomised to the enriched formula. Therefore, urgent evidence was needed to assess whether the metabolic risk associated with faster early growth in these infants is counterbalanced by any longer-term cognitive benefit. The data presented in this chapter contribute important findings to this debate. With low attrition rates, they provide firm evidence that the metabolic risk in these infants is not counterbalanced by any longer-term cognitive benefit as measured by academic performance nor by a reduction in the proportion of children qualifying for special educational needs support. Taken together, no evidence of cognitive benefit and potential for metabolic harm thus supports current

recommendations against promoting catch-up growth in term infants born SGA (Singhal, 2015).

8.5.2.3 LCPUFA-supplemented infant formula in infants born preterm or at term

LCPUFAs such as docosahexaenoic acid (DHA) and arachidonic acid (AA) play a structural role in brain and eye development, and it was hypothesised that adding preformed LCPUFAs to infant formulas would improve visual and cognitive outcomes for formula fed infants (Gibson and Makrides, 2001). LCPUFAs mainly accumulate during the third trimester of pregnancy and early infancy. Breast milk contains DHA and AA, while, historically, infant formula contained only their precursor molecules. This generated concern that formula-fed infants were at risk of developmental deficiencies (Neuringer and Connor, 1986, Woods et al., 1996). If infants cannot synthesise enough DHA and AA from the precursors supplied in formula, the concern was that the lack of preformed LCPUFA in formula would lead to an inadequate supply of these nutrients resulting in sub-optimal brain development. This concern applied particularly to preterm infants, who are not only born with fewer LCPUFA reserves but may also have less capacity to convert fatty acid precursors into DHA and AA, compared to term infants. In addition, it was hypothesised that preterm infants who often undergo periods of negative energy balance might use the precursor fatty acids as an energy supply rather than metabolising the precursors to DHA and AA (Lapillonne et al., 2013).

However, evidence for supplementation with LCPUFAs remained controversial: Cochrane reviews with meta-analyses reported no benefits for vision or measures of cognition in term or preterm infants (Jasani et al., 2017, Moon et al., 2017). Yet, these reviews focused on early (<age 2.5 years) cognitive measures, which were already argued to be too noisy to detect any benefit (Sun et al., 2015). My own systematic review therefore focused on long-term cognitive outcomes such as IQ scores in preterm and term infants. My review was consistent with no cognitive benefit of LCPUFA-supplemented infant formulas, but all included studies were severely limited by attrition (Verfuerden et al.,

2020). By linking trials investigating LCPUFA supplemented infant formula in term and preterm infants to school data without the need for participants to consent, this chapter addresses both of these gaps. I obtained a well-validated, real-world outcome with meaningful predictive properties for adult life and high linkage rates. Together with previous evidence, the results strongly indicate that substantial benefits do not emerge in childhood. Instead, I found weak evidence of harm in children randomised to LCPUFA supplemented formulas, with lower English and Maths scores at age 11 in both preterm and term infants.

It is unclear why LCPUFA-supplemented infant formula might adversely affect academic performance. DHA content of human milk is variable and heavily influenced by maternal diet (notably fish consumption) (Aumeistere et al., 2018). This natural variability in breast milk makes the optimal dose of DHA in infant formula uncertain. Furthermore, it is likely that LCPUFAs derived from other sources than human breast milk, and in isolation from other components present in human breast milk, such as β-carotene (Zielinska et al., 2019), have different biological properties compared to LCPUFAs naturally occurring in human breast milk. The LCPUFAs in the linked studies were derived from egg, which inevitably involves adding other unintended components along with the DHA and AA. Given potential associations between LCPUFA source and cognitive outcomes, long-term follow-up of trials testing infant formulas with other LCPUFA sources is recommended.

The findings in this chapter, are particularly important in light of the recent mandate to add one type of LCPUFA, DHA, to all infant and follow-on formulas in the EU (EU Commission, 2016). Combined with previous evidence the findings should prompt the reappraisal of such legislation as a mandate might not only have the potential for considerable harm but also it also inhibits future research by limiting equipoise.

8.5.2.4 High-iron follow-on formula in healthy term infants

Iron-fortification of follow-on formulas is used to reduce risks of iron deficiency during the complementary feeding period when infants have high requirements for normal growth and development. Iron deficiency is associated with adverse

effects on brain development and cognition (Halterman et al., 2001, Lozoff and Georgieff, 2006). However, iron is also a potent pro-oxidant, and non-absorbed iron in the gut may have adverse effects on the microbiome (Jaeggi et al., 2015). This means that, on balance, iron supplementation of iron-replete infants may have adverse effects such as increased risk of infection and impaired growth. One trial reported adverse effects of fortified follow-on formula on cognitive outcomes at ten years and 16 years (Gahagan et al., 2019). However, the interpretation of these findings remains uncertain due to significant participant attrition.

In this chapter, I compared healthy term infants randomised to high-iron follow-on formula (12 mg/l) with those randomised to low-iron formula (0.09 mg/l) and found no benefit of iron supplementation for academic performance. One might criticise that baseline iron status was not assessed in the IRON trial. However, this reflects the reality of formula purchasing decisions, where most parents are unaware of their infant's iron status. Also, just like most infants in the UK (McAndrew et al., 2012), the majority of participants in the IRON trial had been fed iron-fortified infant formulas before randomisation. This means they were a 'low risk' group for iron deficiency at that point. The finding that high-iron fortification of follow-on formula is neither associated with significant cognitive benefits nor harms contributes important data to an evidence gap highlighted by the European Food Safety Authority and the European Society for Paediatric Gastroenterology Hepatology and Nutrition. These organisations so far refrained from drawing conclusions on the optimum or maximum iron content in follow-on formulas for healthy infants because they considered the evidence on cognitive effects associated with iron supplementation to be too limited (EFSA Panel on Dietetic Products and Allergies, 2014, Domellöf et al., 2014).

8.5.2.5 Negative control association: infant formula with palmitate in the sn-2 position in healthy term infants

In human milk, palmitate (the primary source of saturated fatty acids) is formed so that most triglycerides are attached to the sn-2 position (Innis, 2011). In contrast, the majority of palmitate supplied in cows' milk-based infant formula

is usually attached to the 1 or 3 position. This affects how palmitate is metabolised, influencing the digestion and absorption of calcium (Miles and Calder, 2017). Because it is biologically unlikely that the addition of palmitate to infant formula in the sn-2 position influences cognitive ability and there was no evidence of cognitive effects from RCTs (Chapter 6), the PALM trial was included as a negative control trial. As expected, no association was found between formula with palmitate in the sn-2 position and academic performance. Given that the other trials also found no clear association, the negative control carries only limited information but provides weak evidence that my methods did not bias effect sizes upwards or downwards.

8.5.3 Risk of bias from participant attrition

In this study, Maths grades at age 16 years were ascertained for an average of 86% of randomised participants. In contrast, previous conventional follow-ups of the analysis trials ascertained cognitive outcomes for only 21.95% of trial participants after the age of two years (**Table 2.5**, page 47).

In addition, observed participant characteristics in this study were balanced, indicating no major differences between treatment groups. Sensitivity analyses also indicated that the analyses were robust to different missing data assumptions. Together these observations indicate that my analysis findings are unlikely to be significantly biased by participant attrition.

8.5.4 Predictive ability of early cognitive measures

Multiple previous studies have indicated that early developmental measures, such as Bayley Scores, are poor predictors of later cognitive development (Sun et al., 2015, Colombo, 2018). Poor agreement between early and late measures was also seen in previous cognitive follow-ups of the trials in this thesis; notably the LCPUFA trials. IQ scores at age 4-6 years and at age 17 years found evidence of harms associated with the modified formula, while Bayley Scores at age 18 months found no difference in the LCPUFAT trial and a slight benefit in the LCPUFAP trial (see **Table 2.5** in Chapter 2 and section 8.4.1). Until now it

remained uncertain whether early measures are innately poor predictors of later cognitive ability, or whether the disagreement between early and later measures stemmed from attrition bias in the later measures.

My study has now confirmed the presence of weak later harms at age 11 for LCPUFA in absence of elevated risk from attrition. It therefore seems reasonable to conclude that early developmental measures are intrinsically poor predictors of later cognitive ability and that disagreement is unlikely to stem from attrition.

There are two potential reasons for why short-term developmental measures might not reflect later cognitive outcomes: a) they are inherently too noisy, or b) it takes several years for cognitive effects to emerge and manifest themselves. Both reasons would support the need to monitor the cognitive effects of infant formula modifications beyond infancy to confirm the absence of harm or presence of benefits. This conclusion could not have been generated without unconsented linkage to administrative data, demonstrating the substantial added value of the method.

8.5.5 Study strengths

The study presented in this chapter had several strengths. Firstly, the analyses of academic performance were based on randomised controlled trials with good internal and external validity. Randomisation minimised the risk of confounding, so it was in a position to investigate causal relationships between modified infant formula and academic performance. Results from the negative control trial provided weak evidence that the methods used in this PhD study did not systematically bias the associations upwards or downwards.

Secondly, primary outcomes were ascertained for an average of 86% of participants aged 16 years (see paragraph on risk of attrition above). The high linkage rate in my study significantly increases the robustness of findings and reduces the likelihood of selection bias through attrition.

Thirdly, all outcomes and analyses for this study were prespecified and prospectively registered and peer-reviewed, further minimising risk of bias during the analysis and reporting phase. The methods and results were reported

according to the CONSORT checklist, ensuring that study plans and assumptions were presented clearly, in advance, and were open for critique.

Finally, the primary and secondary outcomes are well-validated by their replicability and predictive value for earning and achievement trajectories in adulthood. The exam boards were blinded to the participant's treatment allocation, making it unlikely that any measurement error depends on the formula group. Furthermore, GCSE Maths grades are less biased, more accurate and more relevant for children's futures than IQ tests (see section 1.2.4, page 24 and section 3.4.1.1, page 56 for discussion). Using unconsented linkage to mandatory high-stakes national school exams, this study allowed for an unbiased ascertainment of long-term cognitive ability.

8.5.6 Study limitations

Several limitations apply to this study. The first is related to the use of data linkage to ascertain outcomes. Linkage error (missed matches and false matches) can introduce bias. All records in the five analysis trials were expected to link (aside from a small percentage participants who died, could have emigrated, or visited schools not interacting with the NPD). I therefore had information on the number of missed matches within each trial. **Table 5.8** in Chapter 5 showed that the rate of missed NPD matches was comparatively low (between 3% and 13% depending on the trial). Missed matches were not statistically significantly different from the initially randomised samples based on a range of key characteristics. There was also no evidence that missed matches depended on randomised group (**Table 5.9**). Assuming that observed participant characteristics also reflect unobserved characteristics, the most likely impact of missed matches was, therefore, a reduction in the statistical power. Quantifying the impact of false matches, that is participants who linked to the wrong pupil record, is harder. Two factors reduced the likelihood of false matches in the analysis sample. First, the fact that there were no links considered to be false matches among negative control participants, and second, that I excluded additional links I considered implausible based on agreement patterns and match weights. Given that all parties involved in ascertaining the outcome were blinded

to the formula allocation, it is also unlikely that false matches depended on the formula group. Any false matches unconditional on the formula allocation are likely to introduce random noise and bias the associations towards zero. However, the consistency of academic performance with previously measured cognitive outcomes and findings from external literature provide support for the credibility of the observed null-associations.

The second limitation is that random error in the precision of estimates could have increased the risk of type I and II errors. Some trial data showed heteroscedasticity (the variance of residuals was not constant around the regression line); this led me to use robust standard errors in all regression analyses. Robust standard errors are generally larger than non-robust standard errors leading to wider confidence intervals. To test how sensitive results were to heteroscedastic data, I compared the estimates to those obtained through ordinal logistic regression (data available on request), which does not assume homoscedasticity and therefore does not require robust standard errors. Findings from ordinal logistic regression were near identical and did not change the conclusions of the analyses. This suggests that heteroscedasticity was unlikely to have introduced substantial bias.

Finally, the trials were conducted several decades ago, so one might criticise the generalisability of the findings to present-day infants. While the composition of the formulas has not changed greatly, infant characteristics and care settings have changed (Stoll et al., 2015). Nowadays, a larger number of sick and small preterm infants survive (Glass et al., 2015), and these infants may have different sensitivities to the nutritional modifications investigated in this chapter. The NEP-PD, NETSGA and LCPUFAP trial included infants born as young as 25 weeks and as light as 630 grams. Exploratory subgroup analyses conducted for this thesis (appendix p 50), although underpowered, provide weak evidence that the benefit from the modified formulas was not different for more premature infants. While I found no later cognitive benefit of nutrient-enriched post-discharge formula in preterm infants, it is important to note that this does not exclude a role for this intervention in treating or preventing the significant undernutrition often seen in preterm infants at the time of hospital discharge and beyond.

8.5.7 Implications for regulators and manufacturers

Findings from this chapter show that large benefits of the tested formula modifications on cognition are highly unlikely. Nevertheless, small benefits or harms associated with global use of infant formula modifications could still have public health impact at a population level. However, to detect or exclude benefits or harms smaller than possible with the data used in this thesis, say of 0.2 SD, with 80% probability 95% confidence and σ of 0.95, RCTs with long-term outcome data on at least 710 participants would be needed. For now, based on the combined evidence of this thesis and previously published studies, claims of cognitive benefits by manufacturers (see **Fig. 8.10** on the next page for examples) are not justified.

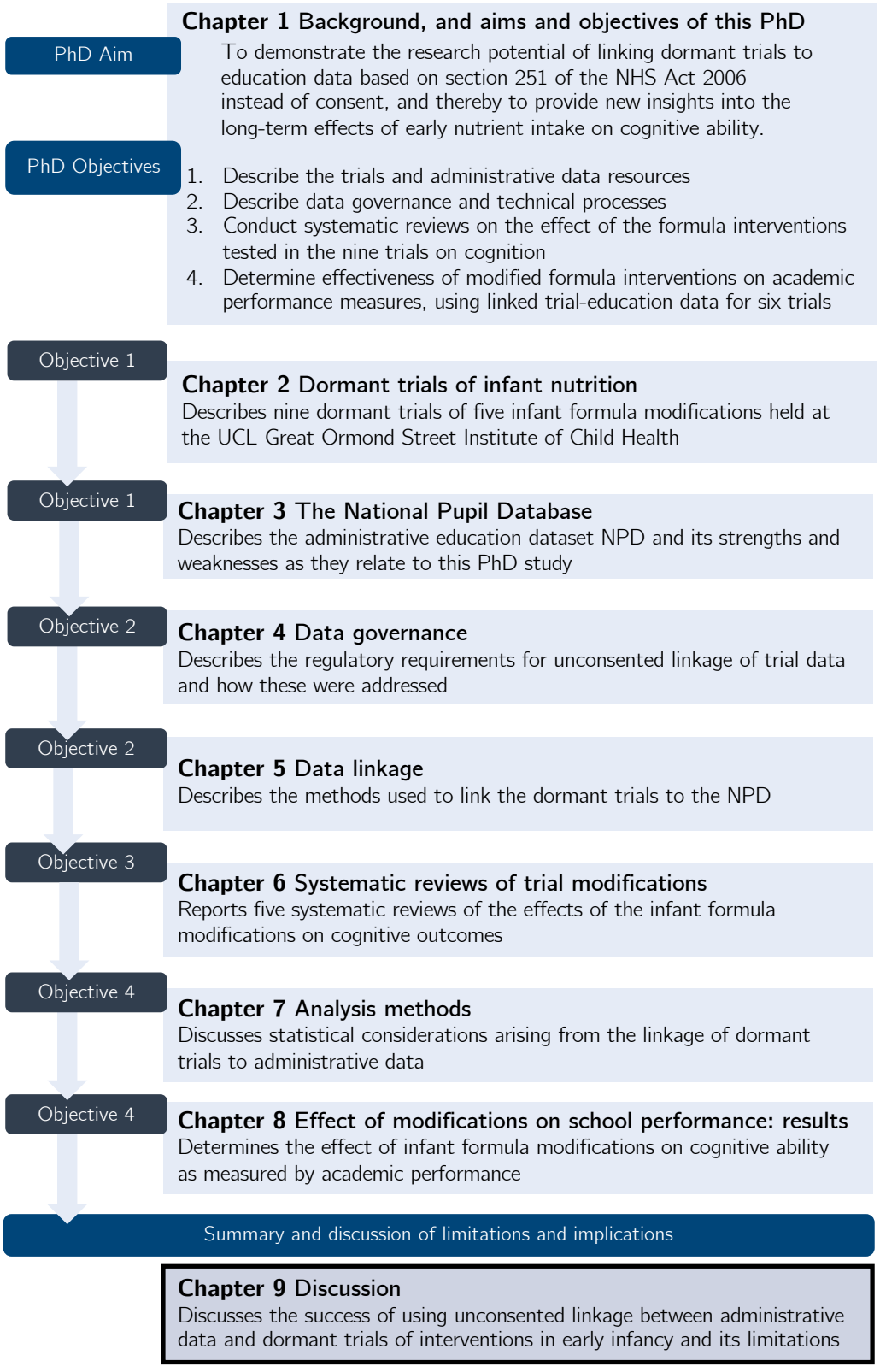
8.5.8 Conclusion

The nutritional composition of infant formula has been proposed to have life-time effects on cognitive development (Lucas, 2005, Lucas et al., 1998). This is the first time, to my knowledge, that a series of infant formula trials has been linked to an administrative education database to investigate this hypothesis with minimal participant attrition. Findings from this chapter suggest that a large benefit of the investigated infant formula modifications on academic performance is highly unlikely. They provide weak evidence of harm for one type of modification. Marketing of these formulas for cognitive benefit is not supported by the available evidence.

Fig. 8.10: Claims that formula modifications support cognitive development and learning ability

8.6 Key points from Chapter 8

- I determined the effects of five types of modified infant formula on academic performance using linked trial-school data with complete primary outcome data on 86% of participants.
- The differences in academic performance between modified and standard formulas were consistent with differences measured in the original trials, and external literature, in failing to find a benefit of the infant formula modifications on cognitive outcomes.
- There is weak evidence that term and preterm-born infants randomised to LCPUFA-supplemented infant formula performed worse in Maths and English at age 11 (secondary outcomes).
- The reported data, for the first time, show the long-term cognitive effects of randomising infants to nutritionally modified formula without significant participant attrition.
- Further trials on the cognitive effects of these formula modifications intended to analyse subgroups or to exclude smaller benefits or harms would require much larger participant numbers and therefore significant resources.


Chapter 1 Background, and aims and objectives of this PhD
PhD Aim
To demonstrate the research potential of linking dormant trials to education data based on section 251 of the NHS Act 2006 instead of consent, and thereby to provide new insights into the long-term effects of early nutrient intake on cognitive ability.
PhD Objectives
1. Describe the trials and administrative data resources
2. Describe data governance and technical processes
3. Conduct systematic reviews on the effect of the formula interventions tested in the nine trials on cognition
4. Determine effectiveness of modified formula interventions on academic performance measures, using linked trial-education data for six trials
Objective 1
Chapter 2 Dormant trials of infant nutrition
Describes nine dormant trials of five infant formula modifications held at the UCL Great Ormond Street Institute of Child Health
Objective 1
Chapter 3 The National Pupil Database
Describes the administrative education dataset NPD and its strengths and weaknesses as they relate to this PhD study
Objective 2
Chapter 4 Data governance
Describes the regulatory requirements for unconsented linkage of trial data and how these were addressed
Objective 2
Chapter 5 Data linkage
Describes the methods used to link the dormant trials to the NPD
Objective 3
Chapter 6 Systematic reviews of trial modifications
Reports five systematic reviews of the effects of the infant formula modifications on cognitive outcomes
Objective 4
Chapter 7 Analysis methods
Discusses statistical considerations arising from the linkage of dormant trials to administrative data
Objective 4
Chapter 8 Effect of modifications on school performance: results
Determines the effect of infant formula modifications on cognitive ability as measured by academic performance
Summary and discussion of limitations and implications
Chapter 9 Discussion
Discusses the success of using unconsented linkage between administrative data and dormant trials of interventions in early infancy and its limitations

CHAPTER 9 Discussion

9.1 Chapter content

This concluding chapter recalls the state of the evidence before this thesis and summarises my account of the contributions my thesis made. It also discusses the wider implications of the work presented in this thesis and proposes avenues for further research.

9.2 Rationale and thesis aim

Due to participant attrition and funding constraints, international infant formula guidelines largely rely on RCT evidence with short-term endpoints. However, short-term endpoints for cognitive ability poorly predict the long-term efficacy and safety of formula modifications. Harms and benefits may emerge later or may not be detectable using early developmental measures. The aim of this thesis was twofold: to demonstrate the processes and added value of linking a series of nine dormant trials without participant consent to administrative records to measure long-term intervention effects, and by doing this, to determine the effect of nutritionally modified infant formulas on long-term cognitive ability as measured by academic performance.

9.3 Thesis objectives

The section below outlines the key findings and conclusions of this PhD thesis under each of the four objectives.

9.3.1 Objective 1: Describe the trial and administrative data resources

For objective 1, I described the trial data available for this PhD thesis and the National Pupil Database, an administrative data set that collects information on English pupils and their educational attainment. Challenges of using administrative data for linkage to trial data include that data collection periods of administrative datasets vary and may not completely cover the participants'

school trajectory. Before conducting the linkage, it is therefore important to verify that participants' school trajectories coincide with data collection periods of administrative school data. Participants in the early nutrient-enriched formula trials (NEP-1 and NEP-2) were too old, and some participants in the nucleotide enriched formula trial (NUCLEO) were too young to link to academic performance outcomes. Nevertheless, these trials were sent for linkage regardless to act as negative controls and proved useful in informing the assessment of match quality and adding information to multiple imputation algorithms. Other measures could become available for these trials in the future, such as GCSE records for the NUCLEO trial or the individual learning record or earnings data for NEP-1 and 2. Therefore, the data from the NEP-1, NEP-2, and NUCLEO trial was prepared, so that linkage of these three trials to available administrative datasets is imminently possible.

9.3.2 Objective 2: Describe data governance and technical processes

One key conclusion from this PhD study is that it is ethical, feasible and timely to use unconsented linkage of dormant trials to administrative data to determine long-term effects of early interventions as demonstrated by the approval of national ethics and data governance committees. Innovative use of administrative data for research is highly encouraged by the UK government (Lugg-Widger et al., 2018). Yet, data governance is often not transparent, with data holders, the confidentiality advisory group (CAG) and ethics committees relying on successful precedents to approve new applications. In receiving favourable CAG and ethics review and data holder approval, this PhD thesis was successful in establishing such a precedent. The acceptability and likely benefit of this project were also confirmed by former trial participants, who understood and appreciated the necessity of the approach (Chapter 4).

Another key finding is the importance of how identification data is stored in dormant trials. For data security purposes, identification data in the dormant trials was stored on paper in consent forms and notebooks, with a significant share of handwritten information. To reactivate dormant trials, it was therefore

necessary to digitise and review the accuracy of these identifiers. Repeated follow-ups, twins and, in rare cases, children who participated in multiple trials presented an opportunity to cross-check the validity of identifiers. The time and cost involved in the digitisation of identifiers should not be underestimated. For this thesis, the digitisation of child identifiers took ten months and incurred £37,966 in total staff costs. Still, the costs of unconsented follow-up through administrative records were much lower than those needed for conventional follow-up approaches, given the need to employ researchers, pay travel expenses and book venues (Llewellyn-Bennett et al., 2018).

The quality of identifiers in administrative datasets should also be considered. The National Pupil Database, which was used in this thesis to provide data on participants' academic performance, undergoes regular and well-documented data quality checks making it a useful data resource for long-term follow-up.

Another key conclusion falling under objective 2 is that data linkage can be conducted both transparently and securely by requesting multiple probable matches for each participant along with information on identifier agreement strength from the data linkage party. The best match can then be chosen in-house using probabilistic methods without the need to see identification data.

9.3.3 Objective 3: Conduct systematic reviews on the effect of the five formula modifications investigated in this thesis

In Chapter 6, I systematically reviewed the available evidence from RCTs of the effect of infant formula modifications (nutrient enrichment, LCPUFA, nucleotides, iron, and sn-2 palmitate) on measures of cognitive ability. I confirmed that high participant drop-out rates are not limited to the trials examined in this thesis but extend to all trials that investigated the long-term cognitive effects of these formulas. Where drop-out rates were low, such as in the measurement of short-term developmental progress, the outcomes were poorly predictive of later cognitive function and showed no cognitive benefits. As a consequence, there was a lack of evidence on the long-term harms or benefits for cognition on all of the examined interventions.

9.3.4 Objective 4: Determine the effectiveness of modified infant formulas for improving academic performance measures, using linked trial-education data for six dormant infant formula trials

The findings of the analysis presented in Chapter 8 were in agreement with outcomes measured in the original trials, and other RCT literature, in failing to detect any benefit of infant formula modification on cognitive outcomes. None of the infant formula modifications examined in this thesis benefitted the primary outcome of performance in GCSE Maths at age 16, and none improved any secondary outcomes of academic performance or reduced the likelihood of qualifying for special educational needs support. I found weak evidence that LCPUFA supplemented infant formula reduced performance for term and preterm born children for two secondary academic performance outcomes.

The upper confidence intervals of the effectiveness estimates exclude improvements larger than 0.27 SD in Maths performance at age 16 years for any of the modified infant formulas, with 5 out of 6 modifications excluding improvements larger than 0.12 SD. Further trials for these interventions would therefore have to spend significant resources to confirm or exclude small effects. For now, the combined evidence of my findings with previously published evidence suggests that the examined infant formula modifications should not be marketed on the basis of cognitive benefits in the studied populations.

9.4 Feasibility and generalisability of methods

I argue that using administrative data to extend early intervention studies without consent is highly feasible. This study sets a precedent for information governance regulators for further studies involving unconsented linkage to dormant trials or cohorts. The methods used in this PhD project were judged to be ethical and feasible, as demonstrated by the approval of national ethics and data governance committees, and are likely to generate public benefit. In terms of participant follow-up and speed of outcome retrieval, using unconsented linkage of administrative data outperformed traditional follow-up methods. The

key criteria for a successful replication of these methods in another research context are 1) the availability of good quality identifiers from a time period that corresponds with data collection periods in administrative data resources, 2) evidence that consent-based methods would not answer the research question, and 3) most importantly, that there is clear potential for public benefit from the research. This benefit must outweigh potential harms to public trust from processing the data without consent.

9.5 Impact and timeliness

In the UK and many other high-income countries, most babies receive formula milk in their first year of life, even if they are initially breastfed (McAndrew et al., 2012, Victora et al., 2016). Unlike other early interventions to support cognitive development, infant formula modifications are highly scalable, and modifications are reviewed and regulated centrally[†††], thereby rapidly reaching a worldwide population of infants. The work presented in this thesis addresses the problems of attrition by presenting a novel method to ascertain long-term effects on cognitive ability, a key outcome for parents and policymakers. The finding of no benefit and potential for harm for some formula modifications in this thesis demonstrates that monitoring long-term effects should be done routinely to detect safety issues and ensure optimal development for all formula-fed infants. The use of unconsented linkage of previously conducted trials to administrative data makes it eminently feasible to provide new, timely and important answers from data that is already available, minimising costs, waiting times, and attrition bias. This conclusion aligns well with the call for the wider and more creative use of administrative data in the UK (Dibben et al., 2009).

[†††] by the Codex Alimentarius CODEX ALIMENTARIUS COMMISSION 2007. Standard for infant formula and formulas for special medical purposes intended for infants. *Codex Stan,* 72**,** 1981. and, in Europe, by the European Directive on infant formulae and follow-on formulae COMMISSION, E. 2006. Commission Directive 2006/141/EC of 22 December 2006 on infant formulae and follow-on formulae and amending Directive 1999/21/EC. *Official Journal of the European Union,* 49**,** 1-33.

9.6 Limitations

The work presented in this thesis was subject to the strengths and limitations of the chosen methods, which have been discussed in the relevant chapters. However, some overarching limitations apply to this thesis as a whole and warrant discussion.

First, is the fact that the trials in this PhD project were chosen opportunistically based on availability. In an ideal world, I would have systematically selected a group of formula trials, for instance, trials which tested the same formula. This would have enabled me to perform individual patient data meta-analyses to determine whether factors such as dose or infant sex modify the effect of the intervention on later cognitive outcomes to formulate more nuanced implications for clinical practice. Currently, no specific register of infant formula trials exists. However, efforts are underway to create such a databank with a focus on critical appraisal of the trials and their consistency with product claims (Robert Boyle, 2020).

Second is the absence of a gold standard dataset for the linkage of trial and school data. A gold-standard dataset can be an external data source with well-validated data corresponding to the information being linked. For instance, if a subset of trials had in addition to participant names and addresses also collected participants' pupil numbers and school details – allowing near certain linkage – this would have made this subset a gold-standard dataset. An alternative to such a dataset is to manually review the actual identifiers and characteristics of a subset of linked and unlinked data to determine the true match status of each pair (Monga and Patrick, 2001, Gill, 1997, Belin and Rubin, 1995). This would have allowed me to verify and evaluate the quality of the linkage and calculate sensitivity and specificity, comparing the true match status of each record pair (in the gold-standard) with the link status of each candidate pair. For example, I could have identified obvious errors in name spellings or nicknames or links that are implausible given the data that I have from the trials. (Zingmond et al., 2004). Due to strict data protection regulations, I could not see the identifiers at the same time as the clinical data, and a gold-standard dataset was not available. Instead, this thesis made use of multiple possible

candidate matches per participant, together with indicators of linkage strength to calculate probabilistic match weights. While this method did not allow the direct evaluation of linkage error it allowed informed approximations of linkage error, while minimising the risk of personal data disclosure.

Third, academic performance analyses in this thesis were only powered to detect between-group differences greater than 0.3 SD – smaller differences are traditionally not considered clinically important in infant formula research (McLeod et al., 2016). However, infant formula trials investigating cognitive outcomes also traditionally focus on outcomes for which small changes are unlikely to have any significant long-term consequences (e.g., Bayley or IQ scores). It could be argued that much smaller average SD differences might have a substantial public health impact at a population level if the outcome is academic performance (Kraft, 2018) and that this PhD study was therefore underpowered. A small change in performance in GCSE examinations can lead to a difference in final grades – with recent research illustrating how falling just above or just below key grade boundaries can have important consequences for academic progression (Machin et al., 2020). In fact, educational researchers have proposed that effect sizes ≥ 0.20 SD should be considered "of policy interest" (Hedges and Hedberg, 2007), are "substantively important" (WWC, 2014), or have "educational significance" (Bloom et al., 2008). Lipsey and colleagues even argue that effect sizes of 0.25 SD in education research should be considered "large" (Lipsey et al., 2012). While this is a valid point, extremely large trials would have to be conducted to confirm or rule out effect sizes of such a small SD magnitude. Whether future trials with higher power are desired to confidently exclude 'important' effects might therefore depend on factors like feasibility of recruiting and following up large sample sizes, cost, and the value of information that such a trial would provide.

Finally, one might criticise that potential cognitive benefits of the formula modification could have been diluted over time, masking the true effectiveness of the modifications. One mechanism of such a dilution could be that infants in the standard formula groups eventually catch up and raise their cognitive ability to the level of the modified formula group with the help of (unobserved) external interventions such as private tutoring. Another mechanism might be that initial

benefits to cognitive ability in the modified formula groups eventually wear out. If the first mechanism were true, the analytic approach I used in this PhD study (focusing on long-term outcomes) could be considered inappropriate to evaluate potential formula benefits. This is because unobserved external intervention effects within the comparison group would make a true benefit impossible to measure. In a meta-analysis of 7,584 participants across 39 randomised controlled trials, Protzko (2015) suggests that indeed the second mechanism is most likely. The author found that after an intervention raises intelligence, the effects fade away because children in the experimental group lose their cognitive advantage and not because those in the control group catch up. For example, infants with an initial cognitive advantage might not self-select into more demanding environments or are not put on a higher trajectory of learning and, therefore, their cognitive advantage is not sustained. This finding supports my analytical approach. In addition, none of the trials I investigated found a strong signal of cognitive advantage for the modified formula group in early childhood (before dilution could have taken place). While this could still be due to low sensitivity of early measures, I argue that evidence of no benefit in multiple developmental points during childhood plus no evidence of benefit on academic performance (known to be predictive of higher education and employment) makes it unlikely that further follow-up of these formula modifications will add new knowledge – unless it is considered essential to narrow the confidence intervals to exclude harm or to confirm small benefits.

9.7 Using unconsented RCT-admin data linkage for other important research questions

The work presented in this thesis opens up several avenues for further research. The scenario considered in this thesis, of extending dormant infant formula trials with school data, can be seen as just one example of reactivating dormant trials with administrative data. This thesis focused on potential cognitive benefits of the formula modifications. However, there are also indications of metabolic harms associated with some of the formula modifications. For example there is RCT

evidence that nutrient-enriched formula is associated with adverse effects on body fat and blood pressure at age 8 years (Singhal et al., 2007). The same approach of unconsented linkage could be followed to link the trials to GP and hospital data to investigate whether adverse effects are also reflected in higher incidence of cardio-metabolic symptoms, diagnoses, and procedures. In fact, the request to link the data to health records was included in the initial application to CAG and has already been approved (appendix p 18). Compared to indicators of cognitive ability measured in school data, cardio-metabolic outcomes will take longer to emerge as most participants are currently too young to have had regular interactions with the healthcare system. It would be feasible, however, to undertake earlier analyses among female participants of childbearing age as blood pressure and hypertension metrics are routinely recorded during pregnancy.

There are several pre-requisites for unconsented linkage of dormant trials to administrative data. These are (1) that the trials are of high internal and external validity, are adequately powered, and that they have retained high-quality identifiers; (2) that the research question is likely to lead to important benefits for the public, meaning that the outcomes in admin data are definitive and clinically relevant (such as academic performance) or represent definitive measures of disease such as treatment (or hospitalisation) for hypertension, diabetes, or respiratory disease (e.g., asthma); (3) there is evidence that previously collected measures (especially in children and for early developmental measures) could not adequately capture the outcomes of interest or have poor predictive validity for important long-term outcomes such as brain function, and lung function (Zivanovic et al., 2014); and (4) that long-term consented follow-up without attrition is not feasible.

Examples of areas that could benefit from unconsented linkage to health and education data include the extension of other interventions in areas that struggle with high participant drop-out. For instance, trials that target disease prevention rather than treatment and therefore struggle with participant motivation, and consequently, the ascertainment of important outcomes that might only emerge years or even decades from the initial intervention.

Another practical extension of this PhD study could be the systematic assessment of correlations between short-term, medium-term and long-term

outcomes to collectively inform the predictive values for short-term and medium-term outcomes in trials with limited time and funding or in settings where administrative data is not yet available (Bernard, 2020, Lucas, 1998).

Finally, retrospective unconsented linkage to administrative data can also facilitate the re-analysis of important trials by providing alternative outcomes as sensitivity analyses (Henry and Fitzpatrick, 2015). To facilitate such analyses, participant identifiers themselves could be treated as a public good. They could be shared for re-use with other researchers – similarly to the widespread practice of sharing (deidentified) trial data, but in addition also considered by an approval body separately from the original trial investigators.

The previous examples highlight one of the main challenges to overcome: the practice of storing identifiers on paper and their destruction after a set time, both frequently requested by ethics committees. These practices undoubtedly limit the value and public benefit from data that was already collected. Data retention policies should be re-examined to enable the collection and preservation of linkable identifiers, especially those which enable direct linkage to national records such as pupil numbers or national health insurance numbers (Henry and Fitzpatrick, 2015).

9.8 Concluding remarks

This thesis demonstrates the processes and added value of unconsented linkage of a series of nine dormant trials held at the UCL Great Ormond Street Institute of Child Health to administrative education data to measure long-term cognitive ability. Evidence from across this thesis has shown no benefit of the investigated infant formula modifications on cognitive ability, and weak evidence of lower cognitive performance associated with the use of LCPUFA supplemented infant formula in term and preterm infants. Findings from this PhD study represent the best available evidence on long-term cognitive effects of nutrient-enriched post discharge formula, LCPUFA supplemented infant formula, high-iron follow-on formula, and infant formula supplemented with sn-2 palmitate. There is broad scope to extend the use of unconsented linkage to administrative data to areas of clinical importance within and outside the field of infant formula research.

9.9 Outputs

Publications

- Verfürden M, Gilbert R, Jerrim J, Lucas A, Fewtrell M. Long-term effectiveness of modified infant formula on cognition: linkage of five dormant randomised controlled trials to academic performance data (*submitted*)
- Verfürden M, Harron K, Jerrim J, Fewtrell M, Gilbert R. Infant formula composition and educational performance: a protocol to extend follow-up for a set of randomised controlled trials using linked administrative education records. BMJ Open 2020; 10(7): e035968.
- Verfürden ML, Dib S, Jerrim J, Fewtrell M, Gilbert RE. Effect of long-chain polyunsaturated fatty acids in infant formula on long-term cognitive function in childhood: A systematic review and meta-analysis of randomised controlled trials. *PLOS ONE* 2020; 15(11): e0241800.

Presentations

- Invited oral presentation at the Third meeting of the Routine Data Working Group of the National Institute for Health Research (Bristol in January 2020)
- Invited oral presentation at Methods in Action class of the Network of Applied Statisticians in Health (London in January 2020)
- Oral presentation at Society for the Study of Human Biology conference (Oxford in August 2019)
- Oral presentation at the International Population Data Linkage Conference (Canada (Banff) in September 2018)
- Oral presentation at the Administrative Data Research Network Conference (Northern Ireland (Belfast) in June 2018)
- Invited oral presentation at the offices of the Education Endowment Foundation (London in February 2018)
- Oral presentation ESRC Winter Conference (London in December 2017)

Resources

- GitHub page containing annotated Stata code to follow data preparation and analysis: https://github.com/MaxVerfuerden/PhD

References

ABEL, K., HEUVELMAN, H., WICKS, S., RAI, D., EMSLEY, R., GARDNER, R. & DALMAN, C. 2017. Gestational age at birth and academic performance: population-based cohort study. *Int J Epidemiol,* 46**,** 324-335.

AGOSTONI, C., TROJAN, S., BELLU, R., RIVA, E., BRUZZESE, M. G. & GIOVANNINI, M. 1997. Developmental quotient at 24 months and fatty acid composition of diet in early infancy: a follow up study. *Archives of Disease in Childhood,* 76**,** 421-424.

AGOSTONI, C., TROJAN, S., BELLU, R., RIVA, E. & GIOVANNINI, M. 1995. Neurodevelopmental quotient of healthy term infants at 4 months and feeding practice: the role of long-chain polyunsaturated fatty acids. *Pediatr Res,* 38**,** 262-6.

AKYOL, Ş. P., KRISHNA, K. & WANG, J. 2018. Taking PISA seriously: How accurate are low stakes exams? : National Bureau of Economic Research.

ANDERSON, J. W., JOHNSTONE, B. M. & REMLEY, D. T. 1999. Breast-feeding and cognitive development: a meta-analysis. *Am J Clin Nutr,* 70**,** 525-35.

ANDRES, A., CLEVES, M. A., BELLANDO, J. B., PIVIK, R. T., CASEY, P. H. & BADGER, T. M. 2012. Developmental status of 1-year-old infants fed breast milk, cow's milk formula, or soy formula. *Pediatrics,* 129**,** 1134-40.

ATTALI, Y. 2016. Effort in Low-Stakes Assessments: What Does It Take to Perform as Well as in a High-Stakes Setting? *Educational and psychological measurement,* 76**,** 1045-1058.

AUESTAD, N., HALTER, R., HALL, R. T., BLATTER, M., BOGLE, M. L., BURKS, W., ERICKSON, J. R., FITZGERALD, K. M., DOBSON, V., INNIS, S. M., SINGER, L. T., MONTALTO, M. B., JACOBS, J. R., QIU, W. & BORNSTEIN, M. H. 2001. Growth and development in term infants fed long-chain polyunsaturated fatty acids: a double-masked, randomized, parallel, prospective, multivariate study. *Pediatrics,* 108**,** 372-81.

AUESTAD, N., SCOTT, D. T., JANOWSKY, J. S., JACOBSEN, C., CARROLL, R. E., MONTALTO, M. B., HALTER, R., QIU, W., JACOBS, J. R., CONNOR, W. E., CONNOR, S. L., TAYLOR, J. A., NEURINGER, M., FITZGERALD, K. M. & HALL, R. T. 2003. Visual, cognitive, and language assessments at 39 months: a follow-up study of children fed formulas containing long-chain polyunsaturated fatty acids to 1 year of age. *Pediatrics,* 112**,** e177-83.

AUMEISTERE, L., CIPROVIČA, I., ZAVADSKA, D. & VOLKOVS, V. 2018. Fish intake reflects on DHA level in breast milk among lactating women in Latvia. *International breastfeeding journal,* 13**,** 1-8.

BARKER, D. J. 1990. The fetal and infant origins of adult disease. *BMJ: British Medical Journal,* 301**,** 1111.

BARKER, D. J. 1992. The fetal origins of adult hypertension. *Journal of Hypertension,* 10**,** S45.

BARKER, D. J. & OSMOND, C. 1986. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet,* 1**,** 1077-81.

BELIN, T. R. & RUBIN, D. B. 1995. A method for calibrating false-match rates in record linkage. *Journal of the American Statistical Association,* 90**,** 694-707.

BEN, X.-M., ZHOU, X.-Y., ZHAO, W.-H., YU, W.-L., PAN, W., ZHANG, W.-L., WU, S.-M., VAN BEUSEKOM, C. M. & SCHAAFSMA, A. 2004a. Supplementation of milk formula with galacto-oligosaccharides improves intestinal micro-flora and fermentation in term infants. *Chinese medical journal,* 117**,** 927-931.

BEN, X. M., ZHOU, X. Y., ZHAO, W. H., YU, W. L., PAN, W., ZHANG, W. L., WU, S. M., VAN BEUSEKOM, C. M. & SCHAAFSMA, A. 2004b. Growth and development of term infants fed with milk with long-chain polyunsaturated fatty acid supplementation. *Chin Med J (Engl),* 117**,** 1268-70.

BERNARD, D. R. 2020. Estimating long-term treatment effects without long-term outcome data. *Global Priorities Institute Working Papers,* .

BEST, J. R. & MILLER, P. H. 2010. A developmental perspective on executive function. *Child development,* 81**,** 1641-1660.

BINET, A. & SIMON, T. 1916. New methods for the diagnosis of the intellectual level of subnormals.(L'Année Psych., 1905, pp. 191-244).

BIRCH, E. E., GARFIELD, S., CASTANEDA, Y., HUGHBANKS-WHEATON, D., UAUY, R. & HOFFMAN, D. 2007. Visual acuity and cognitive outcomes at 4 years of age in a double-blind, randomized trial of long-chain polyunsaturated fatty acid-supplemented infant formula. *Early Hum Dev,* 83**,** 279-84.

BIRCH, E. E., GARFIELD, S., HOFFMAN, D. R., UAUY, R. & BIRCH, D. G. 2000. A randomized controlled trial of early dietary supply of long-chain polyunsaturated fatty acids and mental development in term infants. *Dev Med Child Neurol,* 42**,** 174-81.

BLOOM, H. S., HILL, C. J., BLACK, A. R. & LIPSEY, M. W. 2008. Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness,* 1**,** 295.

BOTTING, N., POWLS, A., COOKE, R. W. & MARLOW, N. 1998. Cognitive and educational outcome of very-low-birthweight children in early adolescence. *Dev Med Child Neurol,* 40**,** 652-60.

BOUWSTRA, H., DIJCK-BROUWER, D. A., BOEHM, G., BOERSMA, E. R., MUSKIET, F. A. & HADDERS-ALGRA, M. 2005. Long-chain polyunsaturated fatty acids and neurological developmental outcome at 18 months in healthy term infants. *Acta Paediatr,* 94**,** 26-32.

BRONSKY, J., CAMPOY, C., EMBLETON, N., FEWTRELL, M., MIS, N. F., GERASIMIDIS, K., HOJSAK, I., HULST, J., INDRIO, F., LAPILLONNE, A., MOLGAARD, C., MOLTU, S. J., VERDUCI, E., VORA, R. & DOMELLÖF, M. 2019. Palm Oil and Beta-palmitate in Infant Formula: A Position Paper by the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) Committee on Nutrition. *J Pediatr Gastroenterol Nutr,* 68**,** 742-760.

BRUNSER, O., ESPINOZA, J., ARAYA, M., CRUCHET, S. & GIL, A. 1994. Effect of dietary nucleotide supplementation on diarrhoeal disease in infants. *Acta Paediatrica,* 83**,** 188-191.

CARPENTER, J. & KENWARD, M. 2012. *Multiple imputation and its application*, John Wiley & Sons.

CARVER, J. D., PIMENTEL, B., COX, W. I. & BARNESS, L. A. 1991. Dietary nucleotide effects upon immune function in infants. *Pediatrics,* 88**,** 359-363.

CASTANYS-MUÑOZ, E., KENNEDY, K., CASTAÑEDA-GUTIÉRREZ, E., FORSYTH, S., GODFREY, K. M., KOLETZKO, B., OZANNE, S. E., RUEDA, R., SCHOEMAKER, M., VAN DER BEEK, E. M., VAN BUUREN, S. & ONG, K. K. 2017. Systematic review indicates postnatal growth in term infants born small-for-gestational-age being associated with later neurocognitive and metabolic outcomes. *Acta Paediatrica,* 106**,** 1230-1238.

CODEX ALIMENTARIUS COMMISSION 2007. Standard for infant formula and formulas for special medical purposes intended for infants. *Codex Stan,* 72**,** 1981.

COLOMBO, J. 2018. Assessing Neurocognitive Development in Studies of Nutrition. *Recent Research in Nutrition and Growth. Nestlé Nutr Inst Workshop.* Switzerland: Nestlé Nutrition Institute,.

COLOMBO, J. & CARLSON, S. E. 2012. Is the measure the message: the BSID and nutritional interventions. *Pediatrics,* 129**,** 1166-1167.

COLOMBO, J., GUSTAFSON, K. M. & CARLSON, S. E. 2019. Critical and Sensitive Periods in Development and Nutrition. *Annals of Nutrition and Metabolism,* 75(suppl 1)**,** 34-42.

COLOMBO, J., JILL SHADDY, D., KERLING, E. H., GUSTAFSON, K. M. & CARLSON, S. E. 2017. Docosahexaenoic acid (DHA) and arachidonic

acid (ARA) balance in developmental outcomes. *Prostaglandins Leukot Essent Fatty Acids,* 121**,** 52-56.

COMMISSION, E. 2006. Commission Directive 2006/141/EC of 22 December 2006 on infant formulae and follow-on formulae and amending Directive 1999/21/EC. *Official Journal of the European Union,* 49**,** 1-33.

COOKE, R. J., EMBLETON, N. D., GRIFFIN, I. J., WELLS, J. C. & MCCORMICK, K. P. 2001. Feeding preterm infants after hospital discharge: growth and development at 18 months of age. *Pediatr Res,* 49**,** 719-22.

COOKE, R. W. & FOULDER-HUGHES, L. 2003. Growth impairment in the very preterm and cognitive and motor performance at 7 years. *Arch Dis Child,* 88**,** 482-7.

COSGROVE, M., DAVIES, D. P. & JENKINS, H. R. 1996. Nucleotide supplementation and the growth of term small for gestational age infants. *Archives of Disease in Childhood. Fetal and Neonatal Edition,* 74**,** F122-F125.

CROSS, L., CARSON, L. E., JEWELL, A., HESLIN, M., OSBORN, D., DOWNS, J. & STEWART, R. 2020. Guidance for researchers wanting to link NHS data using non-consent approaches: a thematic analysis of feedback from the Health Research Authority Confidentiality Advisory Group: A thematic analysis of feedback from the Health Research Authority Confidentiality Advisory Group. *International Journal of Population Data Science,* 5.

DE JONG, C., KIKKERT, H. K., FIDLER, V. & HADDERS-ALGRA, M. 2012. Effects of long-chain polyunsaturated fatty acid supplementation of infant formula on cognition and behaviour at 9 years of age. *Dev Med Child Neurol,* 54**,** 1102-8.

DE JONG, C., KIKKERT, H. K., SEGGERS, J., BOEHM, G., DECSI, T. & HADDERS-ALGRA, M. 2015. Neonatal fatty acid status and neurodevelopmental outcome at 9 years. *Early Human Development,* 91**,** 587-591.

DEARY, I. J. 2020. *Intelligence: A very short introduction*, Oxford University Press.

DEARY, I. J. & BATTY, G. D. 2007. Cognitive epidemiology. *Journal of epidemiology and community health,* 61**,** 378-384.

DEEKS, J. J., HIGGINS, J., ALTMAN, D. G. & GREEN, S. 2011. Cochrane handbook for systematic reviews of interventions version 5.1. 0 (updated March 2011). *The Cochrane Collaboration***,** 2.

DEPARTMENT FOR EDUCATION. 2017a. *Freedom of Information Request by Jen Persson* [Online]. Available: https://www.whatdotheyknow.com/request/pupil_data_national_pupil_databa_3?unfold=1 [Accessed 28 Mar 2020].

DEPARTMENT FOR EDUCATION 2017b. Statistical Notice: Statement of administrative sources.

DEPARTMENT FOR EDUCATION. 2020a. *NPD Data Tables* [Online]. Available: https://find-npd-data.education.gov.uk/data_tables/last [Accessed 26 Mar 2020].

DEPARTMENT FOR EDUCATION. 2020b. *Technical specification and validation rules* [Online]. Available: https://www.gov.uk/guidance/complete-the-school-census [Accessed 28 Mar 2020].

DEPARTMENT FOR EDUCATION. 2021. *Find and explore data in the National Pupil Database* [Online]. Available: https://find-npd-data.education.gov.uk/categories [Accessed 27 Feb 2021].

DIBBEN, C., GOWANS, H., ELLIOT, M., ANTTILA, C., BOYLE, P., KAYE, J., MCLENNAN, D., NOBLE, M., SMITH, G. & WILKINSON, K. 2009. Encouraging the wider and more creative use of administrative data in the UK-the'Administrative Data Liaison Service'.

DICKENS, J. 2015. Criminal check needed for access to National Pupil Database. Available from: https://schoolsweek.co.uk/criminal-check-needed-for-access-to-national-pupil-database/ [Accessed 28 Mar 2020].

DOIDGE, J. C. & HARRON, K. L. 2019. Reflections on modern methods: linkage error bias. *Int J Epidemiol,* 48**,** 2050-2060.

DOMELLÖF, M., BRAEGGER, C., CAMPOY, C., COLOMB, V., DECSI, T., FEWTRELL, M., HOJSAK, I., MIHATSCH, W., MOLGAARD, C., SHAMIR, R., TURCK, D. & VAN GOUDOEVER, J. 2014. Iron requirements of infants and toddlers. *J Pediatr Gastroenterol Nutr,* 58**,** 119-29.

DOMELLÖF, M., LÖNNERDAL, B., ABRAMS, S. A. & HERNELL, O. 2002. Iron absorption in breast-fed infants: effects of age, iron status, iron supplements, and complementary foods. *The American journal of clinical nutrition,* 76**,** 198-204.

DÖRNER, G. 1976. *Hormones and brain differentiation*, North Holland.

DROVER, J. R., FELIUS, J., HOFFMAN, D. R., CASTANEDA, Y. S., GARFIELD, S., WHEATON, D. H. & BIRCH, E. E. 2012. A randomized trial of DHA intake during infancy: school readiness and receptive vocabulary at 2-3.5 years of age. *Early Hum Dev,* 88**,** 885-91.

DROVER, J. R., HOFFMAN, D. R., CASTANEDA, Y. S., MORALE, S. E., GARFIELD, S., WHEATON, D. H. & BIRCH, E. E. 2011. Cognitive function in 18-month-old term infants of the DIAMOND study: a randomized, controlled clinical trial with multiple dietary levels of docosahexaenoic acid. *Early Hum Dev,* 87**,** 223-30.

DUCKWORTH, A. L., QUINN, P. D., LYNAM, D. R., LOEBER, R. & STOUTHAMER-LOEBER, M. 2011. Role of test motivation in

intelligence testing. *Proceedings of the National Academy of Sciences,* 108**,** 7716-7720.

EFSA NDA PANEL 2014. Scientific opinion on the essential composition of infant and follow-on formulae. *EFSA J,* 12**,** 3760.

EFSA PANEL ON DIETETIC PRODUCTS, N. & ALLERGIES 2014. Scientific Opinion on the essential composition of infant and follow-on formulae. *EFSA Journal,* 12**,** 3760.

EMBLETON, N. 2013. Optimal nutrition for preterm infants: Putting the ESPGHAN guidelines into practice. *Journal of Neonatal nursing,* 19**,** 130-133.

EMBLETON, N. D., WOOD, C. L., PEARCE, M. S., BRUNSKILL, G. & GRAHAME, V. 2021. Early diet in preterm infants and later cognition: 10-year follow-up of a randomized controlled trial. *Pediatric Research*.

EU COMMISSION 2016. Commission Delegated Regulation (EU) 2016/127 of 25 September 2015 supplementing Regulation (EU) No 609/2013 of the European Parliament and of the Council as regards the specific compositional and information requirements for infant formula and follow-on formula and as regards requirements on information relating to infant and young child feeding. *OJEC,* 59**,** 1-29.

EUROPEAN MEDICINES AGENCY 2015. Guideline on adjustment for baseline covariates in clinical trials. Author London, England.

EUROVOC 2020. EuroVoc – Multilingual thesaurus of the European Union.

FEWTRELL, M. S., ABBOTT, R. A., KENNEDY, K., SINGHAL, A., MORLEY, R., CAINE, E., JAMIESON, C., COCKBURN, F. & LUCAS, A. 2004. Randomized, double-blind trial of long-chain polyunsaturated fatty acid supplementation with fish oil and borage oil in preterm infants. *J Pediatr,* 144**,** 471-9.

FEWTRELL, M. S., DOMELLOF, M., HOJSAK, I., HULST, J. M., KENNEDY, K., KOLETZKO, B., MIHATSH, W. & STIJNEN, T. 2016. Attrition in Long-Term Nutrition Research Studies: A Commentary by the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Early Nutrition Research Working Group. *J Pediatr Gastroenterol Nutr,* 62**,** 180-2.

FEWTRELL, M. S., KENNEDY, K., SINGHAL, A., MARTIN, R. M., NESS, A., HADDERS-ALGRA, M., KOLETZKO, B. & LUCAS, A. 2008. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Arch Dis Child,* 93**,** 458-61.

FEWTRELL, M. S., MORLEY, R., ABBOTT, R. A., SINGHAL, A., ISAACS, E. B., STEPHENSON, T., MACFADYEN, U. & LUCAS, A. 2002. Double-blind, randomized trial of long-chain polyunsaturated fatty acid supplementation in formula fed to preterm infants. *Pediatrics,* 110**,** 73-82.

FEWTRELL, M. S., MORLEY, R., ABBOTT, R. A., SINGHAL, A., STEPHENSON, T., MACFADYEN, U. M., CLEMENTS, H. & LUCAS, A. 2001. Catch-up growth in small-for-gestational-age term infants: a randomized trial. *The American journal of clinical nutrition,* 74**,** 516-523.

FRIEL, J. K., AZIZ, K., ANDREWS, W. L., HARDING, S. V., COURAGE, M. L. & ADAMS, R. J. 2003. A double-masked, randomized control trial of iron supplementation in early infancy in healthy term breast-fed infants. *J Pediatr,* 143**,** 582-6.

GAHAGAN, S., DELKER, E., BLANCO, E., BURROWS, R. & LOZOFF, B. 2019. Randomized Controlled Trial of Iron-Fortified versus Low-Iron Infant Formula: Developmental Outcomes at 16 Years. *J Pediatr,* 212**,** 124-130.e1.

GEORGIEFF, M. K. 2008. The role of iron in neurodevelopment: fetal iron deficiency and the developing hippocampus. *Biochem Soc Trans,* 36**,** 1267-71.

GIBSON, R. A. & MAKRIDES, M. 2001. Long-chain polyunsaturated fatty acids in breast milk: are they essential? *Adv Exp Med Biol,* 501**,** 375-83.

GIL, A. & SANCHEZ-MEDINA, F. 1982. Acid-soluble nucleotides of human milk at different stages of lactation. *Journal of Dairy Research,* 49**,** 301-307.

GILBERT, R., HARRON, K. & DEARDEN, L. 2015. Benefits of, and barriers to, reactivating dormant trials. *BMJ : British Medical Journal,* 351**,** h5298.

GILL, L. E. OX-LINK: the Oxford medical record linkage system. Proc. Int'l Record Linkage Workshop and Exposition, 1997. Citeseer, 15-33.

GLASS, H. C., COSTARINO, A. T., STAYER, S. A., BRETT, C. M., CLADIS, F. & DAVIS, P. J. 2015. Outcomes for extremely premature infants. *Anesthesia and analgesia,* 120**,** 1337-1351.

GOLDSTEIN, H. & HEALY, M. J. 1995. The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 158**,** 175-177.

GOTTFREDSON, L. S. 1997. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Citeseer.

GRANTHAM-MCGREGOR, S., CHEUNG, Y. B., CUETO, S., GLEWWE, P., RICHTER, L. & STRUPP, B. 2007. Developmental potential in the first 5 years for children in developing countries. *The Lancet,* 369**,** 60-70.

HACK, M., BRESLAU, N., WEISSMAN, B., ARAM, D., KLEIN, N. & BORAWSKI, E. 1991. Effect of Very Low Birth Weight and Subnormal Head Size on Cognitive Abilities at School Age. *New England Journal of Medicine,* 325**,** 231-237.

HAGGER-JOHNSON, G., HARRON, K., GONZALEZ-IZQUIERDO, A., CORTINA-BORJA, M., DATTANI, N., MULLER-PEBODY, B., PARSLOW, R., GILBERT, R. & GOLDSTEIN, H. 2015. Identifying Possible False Matches in Anonymized Hospital Administrative Data without Patient Identifiers. *Health Serv Res,* 50**,** 1162-78.

HALTERMAN, J. S., KACZOROWSKI, J. M., ALIGNE, C. A., AUINGER, P. & SZILAGYI, P. G. 2001. Iron Deficiency and Cognitive Achievement Among School-Aged Children and Adolescents in the United States. *Pediatrics,* 107**,** 1381-1386.

HANLEY, B., DIJANE, J., FEWTRELL, M., GRYNBERG, A., HUMMEL, S., JUNIEN, C., KOLETZKO, B., LEWIS, S., RENZ, H., SYMONDS, M., GROS, M., HARTHOORN, L., MACE, K., SAMUELS, F. & VAN DER BEEK, E. M. 2010. Metabolic imprinting, programming and epigenetics – a review of present priorities and future opportunities. *British Journal of Nutrition,* 104**,** S1-S25.

HARRON, K. 2014. *Evaluating data linkage techniques for the analysis of bloodstream infection in paediatric intensive care.* UCL (University College London).

HARRON, K., DIBBEN, C., BOYD, J., HJERN, A., AZIMAEE, M., BARRETO, M. L. & GOLDSTEIN, H. 2017a. Challenges in administrative data linkage for research. *Big Data Soc,* 4**,** 2053951717745678.

HARRON, K., GILBERT, R., CROMWELL, D. & VAN DER MEULEN, J. 2016. Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS One,* 11**,** e0164667.

HARRON, K., GOLDSTEIN, H. & DIBBEN, C. 2015. *Methodological developments in data linkage*, John Wiley & Sons.

HARRON, K., WADE, A., GILBERT, R., MULLER-PEBODY, B. & GOLDSTEIN, H. 2014. Evaluating bias due to data linkage error in electronic healthcare records. *BMC Med Res Methodol,* 14**,** 36.

HARRON, K., WADE, A., MULLER-PEBODY, B., GOLDSTEIN, H. & GILBERT, R. 2012. Opening the black box of record linkage. *J Epidemiol Community Health,* 66**,** 1198.

HARRON, K. L., DOIDGE, J. C., KNIGHT, H. E., GILBERT, R. E., GOLDSTEIN, H., CROMWELL, D. A. & VAN DER MEULEN, J. H. 2017b. A guide to evaluating linkage quality for the analysis of linked data. *Int J Epidemiol,* 46**,** 1699-1710.

HAWKES, J. S., GIBSON, R. A., ROBERTON, D. & MAKRIDES, M. 2005. Effect of dietary nucleotide supplementation on growth and immune function in term infants: a randomized controlled trial. *European Journal Of Clinical Nutrition,* 60**,** 254.

HAYWARD, H., HUNT, E. & LORD, A. 2014. *The economic value of key intermediate qualifications: estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships*, Department for Education London.

HEDGES, L. V. & HEDBERG, E. C. 2007. Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis,* 29**,** 60-87.

HENRY, D. & FITZPATRICK, T. 2015. Liberating the data from clinical trials. *BMJ : British Medical Journal,* 351**,** h4601.

HUGHES, H. K., LANDA, M. M. & SHARFSTEIN, J. M. 2017. Marketing Claims for Infant Formula: The Need for Evidence. *JAMA Pediatr,* 171**,** 105-106.

IDO, P. A. Z., GALE, R., LAOR, A., DANON, Y. L., STEVENSON, D. K. & SEIDMAN, D. S. 1995. The cognitive outcome of full-term small for gestational age infants at late adolescence. *Obstetrics & Gynecology,* 85**,** 452-456.

INNIS, S. M. 2011. Dietary triacylglycerol structure and its role in infant nutrition. *Adv Nutr,* 2**,** 275-83.

INNIS, S. M. 2016. Palmitic Acid in Early Human Development. *Critical Reviews in Food Science and Nutrition,* 56**,** 1952-1959.

ISAACS, E. B., MORLEY, R. & LUCAS, A. 2009. Early Diet and General Cognitive Outcome at Adolescence in Children Born at or Below 30 Weeks Gestation. *The Journal of Pediatrics,* 155**,** 229-234.

ISAACS, E. B., ROSS, S., KENNEDY, K., WEAVER, L. T., LUCAS, A. & FEWTRELL, M. S. 2011. 10-year cognition in preterms after random assignment to fatty acid supplementation in infancy. *Pediatrics,* 128**,** e890-8.

JAEGGI, T., KORTMAN, G. A. M., MORETTI, D., CHASSARD, C., HOLDING, P., DOSTAL, A., BOEKHORST, J., TIMMERMAN, H. M., SWINKELS, D. W., TJALSMA, H., NJENGA, J., MWANGI, A., KVALSVIG, J., LACROIX, C. & ZIMMERMANN, M. B. 2015. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. *Gut,* 64**,** 731-742.

JASANI, B., SIMMER, K., PATOLE, S. K. & RAO, S. C. 2017. Long chain polyunsaturated fatty acid supplementation in infants born at term. *Cochrane Database Syst Rev,* 3**,** CD000376.

JAY, M. A., MC GRATH-LONE, L. & GILBERT, R. 2019. Data Resource: the National Pupil Database (NPD). *International Journal of Population Data Science,* 4.

JEON, G. W., JUNG, Y. J., KOH, S. Y., LEE, Y. K., KIM, K. A., SHIN, S. M., KIM, S. S., SHIM, J. W., CHANG, Y. S. & PARK, W. S. 2011. Preterm

infants fed nutrient-enriched formula until 6 months show improved growth and development. *Pediatr Int,* 53**,** 683-688.

JONES, G. & SCHNEIDER, W. J. 2006. Intelligence, human capital, and economic growth: A Bayesian averaging of classical estimates (BACE) approach. *Journal of economic growth,* 11**,** 71-93.

JONSDOTTIR, O. H., THORSDOTTIR, I., HIBBERD, P. L., FEWTRELL, M. S., WELLS, J. C., PALSSON, G. I., LUCAS, A., GUNNLAUGSSON, G. & KLEINMAN, R. E. 2012. Timing of the introduction of complementary foods in infancy: a randomized controlled trial. *Pediatrics,* 130**,** 1038-1045.

KENNEDY, K., FEWTRELL, M. S., MORLEY, R., ABBOTT, R., QUINLAN, P. T., WELLS, J. C., BINDELS, J. G. & LUCAS, A. 1999. Double-blind, randomized trial of a synthetic triacylglycerol in formula-fed term infants: effects on stool biochemistry, stool characteristics, and bone mineralization. *Am J Clin Nutr,* 70**,** 920-7.

KOLETZKO, B., BAKER, S., CLEGHORN, G., NETO, U. F., GOPALAN, S., HERNELL, O., HOCK, Q. S., JIRAPINYO, P., LONNERDAL, B. & PENCHARZ, P. 2005. Global standard for the composition of infant formula: recommendations of an ESPGHAN coordinated international expert group. *Journal of pediatric gastroenterology and nutrition,* 41**,** 584-599.

KRAFT, M. A. 2018. Interpreting Effect Sizes of Education Interventions. *Brown University Working Paper.*

KRAMER, M. S., ABOUD, F., MIRONOVA, E., VANILOVICH, I., PLATT, R. W., MATUSH, L., IGUMNOV, S., FOMBONNE, E., BOGDANOVICH, N., DUCRUET, T., COLLET, J.-P., CHALMERS, B., HODNETT, E., DAVIDOVSKY, S., SKUGAREVSKY, O., TROFIMOVICH, O., KOZLOVA, L., SHAPIRO, S. & GROUP, P. O. B. I. T. S. 2008. Breastfeeding and Child Cognitive Development: New Evidence From a Large Randomized Trial. *Archives of General Psychiatry,* 65**,** 578-584.

KUMAR, R. K., SINGHAL, A., VAIDYA, U., BANERJEE, S., ANWAR, F. & RAO, S. 2017. Optimizing Nutrition in Preterm Low Birth Weight Infants—Consensus Summary. *Frontiers in Nutrition,* 4.

LANIGAN, J. & SINGHAL, A. 2009. Early nutrition and long-term health: a practical approach. *Proc Nutr Soc,* 68**,** 422-9.

LAPILLONNE, A., GROH-WARGO, S., LOZANO GONZALEZ, C. H. & UAUY, R. 2013. Lipid Needs of Preterm Infants: Updated Recommendations. *The Journal of Pediatrics,* 162**,** S37-S47.

LEPPÄNEN, M., LAPINLEIMU, H., LIND, A., MATOMÄKI, J., LEHTONEN, L., HAATAJA, L. & RAUTAVA, P. 2014. Antenatal and Postnatal

Growth and 5-Year Cognitive Outcome in Very Preterm Infants. *Pediatrics,* 133**,** 63-70.

LEWIS, D. S., BERTRAND, H. A., MCMAHAN, C. A., MCGILL, H. C., JR., CAREY, K. D. & MASORO, E. J. 1986. Preweaning food intake influences the adiposity of young adult baboons. *J Clin Invest,* 78**,** 899-905.

LEWIS, J. A. & MACHIN, D. 1993. Intention to treat--who should use ITT? *Br J Cancer,* 68**,** 647-50.

LIAO, K.-Y., WU, T.-C., HUANG, C.-F., LIN, C.-C., HUANG, I.-F. & WU, L. 2011. Profile of nucleotides and nucleosides in Taiwanese human milk. *Pediatrics & Neonatology,* 52**,** 93-97.

LIN, L., AMISSAH, E., GAMBLE, G. D., CROWTHER, C. A. & HARDING, J. E. 2019. Impact of macronutrient supplements for children born preterm or small for gestational age on developmental and metabolic outcomes: A systematic review and meta-analysis. *PLoS Med,* 16**,** e1002952.

LIND, T., LONNERDAL, B., STENLUND, H., GAMAYANTI, I. L., ISMAIL, D., SESWANDHANA, R. & PERSSON, L. A. 2004. A community-based randomized controlled trial of iron and zinc supplementation in Indonesian infants: effects on growth and development. *Am J Clin Nutr,* 80**,** 729-36.

LIPSEY, M. W., PUZIO, K., YUN, C., HEBERT, M. A., STEINKA-FRY, K., COLE, M. W., ROBERTS, M., ANTHONY, K. S. & BUSICK, M. D. 2012. Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research.*

LIPSITCH, M., TCHETGEN, E. T. & COHEN, T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.),* 21**,** 383.

LLEWELLYN-BENNETT, R., EDWARDS, D., ROBERTS, N., HAINSWORTH, A. H., BULBULIA, R. & BOWMAN, L. 2018. Post-trial follow-up methodology in large randomised controlled trials: a systematic review. *Trials,* 19**,** 298.

LOZOFF, B., CASTILLO, M., CLARK, K. M. & SMITH, J. B. 2012. Iron-fortified vs low-iron infant formula: developmental outcome at 10 years. *Arch Pediatr Adolesc Med,* 166**,** 208-15.

LOZOFF, B. & GEORGIEFF, M. K. 2006. Iron Deficiency and Brain Development. *Seminars in Pediatric Neurology,* 13**,** 158-165.

LUCAS, A. 1991. Programming by early nutrition in man. *Ciba Found Symp,* 156**,** 38-50; discussion 50-5.

LUCAS, A. 1998. Programming by Early Nutrition: An Experimental Approach. *The Journal of Nutrition,* 128**,** 401S-406S.

LUCAS, A. 2005. Long-term programming effects of early nutrition -- implications for the preterm infant. *J Perinatol,* 25 Suppl 2**,** S2-6.

LUCAS, A., FEWTRELL, M. S., MORLEY, R., SINGHAL, A., ABBOTT, R. A., ISAACS, E., STEPHENSON, T., MACFADYEN, U. M. & CLEMENTS, H. 2001. Randomized trial of nutrient-enriched formula versus standard formula for postdischarge preterm infants. *Pediatrics,* 108**,** 703-11.

LUCAS, A., MAKRIDES, M. & ZIEGLER, E. Importance of growth for health and development. Nestle Nutr. Inst. Workshop Series Pediatr. Program, 2010. 251.

LUCAS, A., MORLEY, R. & COLE, T. J. 1998. Randomised trial of early diet in preterm babies and later intelligence quotient. *Bmj,* 317**,** 1481-1487.

LUCAS, A., MORLEY, R., COLE, T. J., GORE, S. M., DAVIS, J. A., BAMFORD, M. F. & DOSSETOR, J. F. 1989. Early diet in preterm babies and developmental status in infancy. *Arch Dis Child,* 64**,** 1570-8.

LUCAS, A., MORLEY, R., COLE, T. J., LISTER, G. & LEESON-PAYNE, C. 1992. Breast milk and subsequent intelligence quotient in children born preterm. *Lancet,* 339**,** 261-4.

LUCAS, A. & SAMPSON, H. 2006. Infant nutrition and primary prevention: current and future perspectives. *Primary Prevention by Nutrition Intervention in Infancy and Childhood.* Karger Publishers.

LUCAS, A., SINGHAL, A., FEWTRELL, M., ISAACS, E., FAREWELL, V., HOF, M. V. T., MORLEY, R., BROWN, C., STEPHENSON, T., MCHUGH, G. & ALIAS-JONES, A. unpublished. Long chain polyunsaturated fatty acids in infant formula and intelligence quotient at 4-5 years: a randomised trial.

LUCAS, A., STAFFORD, M., MORLEY, R., ABBOTT, R., STEPHENSON, T., MACFADYEN, U., ELIAS-JONES, A. & CLEMENTS, H. 1999. Efficacy and safety of long-chain polyunsaturated fatty acid supplementation of infant-formula milk: a randomised trial. *Lancet,* 354**,** 1948-54.

LUGG-WIDGER, F., ANGEL, L., CANNINGS-JOHN, R., HOOD, K., HUGHES, K., MOODY, G. & ROBLING, M. 2018. Challenges in accessing routinely collected data from multiple providers in the UK for primary studies: managing the morass. *International Journal of Population Data Science,* 3.

MACHIN, S., MCNALLY, S. & RUIZ-VALENZUELA, J. 2020. Entry through the narrow door: The costs of just failing high stakes exams. *Journal of Public Economics,* 190**,** 104224.

MAKRIDES, M., NEUMANN, M., SIMMER, K., PATER, J. & GIBSON, R. 1995. Are long-chain polyunsaturated fatty acids essential nutrients in infancy? *Lancet,* 345**,** 1463-8.

MAKRIDES, M., NEUMANN, M. A., SIMMER, K. & GIBSON, R. A. 1999. Dietary long-chain polyunsaturated fatty acids do not influence growth of term infants: A randomized clinical trial. *Pediatrics,* 104**,** 468-75.

MAKRIDES, M., SMITHERS, L. G. & GIBSON, R. A. 2010. Role of long-chain polyunsaturated fatty acids in neurodevelopment and growth. *Nestle Nutr Workshop Ser Pediatr Program,* 65**,** 123-33; discussion 133-6.

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.

MCLEOD, L. D., CAPPELLERI, J. C. & HAYS, R. D. 2016. Best (but oft-forgotten) practices: expressing and interpreting associations and effect sizes in clinical outcome assessments. *The American journal of clinical nutrition,* 103**,** 685-693.

MILES, E. A. & CALDER, P. C. 2017. The influence of the position of palmitate in infant formula triacylglycerols on health outcomes. *Nutr Res,* 44**,** 1-8.

MOFFATT, M. E., LONGSTAFFE, S., BESANT, J. & DURESKI, C. 1994. Prevention of iron deficiency and psychomotor decline in high-risk infants through use of iron-fortified infant formula: a randomized clinical trial. *J Pediatr,* 125**,** 527-34.

MONGA, H. & PATRICK, T. B. 2001. Error estimation in linking heterogeneous data sources. *Health Informatics Journal,* 7**,** 135-137.

MOON, K., RAO, S., PATOLE, S., SIMMER, K. & SCHULZKE, S. 2017. Longchain polyunsaturted fatty acid supplementation in preterm infants: Updated cochrane review. *Journal of Paediatrics and Child Health,* 53 (Supplement 2)**,** 69-70.

MORLEY, R., ABBOTT, R., FAIRWEATHER-TAIT, S., MACFADYEN, U., STEPHENSON, T. & LUCAS, A. 1999. Iron fortified follow on formula from 9 to 18 months improves iron status but not development or growth: a randomised trial. *Arch Dis Child,* 81**,** 247-52.

MORLEY, R., FEWTRELL, M. S., ABBOTT, R. A., STEPHENSON, T., MACFADYEN, U. & LUCAS, A. 2004. Neurodevelopment in children born small for gestational age: a randomized trial of nutrient-enriched versus standard formula and comparison with a reference breastfed group. *Pediatrics,* 113**,** 515-21.

MORRIS, T. P., WHITE, I. R. & ROYSTON, P. 2014. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology,* 14**,** 75.

MORRIS, Z. S., WOODING, S. & GRANT, J. 2011. The answer is 17 years, what is the question: understanding time lags in translational research. *Journal of the Royal Society of Medicine,* 104**,** 510-520.

MORROW-TLUCAK, M., HAUDE, R. H. & ERNHART, C. B. 1988. Breastfeeding and cognitive development in the first 2 years of life. *Social Science & Medicine,* 26**,** 635-639.

NEURINGER, M. & CONNOR, W. E. 1986. n-3 fatty acids in the brain and retina: evidence for their essentiality. *Nutrition Reviews,* 44**,** 285-294.

NEWELL, D. J. 1992. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol,* 21**,** 837-41.

OFFICE FOR NATIONAL STATISTICS. 2016. *Review of the Dissemination of Health Statistics: Confidentiality Guidance* [Online]. Available: https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/healthstatistics [Accessed 17 Feb 2021].

PAIXAO, E. S., CAMPBELL, O. M. R., RODRIGUES, L. C., TEIXEIRA, M. G., COSTA, M., BRICKLEY, E. B. & HARRON, K. 2019. Validating linkage of multiple population-based administrative databases in Brazil. *PLoS One,* 14**,** e0214050.

PICKERING, L. K., GRANOFF, D. M., ERICKSON, J. R., MASOR, M. L., CORDLE, C. T., SCHALLER, J. P., WINSHIP, T. R., PAULE, C. L. & HILTY, M. D. 1998. Modulation of the immune system by human milk and infant formula containing nucleotides. *Pediatrics,* 101**,** 242-249.

PROTZKO, J. 2015. The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence,* 53**,** 202-210.

RAO, R. & GEORGIEFF, M. K. Early nutrition and brain development. The effects of early adversity on neurobehavioral development. Minnesota Symposium on Child Psychology, 2000. Erlbaum Associates Hillsdale, NJ, 1-30.

RHEAD, S. & BLACK, B. 2018. Marking consistency metrics: An update. *In:* OFQUAL (ed.). Coventry: Ofqual.

ROBERT BOYLE. 2020. *RE: Personal communication about plan to start a database on infant formula trials at Imperial College London.* Type to VERFUERDEN, M.

ROSEMAN, M., TURNER, E. H., LEXCHIN, J., COYNE, J. C., BERO, L. A. & THOMBS, B. D. 2012. Reporting of conflicts of interest from drug trials in Cochrane reviews: cross sectional study. *Bmj,* 345**,** e5155.

ROTHMAN, K. J., GREENLAND, S. & LASH, T. L. 2008. *Modern epidemiology*, Lippincott Williams & Wilkins.

RUBIN, D. B. 2004. *Multiple imputation for nonresponse in surveys*, John Wiley & Sons.

RUYS, C. A., BRÖRING, T., VAN SCHIE, P. E. M., VAN DE LAGEMAAT, M., ROTTEVEEL, J., FINKEN, M. J. J., OOSTROM, K. J. & LAFEBER, H. N. 2018. Neurodevelopment of children born very preterm and/or with a very low birth weight: 8-Year follow-up of a nutritional RCT. *Clin Nutr ESPEN,* 30**,** 190-198.

SAMMONS, P., SYLVA, K., MELHUISH, E., SIRAJ, I., TAGGART, B., TOTH, K. & SMEES, R. 2014. Influences on students' GCSE attainment and progress at age 16. *Department for Education*.

SCHALLER, J. P., KUCHAN, M. J., THOMAS, D. L., CORDLE, C. T., WINSHIP, T. R., BUCK, R. H., BAGGS, G. E. & WHEELER, J. G. 2004. Effect of dietary ribonucleotides on infant immune status. Part 1: Humoral responses. *Pediatric research,* 56**,** 883-890.

SCOTT, D. T., JANOWSKY, J. S., CARROLL, R. E., TAYLOR, J. A., AUESTAD, N. & MONTALTO, M. B. 1998. Formula supplementation with long-chain polyunsaturated fatty acids: are there developmental benefits? *Pediatrics,* 102**,** E59.

SINGHAL, A. 2015. Should we promote catch-up growth or growth acceleration in low-birthweight infants? *Low-birthweight baby: born too soon or too small,* 81**,** 51-60.

SINGHAL, A., COLE, T. J., FEWTRELL, M., KENNEDY, K., STEPHENSON, T., ELIAS-JONES, A. & LUCAS, A. 2007. Promotion of faster weight gain in infants born small for gestational age: is there an adverse effect on later blood pressure? *Circulation,* 115**,** 213-20.

SINGHAL, A., COLE, T. J. & LUCAS, A. 2001. Early nutrition in preterm infants and later blood pressure: two cohorts after randomised trials. *Lancet,* 357**,** 413-9.

SINGHAL, A., KENNEDY, K., LANIGAN, J., CLOUGH, H., JENKINS, W., ELIAS-JONES, A., STEPHENSON, T., DUDEK, P. & LUCAS, A. 2010a. Dietary nucleotides and early growth in formula-fed infants: a randomized controlled trial. *Pediatrics,* 126**,** e946-53.

SINGHAL, A., KENNEDY, K., LANIGAN, J., FEWTRELL, M., COLE, T. J., STEPHENSON, T., ELIAS-JONES, A., WEAVER, L. T., IBHANESEBHOR, S., MACDONALD, P. D., BINDELS, J. & LUCAS, A. 2010b. Nutrition in infancy and long-term risk of obesity: evidence from 2 randomized controlled trials. *The American Journal of Clinical Nutrition,* 92**,** 1133-1144.

SINGHAL, A., MACFARLANE, G., MACFARLANE, S., LANIGAN, J., KENNEDY, K., ELIAS-JONES, A., STEPHENSON, T., DUDEK, P. & LUCAS, A. 2008. Dietary nucleotides and fecal microbiota in formula-fed infants: a randomized controlled trial. *Am J Clin Nutr,* 87**,** 1785-92.

SPALDING, D. 1873. Instinct, with original observations on young animals. *Macmillan's Mag,* 27**,** 282-293.

STOLL, B. J., HANSEN, N. I., BELL, E. F., WALSH, M. C., CARLO, W. A., SHANKARAN, S., LAPTOOK, A. R., SÁNCHEZ, P. J., VAN MEURS, K. P. & WYCKOFF, M. 2015. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *Jama,* 314**,** 1039-1051.

SUN, H., COMO, P. G., DOWNEY, L. C., MURPHY, D., ARIAGNO, R. L. & RODRIGUEZ, W. 2015. Infant formula and neurocognitive outcomes: impact of study end-point selection. *J Perinatol,* 35**,** 867-74.

UCL. 2021. *RedCap data collection service* [Online]. Available: www.ucl.ac.uk/isd/it-for-slms/redcap-research-data-collection-service [Accessed 26 Jan 2021].

VERFUERDEN, M. L., DIB, S., JERRIM, J., FEWTRELL, M. & GILBERT, R. E. 2020. Effect of long-chain polyunsaturated fatty acids in infant formula on long-term cognitive function in childhood: A systematic review and meta-analysis of randomised controlled trials. *PLOS ONE,* 15**,** e0241800.

VERFÜRDEN, M., HARRON, K., JERRIM, J., FEWTRELL, M. & GILBERT, R. 2020. Infant formula composition and educational performance: a protocol to extend follow-up for a set of randomised controlled trials using linked administrative education records. *BMJ Open,* 10**,** e035968.

VICTORA, C. G., BAHL, R., BARROS, A. J., FRANÇA, G. V., HORTON, S., KRASEVEC, J., MURCH, S., SANKAR, M. J., WALKER, N. & ROLLINS, N. C. 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet,* 387**,** 475-490.

WALSH, V., BROWN, J. V. E., ASKIE, L. M., EMBLETON, N. D. & MCGUIRE, W. 2019. Nutrient-enriched formula versus standard formula for preterm infants. *Cochrane Database of Systematic Reviews.*

WALTER, T., PINO, P., PIZARRO, F. & LOZOFF, B. 1998. Prevention of iron-deficiency anemia: comparison of high- and low-iron formulas in term healthy infants after six months of life. *J Pediatr,* 132**,** 635-40.

WILLATTS, P., FORSYTH, J. S., DIMODUGNO, M. K., VARMA, S. & COLVIN, M. 1998. Effect of long-chain polyunsaturated fatty acids in infant formula on problem solving at 10 months of age. *Lancet,* 352**,** 688-91.

WILLATTS, P., FORSYTH, S., AGOSTONI, C., CASAER, P., RIVA, E. & BOEHM, G. 2013. Effects of long-chain PUFA supplementation in infant formula on cognitive function in later childhood. *Am J Clin Nutr,* 98**,** 536S-42S.

WOODS, J., WARD, G. & SALEM, N. 1996. Is Docosahexaenoic Acid Necessary in Infant Formula? Evaluation of High Linolenate Diets in the Neonatal Rat. *Pediatric Research,* 40**,** 687-694.

WORLD HEALTH ORGANIZATION 2008. Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia. *In:* BENOIST, B. D., MCLEAN, E., EGLL, I. & COGSWELL, M. (eds.) *Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.*

WWC 2014. WWC procedures and standards handbook (version 3.0). *Washington, DC: US Department of Education, Institute of Education Sciences, What Works Clearinghouse*, 23.

YAU, K.-I. T., HUANG, C.-B., CHEN, W., CHEN, S.-J., CHOU, Y.-H., HUANG, F.-Y., KUA, K. E., CHEN, N., MCCUE, M. & ALARCON, P. A. 2003. Effect of nucleotides on diarrhea and immune responses in healthy term infants in Taiwan. *Journal of pediatric gastroenterology and nutrition,* 36, 37-43.

YOUNG, L., EMBLETON, N. D. & MCGUIRE, W. 2016. Nutrient-enriched formula versus standard formula for preterm infants following hospital discharge. *Cochrane Database of Systematic Reviews*.

YU, V. Y. 1998. The role of dietary nucleotides in neonatal and infant nutrition. *Singapore Med J,* 39, 145-50.

ZIELINSKA, M. A., HAMULKA, J., GRABOWICZ-CHĄDRZYŃSKA, I., BRYŚ, J. & WESOLOWSKA, A. 2019. Association between Breastmilk LC PUFA, Carotenoids and Psychomotor Development of Exclusively Breastfed Infants. *International journal of environmental research and public health,* 16, 1144.

ZINGMOND, D. S., YE, Z., ETTNER, S. L. & LIU, H. 2004. Linking hospital discharge and death records—accuracy and sources of bias. *Journal of clinical epidemiology,* 57, 21-29.

ZIVANOVIC, S., PEACOCK, J., ALCAZAR-PARIS, M., LO, J. W., LUNT, A., MARLOW, N., CALVERT, S. & GREENOUGH, A. 2014. Late outcomes of a randomized trial of high-frequency oscillation in neonates. *N Engl J Med,* 370, 1121-1130.

ZOU, L., PANDE, G. & AKOH, C. C. 2016. Infant Formula Fat Analogs and Human Milk Fat: New Focus on Infant Developmental Needs. *Annu Rev Food Sci Technol,* 7, 139-65.

Appendix

A. Composition of trial formulas

Table A1 Formula composition in NEP-PD trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	68	72
Protein (g)	1.45	1.85
Casein	0.56	0.72
Whey	0.89	1.13
Carbohydrate (g)	6.96	7.24
Lactose (g)	6.96	6.20
Maltodextrin (g)	-	1.04
Fat (g)	3.82	3.96
Minerals		
Calcium (mg)	39	70
Phosphorus (mg)	27	35
Sodium (mg)	17	22
Chloride (mg)	45	45
Potassium (mg)	57	78
Iron (mg)	0.65	0.65
Zinc (mg)	0.34	0.60
Copper (µg)	42	57
Iodine (µg)	4.5	4.5
Magnesium (mg)	5.2	5.2
Manganese (µg)	3.4	5
Vitamins		
Vitamin A (µg)	100	100
Thiamine B_1 (mg)	42	95
Riboflavin B_2 (mg)	55	100
Pantothenic acid B_5 (mg)	0.23	0.40
Pyridoxine B_6 (µg)	35	80
Biotin B_7 (µg)	1.0	1.1
Folate B_9 (µg)	3.4	25
Cyanocobalamin B_{12} (µg)	0.14	0.2
Vitamin C (mg)	6.9	15
Vitamin D (µg)	1.0	1.3
Vitamin E (mg)	0.48	1.5
Vitamin K (µg)	2.7	6.0
Choline (mg)	4.8	5.1
Taurine (mg)	5	5.1
Carnitine (mg)	-	1.1

Table A2 Formula composition in NETSGA trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	68	72
Protein (g)	1.45	1.85
Casein	0.56	0.72
Whey	0.89	1.13
Carbohydrate (g)	6.96	7.24
Lactose (g)	6.96	6.20
Maltodextrin (g)	-	1.04
Fat (g)	3.82	3.96
Minerals		
Calcium (mg)	39	70
Phosphorus (mg)	27	35
Sodium (mg)	17	22
Chloride (mg)	45	45
Potassium (mg)	57	78
Iron (mg)	0.65	0.65
Zinc (mg)	0.34	0.60
Copper (µg)	42	57
Iodine (µg)	4.5	4.5
Magnesium (mg)	5.2	5.2
Manganese (µg)	3.4	5
Vitamins		
Vitamin A (µg)	100	100
Thiamine B_1 (mg)	42	95
Riboflavin B_2 (mg)	55	100
Pantothenic acid B_5 (mg)	0.23	0.40
Pyridoxine B_6 (µg)	35	80
Biotin B_7 (µg)	1.0	1.1
Folate B_9 (µg)	3.4	25
Cyanocobalamin B_{12} (µg)	0.14	0.2
Vitamin C (mg)	6.9	15
Vitamin D (µg)	1.0	1.3
Vitamin E (mg)	0.48	1.5
Vitamin K (µg)	2.7	6.0
Choline (mg)	4.8	5.1
Taurine (mg)	5.0	5.1
Carnitine (mg)	-	1.1

Table A3 Formula composition in LCPUFAP trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	70	70
Protein (g)	2.0	2.0
Casein	0.8	0.8
Whey	1.2	1.2
Carbohydrate (g)	7.7	7.7
Total fat (g)	3.5	3.5
Fatty acid composition (g/ 100 g fat)		
C8:0 caprylic	0.7	0.6
C10:0 capric	1.2	1.1
C12:0 lauric	6.3	4.9
C14:0 myristic	5.6	5.6
C16:0 palmitic	25.8	26.3
C18:0 stearic	8.2	8.5
C18:1 oleic	32.6	32.9
C18:2 n-6 linoleic	10.6	12.0
C18:3 n-6 γ-linolenic	0.1	0.4
C18:3 n-6 α-linolenic	0.7	0.6
C20:0 arachidic	0.4	0.3
C20:1 n-9 eicosanoic	0.2	0.3
C20:4 n-6 AA	-	0.31
C20:5 n-3 eicosapentaenoic	-	0.04
C22:0	0.3	0.2
C22:6 n-3 DHA	-	0.17
C24:0	0.2	0.1
Other fatty acids	7.1	4.62
Cholesterol	-	7.73
Minerals		
Calcium (mg)	70	70
Phosphorus (mg)	35	42
Sodium (mg)	30	27
Potassium (mg)	75	71
Iron (mg)	0.1	0.07
Zinc (mg)	0.39	0.4
Vitamins		
Vitamin A (µg)	63	63
Vitamin D (µg)	2.1	2.1
Vitamin E (mg)	2.0	2.0
Vitamin K (µg)	2.8	3.0

Table A4 Formula composition in LCPUFAT trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	67	67
Protein (g) (casein:whey 40:60)	1.5	1.5
Carbohydrate (g)	7.6	7.7
Lactose (g)	7.6	7.7
Total fat (g)	3.4	3.4
Fatty acid composition (% total fat)		
C8:0 caprylic	0.8	2.0
C10:0 capric	2.4	2.0
C12:0 lauric	1.6	12.3
C14:0 myristic	8.6	5.2
C16:0 palmitic	23.3	25.6
C18:0 stearic	10.2	4.6
C18:1 oleic	32.7	29.7
C18:2 n-6 linoleic	12.4	15.9
C18:3 n-6 α-linolenic	1.1	1.4
C20:4 n-6 AA	-	0.30
C20:5 n-3 eicosapentaenoic	-	0.01
C22:6 n-3 DHA	-	0.32
Cholesterol	<0.5	0.8
Minerals		
Iron (mg)	0.8	0.8

Table A5 Formula composition in IRON trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	65	65
Protein (g)	2.5	2.5
Carbohydrate (g)	8.0	8.0
Lactose (g)	6.4	6.4
Maltodextrin	1.6	1.6
Fat (g)	2.8	2.8
Saturated (%)	42.9	42.9
Unsaturated (%)	57.1	57.1
Minerals		
Calcium (mg)	100	100
Phosphorus (mg)	65	65
Magnesium (mg)	7.5	7.5
Sodium (mg)	30	30
Potassium (mg)	100	100
Chloride (mg)	70	70
Iron (mg)	0.9	12
Zinc (mg)	0.5	0.5
Iodine (µg)	6.9	6.9
Vitamins		
Vitamin A (µg)	6.9	6.9
Thiamine B_1 (mg)	0.8	0.8
Riboflavin B_2 (mg)	0.12	0.12
Niacin B_3 (µg)	0.61	0.61
Pantothenic acid B_5 (µg)	0.24	0.24
Pyridoxine B_6 (µg)	48	48
Biotin B_7 (µg)	1.71	1.71
Folate B_9 (µg)	6.0	6.0
Cyanocobalamin B_{12} (µg)	0.12	0.12
Vitamin C (µg)	6.6	6.6
Vitamin D (µg)	1.2	1.2
Vitamin E (mg)	0.75	0.75
Vitamin K (µg)	6.6	6.6

Table A6 Formula composition in PALM trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	70	73
Protein (g)	1.6	1.6
Carbohydrate (g)	7.1	7.1
Lactose (g)	7.1	7.1
Total fat (g)	3.9	4.2
Fatty acid composition (% total fat)		
C8:0 caprylic	1.6	1.5
C10:0 capric	0.5	0.4
C12:0 lauric	10.3	12.4
C14:0 myristic	4.9	4.7
C16:0 palmitic	19.6 (12% in sn-2)	20.1 (50% in sn-2)
C18:0 stearic	3.9	3.1
C18:1 n-9 oleic	41.3	42
C18:2 n-6 linoleic	12.1	13
C18:3 n-3 α-linolenic	1.9	1.6
Other fatty acids	4.0	1.2
Minerals		
Calcium (mg)	54	57
Phosphorus (mg)	32	33
Sodium (mg)	23	23
Potassium (mg)	76	74
Chloride (mg)	45	44
Magnesium (mg)	6.3	6.3
Manganese (μg)	5.1	5.2
Iron (mg)	0.51	0.57
Zinc (mg)	0.45	0.49
Copper (μg)	44	45
Iodine (μg)	10	10
Vitamins		
Vitamin A (μg)	92	95
Thiamine B_1 (μg)	40	40
Riboflavin B_2 (μg)	100	100
Pyridoxine B_6 (μg)	40	40
Vitamin D (μg)	1.1	1.1
Vitamin K (μg)	5	5
Vitamin E (mg)	1.1	1.1

Table A7 Formula composition in NEP-1 trial

per 100 ml	Banked breast milk	Modified formula
Energy (kcal)	46	80
Protein (g)	1.1	2.0
Carbohydrate (g)	7.1	7.0
Total fat (g)	1.7	4.9
Calcium (mg)	35	70
Phosphorus (mg)	15	35
Sodium (Na) (mg)	16	45
Potassium (mmol/L)	14.3	16.7

Table A8 Formula composition in NEP-2 trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	68	80
Protein (g) (casein: whey 40:60)	1.5	2.0
Carbohydrate (g)	7.0	7.0
Lactose (g)	7.0	6.0
Maltodextrin (g)	-	1.0
Total fat (g)	3.8	4.9
Saturated (%)	39.5	39.5
Unsaturated (%)	60.5	60.5
Minerals		
Calcium (mg)	35	70
Phosphorus (mg)	29	35
Sodium (mg)	19	45
Potassium (mg)	57	65
Iron (μg)	650	40
Zinc (mg)	350	1000
Copper (μg)	43	120
Iodine (μg)	4.5	7
Magnesium (mg)	5.2	5
Manganese (μg)	3.4	3
Vitamins		
Vitamin A (μg)	100	100
Thiamine B_1 (mg)	42	95
Riboflavin B_2 (mg)	55	180
Niacin B_3 (μg)	690	1000
Pantothenic acid B_5 (μg)	230	500
Pyridoxine B_6 (μg)	35	100
Biotin B_7 (μg)	1	2
Inositol B_8 (mg)	5	3.2
Folic Acid B_9 (μg)	3.4	50
Cyanocobalamin B_{12} (μg)	0.14	0.2
Vitamin C (μg)	6.9	28
Vitamin D (μg)	1.0	8.0
Vitamin K (μg)	2.7	7
Vitamin E (mg)	0.48	10
Choline (mg)	5	5.6
Taurine (mg)	-	5.1
Carnitine (mg)	2	1

Table A9 Formula composition in NUCLEO trial

per 100 ml	Standard formula	Modified formula
Energy (kcal)	68	68
Protein (g) (casein:whey 40:60)	1.5	1.5
Casein	0.6	0.6
Whey	0.9	0.9
Carbohydrate as lactose (g)	7.0	7.0
Total fat (g)	3.8	3.8
C18:2 n-6 linoleic (mg)	350	350
C18:3 n-6 γ-linolenic (mg)	33	33
C18:3 n-6 α-linolenic (mg)	44	44
LCPUFAs (mg)	26	26
Nucleotides		
Cytidine monophosphate (mg)	0.3	1.5
Uridine monophosphate (mg)	-	0.5
Adenosine monophosphate (mg)	-	0.6
Guanosine monophosphate (mg)	-	0.2
Inosine monophosphate (mg)	-	0.3
Minerals		
Calcium (mg)	39	39
Magnesium (mg)	5.2	5.2
Manganese (µg)	3.4	3.4
Phosphorus (mg)	27	27
Sodium (mg)	17	17
Potassium (mg)	0.3	0.3
Iron (mg)	0.6	0.6
Zinc (mg)	0.3	0.3
Copper (µg)	42	42
Iodine (µg)	4.5	4.5
Vitamins		
Vitamin A (µg)	100	100
Thiamine B_1 (mg)	42	42
Riboflavin B_2 (mg)	55	55
Niacin B_3 (mg)	0.7	0.7
Pantothenic acid B_5 (µg)	0.2	0.2
Biotin B_7 (µg)	1.0	1.0
Vitamin B_8 (µg)	35	35
Folic Acid B_9 (µg)	3.4	3.4
Cyanocobalamin B_{12} (µg)	0.14	0.14
Vitamin C (mg)	6.9	6.9
Vitamin D (µg)	1.0	1.0
Vitamin E (mg)	0.5	0.5
Vitamin K (µg)	2.7	2.7
Choline (mg)	4.8	4.8
Taurine (mg)	5.0	5.0

B. Indicators of trial quality

Table A10 Risk of bias assessment using Cochrane risk of bias tool. Risk of bias summary showing my judgment about each risk of bias item for the trials (items 1-3).

RCT	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding* (information bias)
Enriched formula vs. standard term formula for preterms 1982-84	+	+	-
Enriched formula vs. banked breast milk for preterms 1982-84	+	+	-
Enriched post-discharge formula vs. standard term formula for preterms 1993-96	+	+	+
Enriched formula vs. standard term formula for SGA terms 1993-96	+	+	+
LCPUFA supplemented formula vs. unsupplemented preterm formula for preterms 1993-96	+	+	+
LCPUFA supplemented formula vs. unsupplemented standard formula for terms 1993-95	+	+	+
Nucleotide supplemented formula vs. unsupplemented standard formula for terms 2000-02	+	+	+
Iron supplemented follow-on formula vs. formula supplemented with low iron vs. cow milk for terms 1993-94	+	+	+
Sn-2 palmitate supplemented formula vs. standard term formula for terms 1995-96	+	+	+

* In this thesis I remained blind to the allocation of each participant until the data analysis plan was peer-reviewed.

Table A11 Cochrane risk of bias items 1-3 for all dormant trials discussed in this thesis

Trial	Randomisation sequence generation	Allocation concealment	Blinding
NEP-1	Stratified randomisation sequence for each centre, within strata defined by birth weight: less than 1200 g and 1200 to 1849 g. A member of the team who was not involved in subsequent aspects of the trial prepared the sequence (DOI: 10.1007/s11745-999-0353-0.)	Treatment assignments were held in sealed numbered opaque envelopes separately in each centre	No attempt was made to blind clinical staff to the type of diet employed as the diets are distinguishable in practice
NEP-2	Stratified randomisation sequence for each centre, within strata defined by birth weight: less than 1200 g and 1200 to 1849 g. A member of the team who was not involved in subsequent aspects of the trial prepared the sequence (DOI: 10.1007/s11745-999-0353-0.)	Treatment assignments were held in sealed numbered opaque envelopes separately in each centre	No attempt was made to blind clinical staff to the type of diet employed as the diets are distinguishable in practice
NEP-PD	The randomisation schedule was generated by permuted blocks of randomised length and was stratified by birth weight (< or >1200 g), whether or not infants required supplemental oxygen for >28 days, and by the number of fetuses (twins or triplets were randomised as one onto the same diet). A member of the team who was not involved in subsequent aspects of the trial prepared randomisation assignments.	Dietary allocations stored in sealed, numbered and opaque envelopes.	The formulas were colour-coded; the codes were held by the formula manufacturers and were not revealed to the investigators until after the principal data analyses were performed. Both formulas were identical in colour and smell.
NETSGA	The randomisation schedule was generated by random permuted blocks; the subjects were stratified by race (white or Asian) and by birth weight centile. Randomisation assignments were prepared by a member of the team who was not involved in subsequent aspects of the study.	Dietary allocations stored in sealed, numbered and opaque envelopes.	The formulas were colour-coded, and the code was held by Farley Health Products and not revealed to the investigators until after the preliminary data analysis. Therefore, parents and study personnel were blinded to the dietary allocation throughout the study, follow-up, and initial data analyses.

LCPUFAP	The allocation schedule for each centre was generated using permuted blocks of randomised length by personnel who were not involved in subsequent aspects of the study. Infants were randomly assigned to receive one trial formula. Randomisation was stratified by birth weight (<1200 g or >1200 g) to increase the likelihood that the smallest, sickest infants would be equally distributed between feed groups. Twins and triplets were randomised separately.	Dietary allocations stored in sealed, numbered and opaque envelopes.	The two formulas were provided in colour-coded containers. The code was held by the formula manufacturers and was not broken until the study and analyses were completed. Therefore, parents and study personnel were blind to the dietary allocation throughout the study, follow-up, and data analysis periods.
LCPUFAT	A random permuted block design, stratified by centre (Nottingham or Leicester) and infant sex, was used.	Dietary allocations stored in sealed, numbered and opaque envelopes.	The subjects and their guardians remained blinded to the formulas used, and the field workers (who performed the cognitive tests) were also unaware of the dietary assignment.
NUCLEO	Random permuted-block design, stratified according to centre (Leicester or Nottingham), was generated by an independent statistician.	Dietary allocations stored in sealed, numbered and opaque envelopes.	All mothers and research staff members were blinded to the identity of the formula.
IRON	The randomisation schedules were prepared by an independent statistician using permuted blocks of random length. Each centre had a separate schedule, and subjects were randomised by the research nurse from consecutively numbered opaque sealed envelopes.	Dietary allocations stored in sealed, numbered and opaque envelopes.	The formula milks were supplied in powdered form; tins of iron-fortified formula were labelled "formula 28", and tins of unfortified formula were labelled "formula 61". The manufacturers did not reveal this code until the study was completed, and all data had been entered and checked.
PALM	Double-blind random permuted block allocation with dietary assignments identified by a barcode.	Dietary allocations stored in sealed, numbered and opaque envelopes.	Formula packaging was identical except for differences in the barcodes; investigators and parents were blinded to the dietary allocation.

C. What might have contributed to missed links

To anticipate the statistical power available for this PhD project ahead of receiving the linked data, I estimated the coverage of NPD attainment modules and the NPD as a whole for each trial. Linkage to the NPD used in this thesis did not require consent. This eliminates drop-out from participants who cannot be reached or decline to respond. As a result, there are only limited reasons for why some of the participants might not be found in any of the NPD datasets:

- Death
- Emigration
- Never interacted with a school that transmits data to the NPD
- Missing identifiers

My estimates are summarised in Table A13. Below, I discuss each factor in more detail.

Death

The background rate for child mortality under 15 years is very low in the UK (11.6 per 100,000 children) but generally higher in infancy (4 per 1,000 children) and for children born small or preterm (Wolfe et al., 2014). As trials attempted follow-up over several years, I expected that the large majority of participant deaths was known and recorded in the trial data. Table A13 shows the number of observed deaths for each trial.

Emigration

The percentage of school-aged children (age <15 years) emigrating out of the UK in the past 20 years was about 0.25% (Office for National Statistics, 2020). Assuming trial participants were representative of the UK population, I expected that fewer than 1 participant per trial to have emigrated. However, this does not apply to the NEP-1 and NEP-2 trials, where a proportion of infants were recruited from families working at a nearby US airbase leading to higher emigration rates than in the other trials.

Beyond anecdotes, I, unfortunately, could not recover information on the identity of these infants, nor was there any information on the number of participants concerned. As a result, I expected emigration to be a driving factor for less than optimal coverage of the NPD in the NEP-1 and NEP-2 trials but not in the other trials.

Never interacting with a school that transmits data to the NPD

As discussed above, the percentage of children who have never interacted with schools that submit data to the NPD is low: about 3%. Applied to the overall number of participants, I expected this to equate to about 77 participants (Table A13).

Missing participant identifiers

On the NPD side, identifiers are confirmed annually at the school census. On the trial side, the risk of data entry errors in the identifier data for each trial goes down as the number of follow-ups (interactions with participants) goes up. This is because name and address information is collected repeatedly over this period, allowing potential mistakes to be corrected. Trial follow-ups during school-age also increase the chance that the addresses on record corresponds to the addresses held in the NPD. Only the IRON and the NUCLEO trial did not attempt to follow up participants after the age of 5 years. Instead, they only had interactions with participants in the first year. Our research team previously analysed the percentage of postcode changes by age group using data from the Patient Demographic Service (PDS). They found that by the age of 6 years, about 70% of children had changed postcodes at least once (Table A12). Based on previously published data (Hagger-Johnson et al., 2015), I assumed conservatively that a missing postcode leads to a 20% probability of a missed link (depending on the completeness of the other variables). Applying this information to the participants from the IRON and NUCLEO studies, I multiplied the number eligible for linkage based on month and year of birth by 0.7 and then by 0.2 to calculate the expected losses from missing postcode information. This led to a conservative estimation o expected losses from missing postcode information because linkage did, in fact, take

into account neighbouring authorities, so a move within the vicinity of the recorded postcode would still enable linkage to the NPD.

KS2 and 4 are the attainment modules with nationally administered, mandatory exams. Table A13 shows the expected coverage for these modules and the NPD as a whole based on the factors discussed above. I arrived at the expected coverage by subtracting the expected losses from each cause from the eligible number linkable to a module based on year of birth, not considering cases where multiple causes might apply to a single participant. The resulting expected coverage was slightly higher compared to those reported in the Avon Longitudinal Study of Parents and Children, where 82% of participants linked to the NPD (Boyd et al., 2013).

Table A12 Number of postcode changes according to the PDS by age group (Table provided by Dr Gareth Hagger Johnson)

	Number of distinct postcodes in PDS since 2004 (or birth)			
	% 1 postcode	% 2 postcodes	% 3 postcodes	%4+ postcodes
Age 0/1	44.5%	33.1%	14.4%	8.0%
Age 5/6	31.0%	31.5%	18.3%	19.2%
Age 18/19	43.0%	24.0%	15.1%	17.9%

Study period: 2011/12

Table A13 Expected number of missed links from death, emigration and never interaction with organisation submitting data to the NPD, by trial

	NEP-1	NEP-2	NEP-PD	NETSGA	LCPUFAP	LCPUFAT	NUCLEO	IRON	PALM
N randomised	423	369	228	299	196	309	196	493	203
Eligible for any NPD module based on date of birth	158	153	228	299	196	309	196	493	203
Eligible for KS2 module based on date of birth	24	-	228	299	196	309	196	493	203
Eligible for KS4 module based on year and month of birth	-	-	228	299	196	309	58	493	203
Exp. loss from death (observed)	46	48	8	0	21	0	0	0	0
Exp. loss from emigration (assumed to be 0.25%)	1	1	1	1	0	1	0	1	1
Exp. loss from never interaction with NPD organisation (assumed 3%)	13	11	7	9	6	9	6	10	6
Exp. loss from not up-to date postcode information ((N x 0.7)* 0.20)							27	69	
Exp. loss from missing / incorrect identifiers (observed missing + 2% assumed to be incorrect)	73	52	9	22	15	13	6	15	9

[redacted]

[redacted]

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

[redacted]
Email: [redacted]
UCL Institute of Child Health
30 Guilford Street
WC1N 1EH

11 February 2019

Dear [redacted]

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Methods of linking dormant trial data to determine the long-term effects of enriched nutrition in infancy.**
IRAS project ID:	**212148**
REC reference:	**17/LO/0556**
Sponsor	**UCL Institute of Child Health**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

How should I continue to work with participating NHS organisations in England and Wales?
You should now provide a copy of this letter to all participating NHS organisations in England and Wales, as well as any documentation that has been updated as a result of the assessment.

This is a single site study sponsored by a partner academic institution, under joint research governance arrangements. The Joint R&D office will confirm to you when the study can start following issue of HRA and HCRW Approval.

It is important that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details of the research management function for each organisation can be accessed here.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within the devolved administrations of Northern Ireland and Scotland.

[redacted]

[redacted]

IRAS project ID	212148

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) has been sent to the coordinating centre of each participating nation. You should work with the relevant national coordinating functions to ensure any nation specific checks are complete, and with each site so that they are able to give management permission for the study to begin.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?
The document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

I am a participating NHS organisation in England or Wales. What should I do once I receive this letter?
You should work with the applicant and sponsor to complete any outstanding arrangements so you are able to confirm capacity and capability in line with the information provided in this letter.

The sponsor contact for this application is as follows:

Name: [redacted]
Tel: [redacted]
Email: [redacted]

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **212148**. Please quote this on all correspondence.

Yours sincerely

[redacted]
Assessor
Email: [redacted]

IRAS project ID	212148

Copy to: *Sponsor and Lead NHS R&D Office Representative: Ms Emma Pendleton, UCL Institute of Child Health*

IRAS project ID	212148

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Confirmation of any other Regulatory Approvals (e.g. NIGB) and all correspondence [CAG form]		
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Appendix D Insurance]	1	23 February 2017
IRAS Application Form [IRAS_Form_28032017]		28 March 2017
Letter from funder [SIRO Letter]		10 March 2017
Letter from sponsor [Sponsor Award Letter]	1	12 July 2016
Other [Appendix B]	1	23 February 2017
Other [Appendix C]	1	23 February 2017
Other [Appendix E Ethics reference numbers]	1	01 March 2017
Other [CAG queries responses]	1	28 March 2017
Other [CAG correspondence]		28 March 2017
Other [Consent forms 1]		28 March 2017
Other [Consent forms 2]		28 March 2017
Other [Consent forms 3]		28 March 2017
Other [Consent forms 4]		28 March 2017
Other [Consent forms 5]		28 March 2017
Other [Consent forms 6]		28 March 2017
Other [Consent forms 7]		28 March 2017
Other [Consent forms 8]		28 March 2017
Other [Consent forms 9]		28 March 2017
Other [Consent forms 10]		28 March 2017
Other [Consent forms 11]		28 March 2017
Other [Consent forms 12]		28 March 2017
Other [Consent forms 13]		28 March 2017
Other [Consent forms 14]		28 March 2017
Other [Consent forms 15]		28 March 2017
Other [Consent forms 16]		28 March 2017
Other [Consent forms 17]		28 March 2017
Other [Consent forms 18]		28 March 2017
Other [Consent forms 19]		28 March 2017
Other [Consent forms 20]		28 March 2017
Other [Consent forms 21]		28 March 2017
Research protocol or project proposal	1	02 March 2017
Research protocol or project proposal [Research Protocol]	1	15 March 2017
Summary CV for Chief Investigator (CI)	1	23 February 2017
Summary CV for Chief Investigator (CI) [CV]	1	23 February 2017
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Appendix A]	1	23 February 2017
17 LO 0556_Application_valid Letter 28.03.17.pdf		28 March 2017

IRAS project ID	212148

17 LO 0556_Favourable_opinion_at_first_review 22.04.17.rtf.pdf		22 April 2017
17CAG0051 Provisional Outcome.pdf		23 May 2017

IRAS project ID	212148

Summary of assessment

The following information provides assurance to you, the sponsor and the NHS in England and Wales that the study, as assessed for HRA and HCRW Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England and Wales to assist in assessing, arranging and confirming capacity and capability.

Assessment criteria

Section	Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	An agreement is not expected as Joint Research Office arrangements are in place between the sponsor and the participating NHS organisation.
4.2	Insurance/indemnity arrangements assessed	Yes	No comments
4.3	Financial arrangements assessed	Yes	The sponsor secured funding from the Great Ormond Street Hospital Children's Charity. A copy of the funding award letter was received.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments

IRAS project ID	212148

Section	Assessment Criteria	Compliant with Standards	Comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	NHS Research Ethics Committee favourable opinion was confirmed by the London - City & East Research Ethics Committee on 22 April 2017.
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England and Wales

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.
This is a single site study; there is therefore one site type. If this study is subsequently extended to other NHS organisation(s) in England or Wales, an amendment should be submitted, with a Statement of Activities and Schedule of Events for the newly participating NHS organisation(s) in England or Wales. The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England and Wales in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. Where applicable, the local LCRN contact should also be copied into this correspondence. If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England and Wales which are not provided in IRAS, the HRA or HCRW websites, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net or HCRW at Research-permissions@wales.nhs.uk. We will work with these organisations to achieve a consistent approach to information provision.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and Wales, and the minimum expectations for education, training and experience that PIs should meet (where applicable).
A Principal Investigator should be in place at participating NHS organisations in England. The Chief Investigator will take on this role at the sole participating site.

IRAS project ID	212148

GCP training is not a generic training expectation, in line with the HRA/HCRW/MHRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
The research has contractual arrangements in place with the participating NHS organisation.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales to aid study set-up.
The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

E. Overview of search terms and strategies for each modification

Searches were last conducted in Sept 2020. The search terms to identify systematic reviews in MEDLINE® and EMBASE were adapted from the Scottish Intercollegiate Guidelines Network.

Tables A14

MEDLINE® terms for cognitive ability
("Cognition"[MeSH Terms] OR "Child Development"[Mesh] OR "Intelligence"[Mesh] OR "Brain/growth and development"[Mesh] OR "Cognition"[TIAB] OR "Child Development"[TIAB] OR "Intelligence"[TIAB] OR "Cognitive function"[TIAB] OR "Learning"[TIAB] OR "Cognitive test"[TIAB] OR "Brain"[TIAB] OR "neuro*development"[TIAB] OR "educational status"[MeSH Terms] OR "education*"[TIAB] OR "educational status"[TIAB] OR "schools"[MeSH Terms] OR "schools"[TIAB] OR "school"[TIAB] OR attainment[TIAB] OR "Bayley"[TIAB])

MEDLINE® terms for infant nutrition
AND ("Infant Formula"[Mesh] OR "Infant Food"[Mesh] OR "Infant Nutritional Physiological Phenomena"[Mesh] OR "Food, Fortified"[MAJR] OR "Nutritional Support"[TIAB] OR "follow*on*formula"[TIAB] OR "supplementation"[TIAB])

MEDLINE® terms for study population
(("infant"[MeSH Terms] OR "child"[MeSH Terms] OR "adolescent"[MeSH Terms]) OR (infant[TIAB] OR child[TIAB] OR adolescent[TIAB]))

MEDLINE® terms for systematic review of RCTs
(Review[ptyp] OR ((systematic review[ti] OR meta-analysis[pt] OR meta-analysis[ti] OR systematic literature review[ti] OR this systematic review[tw] OR pooling project[tw] OR (systematic review[tiab] AND review[pt]) OR meta synthesis[ti] OR (meta analysis[ti] OR meta analyses[ti] OR meta analyse[ti] OR meta analysed[ti] OR meta analyser[ti] OR meta analyses[ti] OR meta analysing[ti] OR meta analysis[ti] OR meta analysis,[ti] OR meta analysisdagger[ti] OR meta analysis of[ti] OR meta analyst[ti] OR meta analysticians[ti] OR meta analysts[ti] OR meta analysys[ti] OR meta analytic[ti] OR meta analytical[ti] OR meta analytically[ti] OR meta analyze[ti] OR meta analyzed[ti] OR meta analyzes[ti] OR meta analyzing[ti]) OR integrative review[tw] OR integrative research review[tw] OR rapid review[tw] OR umbrella review[tw] OR consensus development conference[pt] OR practice guideline[pt] OR drug class reviews[ti] OR "Cochrane Database Syst Rev"[Journal] OR "ACP J Club"[Journal] OR "Health Technol Assess"[Journal] OR "Evid Rep Technol Assess (Summ)"[Journal] OR "JBI Database System Rev Implement Rep"[Journal]) OR (clinical guideline[tw] AND management[tw]) OR ((evidence based[ti] OR "evidence-based medicine"[MeSH Terms] OR (best practice[ti] OR best practices[ti]) OR evidence synthesis[tiab]) AND (review[pt] OR diseases category[mh] OR "behaviour and behaviour mechanisms"[MeSH Terms] OR "therapeutics"[MeSH Terms] OR evaluation studies[pt] OR validation studies[pt] OR guideline[pt] OR pmcbook[All Fields])) OR ((systematic[tw] OR systematically[tw] OR critical[tiab] OR study selection[tw] OR (predetermined[tw] OR inclusion[tw] AND (criteri[tw] OR criteria[tw] OR criteria'[tw] OR criteria'double[tw] OR criteria's[tw] OR criteria'srandomized[tw] OR criteria1[tw] OR criteria2[tw] OR criteriaadult[tw] OR criteriaall[tw] OR criteriaare[tw] OR criteriabased[tw] OR criteriadisulfiram[tw] OR criteriae[tw] OR criteriaeditorials[tw] OR criteriaen[tw] OR criteriaenglish[tw] OR criteriaexclusion[tw] OR criteriafor[tw] OR criteriafora[tw] OR criteriaheath[tw] OR criteriai[tw] OR criteriaincluded[tw] OR criteriaithe[tw] OR criterial[tw] OR

criterialism[tw] OR criteriality[tw] OR criteriall[tw] OR criterially[tw] OR criterials[tw] OR criterian[tw] OR criteriaof[tw] OR criteriar[tw] OR criteriarandomised[tw] OR criteriarpar[tw] OR criterias[tw] OR criteriasof[tw] OR criteriastudies[tw] OR criteriasystematic[tw] OR criteriathe[tw] OR criteriation[tw] OR criteriatrade[tw] OR criteriaum[tw] OR criteriawerehaving[tw] OR criteric[tw] OR criterid[tw] OR criterien[tw] OR criteries[tw] OR criteriia[tw] OR criterin[tw] OR criterio[tw] OR criterioe[tw] OR criteriologic[tw] OR criteriological[tw] OR criteriology[tw] OR criterion[tw] OR criterion'[tw] OR criterion's[tw] OR criterional[tw] OR criterionby[tw] OR criterionis[tw] OR criterionoriented[tw] OR criterions[tw] OR criterior[tw] OR criteriors[tw] OR criterios[tw] OR criteriosa[tw] OR criteriosamente[tw] OR criterioso[tw] OR criterious[tw] OR criteris[tw] OR criterita[tw] OR criterium[tw] OR criterium'[tw] OR criteriums[tw] OR criterization[tw])) OR (exclusion criteria[tw] OR exclusion criterias[tw] OR exclusion criterion[tw] OR exclusion criterions[tw] OR exclusion criterium[tw]) OR main outcome measures[tw] OR standard of care[tw] OR standards of care[tw]) AND (survey[tiab] OR surveys[tiab] OR (overview[tw] OR overview'[tw] OR overview's[tw] OR overview2[tw] OR overviewed[tw] OR overviewer[tw] OR overviewers[tw] OR overviewes[tw] OR overviewing[tw] OR overviewn[tw] OR overviewon[tw] OR overviewpredictive[tw] OR overviewprognostic[tw] OR overviews[tw] OR overviews'[tw] OR overviews''[tw] OR overviewstudy[tw]) OR review[tiab] OR reviews[tiab] OR (search[tw] OR search'[tw] OR search's[tw] OR search010[tw] OR search013[tw] OR search1[tw] OR search5[tw] OR searchability[tw] OR searchable[tw] OR searchableby[tw] OR searchall[tw] OR searchamerica[tw] OR searchand[tw] OR searchback[tw] OR searchbreast[tw] OR searchcoil[tw] OR searchcompare[tw] OR searchdb[tw] OR searchdisease[tw] OR searchdogs[tw] OR searche[tw] OR searcheable[tw] OR searched[tw] OR searched'[tw] OR searched19[tw] OR searchedfor[tw] OR searchedmedline[tw] OR searchedwas[tw] OR searcheed[tw] OR searchen[tw] OR searcher[tw] OR searcher'[tw] OR searcher's[tw] OR searchers[tw] OR searchers'[tw] OR searches[tw] OR searches'[tw] OR searchescohorts[tw] OR searchfor[tw] OR searchform[tw] OR searchgenes[tw] OR searchgtr[tw] OR searchgui[tw] OR searchhes[tw] OR searchin[tw] OR searchin'[tw] OR searching[tw] OR searching'[tw] OR searchinger[tw] OR searchingfor[tw] OR searchingly[tw] OR searchings[tw] OR searchlight[tw] OR searchlight'[tw] OR searchlights[tw] OR searchlighttrade[tw] OR searchline[tw] OR searchlite[tw] OR searchlyte[tw] OR searchmedica[tw] OR searchomyces[tw] OR searchpageeng[tw] OR searchpath[tw] OR searchpaths[tw] OR searchpatterns[tw] OR searchpattool[tw] OR searchpks[tw] OR searchproj[tw] OR searchresult[tw] OR searchs[tw] OR searchshowed[tw] OR searchsmallrna[tw] OR searchsnp[tw] OR searchtesv[tw] OR searchtm[tw] OR searchtrade[tw] OR searchtxt[tw] OR searchtype[tw] OR searchwise[tw] OR searchxlinks[tw]) OR handsearch[tw] OR analysis[ti] OR critique[tiab] OR appraisal[tw] OR (reduction[tw] AND ("risk"[MeSH Terms] OR risk[tw]) AND (("death"[MeSH Terms] OR "death"[All Fields]) OR ("recurrence"[MeSH Terms] OR "recurrence"[All Fields])))) AND (literature[tiab] OR articles[tiab] OR publications[tiab] OR publication[tiab] OR bibliography[tiab] OR bibliographies[tiab] OR published[tiab] OR pooled data[tw] OR unpublished[tw] OR citation[tw] OR citations[tw] OR database[tiab] OR internet[tiab] OR textbooks[tiab] OR references[tw] OR scales[tw] OR papers[tw] OR datasets[tw] OR trials[tiab] OR (meta analyis[tw] OR meta analyisis[tw] OR meta analysable[tw] OR meta analysas[tw] OR meta analyse[tw] OR meta analysed[tw] OR meta analysei[tw] OR meta analysen[tw] OR meta analyser[tw] OR meta analysers[tw] OR meta analyses[tw] OR meta analysescohort[tw] OR meta analysespublication[tw] OR meta analysestype[tw] OR meta analysi[tw] OR meta analysia[tw] OR meta analysic[tw] OR meta analysing[tw] OR meta analysis[tw] OR meta analysis's[tw] OR meta analysis,[tw] OR meta analysis12[tw] OR meta analysis2[tw] OR meta analysisbone[tw] OR meta analysisdagger[tw] OR meta analysises[tw] OR meta analysisevaluating[tw] OR meta analysisif[tw] OR meta analysisindicated[tw] OR meta analysisintroduction[tw] OR meta analysisjr[tw] OR meta analysismethods[tw] OR meta analysismoderate[tw] OR meta analysisof[tw] OR meta analysistrade[tw] OR meta analysisv[tw] OR meta

analysisxs[tw] OR meta analysized[tw] OR meta analyst[tw] OR meta analysticians[tw] OR meta analysts[tw] OR meta analysys[tw] OR meta analytic[tw] OR meta analytical[tw] OR meta analytically[tw] OR meta analytics[tw] OR meta analyzable[tw] OR meta analyze[tw] OR meta analyzed[tw] OR meta analyzes[tw] OR meta analyzing[tw]) OR (clinical[tiab] AND studies[tiab]) OR "treatment outcome"[MeSH Terms] OR treatment outcome[tw] OR pmcbook[All Fields])) NOT (letter[pt] OR newspaper article[pt])) OR (((randomized controlled trial[pt]) OR (controlled clinical trial[pt]) OR (randomized[tiab] OR randomised[tiab]) OR (placebo[tiab]) OR (drug therapy[sh]) OR (randomly[tiab]) OR (trial[tiab]) OR (groups[tiab])) NOT (animals[mh] NOT humans[mh]))))

Enriched post-discharge formula for preterm infants
Embase

#	Searches
1	exp Meta Analysis/
2	((meta adj analy$) or metaanalys$).tw.
3	(systematic adj (review$1 or overview$1)).tw.
4	or/1-3
5	cancerlit.ab.
6	cochrane.ab.
7	embase.ab.
8	(psychlit or psyclit).ab.
9	(psychinfo or psycinfo).ab.
10	(cinahl or cinhal).ab.
11	science citation index.ab.
12	bids.ab.
13	or/5-12
14	reference lists.ab.
15	bibliograph$.ab.
16	hand-search$.ab.
17	manual search$.ab.
18	relevant journals.ab.
19	or/14-18
20	data extraction.ab.
21	selection criteria.ab.
22	20 or 21
23	review.pt.
24	22 and 23
25	letter.pt.
26	editorial.pt.
27	animal/
28	human/
29	27 not (27 and 28)
30	or/25-26,29
31	4 or 13 or 19 or 24
32	31 not 30
33	infant nutrition.tw. or exp infant nutrition/ or nutritional support.tw. or exp nutritional support/ or infant$food.tw. or infant$diet.tw. or baby$food.tw. or baby$diet.tw. or exp baby food/ or exp artificial milk/
34	exp cognition assessment/ or Cognition.tw. or exp cognition/ or exp social cognition/ or IQ.tw. or intelligence.tw. or exp Wechsler adult intelligence scale/ or

	exp intelligence quotient/ or exp Wechsler intelligence scale/ or exp "Wechsler preschool and primary scale of intelligence"/ or exp intelligence/ or exp Wechsler intelligence scale for children/ or exp intelligence test/ or exp emotional intelligence/ or Intelligence.mp. or exp Stanford-Binet Intelligence Scale/ or (neurodevelopment or cognitive).tw. or child$development.tw. or exp child development/ or exp neuropsychological test/ or exp child development/ or exp "Bayley Scales of Infant Development"/ or exp mental development/ or bayley.tw. or exp reading/ or exp school/ or exp learning disorder/ or exp educational status/ or exp education/ or exp achievement/ or exp academic achievement/ or school attainment.tw.
35	(child$ or infant$ or baby or babies or adolescent$ or teenager$).tw.
36	dietary proteins.tw. or exp protein intake/ or energy intake.tw. or exp caloric intake/ or enriched formula.tw. or exp enteric feeding/ or exp diet supplementation/ or preterm$formula.tw.
37	exp premature labor/ or exp prematurity/ or exp low birth weight/ or preterm$.tw. or premie$.tw. or low gestational age.tw. or low birthweight.tw. or exp low birth weight/ or small for gestational age.tw. or smallfordate$.tw. or exp small for date infant/
38	exp hospital discharge/ or post$discharge.tw. or after$discharge.tw. or following$discharge.tw. or hospital$discharge.tw.
39	32 and 33 and 34 and 35 and 36 and 37 and 38

MEDLINE

MEDLINE terms for cognitive ability, study population, premature infants, systematic review and infant nutrition (see above) plus:

Dietary Proteins"[Mesh] OR "Energy Intake"[Mesh] OR "Nutritional Support"[Mesh] OR "Dietary Proteins"[TIAB] OR "Energy Intake"[TIAB] OR "enriched formula"[TIAB] OR "supplemented formula"[TIAB] OR "Nutrient*enriched"[TIAB] OR "preterm*formula"[TIAB]) AND (("Hospital*"[TIAB] OR Discharge[TIAB]) AND ("after"[TIAB] OR "post"[TIAB] OR "following"[TIAB]))

LCPUFA-enriched infant formula for preterm infants

Embase

#	Searches
1	exp Meta Analysis/
2	((meta adj analy$) or metaanalys$).tw.
3	(systematic adj (review$1 or overview$1)).tw.
4	or/1-3
5	cancerlit.ab.
6	cochrane.ab.
7	embase.ab.
8	(psychlit or psyclit).ab.
9	(psychinfo or psycinfo).ab.
10	(cinahl or cinhal).ab.
11	science citation index.ab.
12	bids.ab.
13	or/5-12
14	reference lists.ab.
15	bibliograph$.ab.
16	hand-search$.ab.
17	manual search$.ab.

18	relevant journals.ab.
19	or/14-18
20	data extraction.ab.
21	selection criteria.ab.
22	20 or 21
23	review.pt.
24	22 and 23
25	letter.pt.
26	editorial.pt.
27	animal/
28	human/
29	27 not (27 and 28) (89)
30	or/25-26,29
31	4 or 13 or 19 or 24
32	31 not 30
33	infant nutrition.tw. or exp infant nutrition/ or nutritional support.tw. or exp nutritional support/ or infant$food.tw. or infant$diet.tw. or baby$food.tw. or baby$diet.tw. or exp baby food/ or exp artificial milk/
34	exp cognition assessment/ or Cognition.tw. or exp cognition/ or exp social cognition/ or IQ.tw. or intelligence.tw. or exp Wechsler adult intelligence scale/ or exp intelligence quotient/ or exp Wechsler intelligence scale/ or exp "Wechsler preschool and primary scale of intelligence"/ or exp intelligence/ or exp Wechsler intelligence scale for children/ or exp intelligence test/ or exp emotional intelligence/ or Intelligence.mp. or exp Stanford-Binet Intelligence Scale/ or (neurodevelopment or cognitive).tw. or child$development.tw. or exp child development/ or exp neuropsychological test/ or exp child development/ or exp "Bayley Scales of Infant Development"/ or exp mental development/ or bayley.tw. or exp reading/ or exp school/ or exp learning disorder/ or exp educational status/ or exp education/ or exp achievement/ or exp academic achievement/ or school attainment.tw.
35	(child$ or infant$ or baby or babies or adolescent$ or teenager$).tw.
36	exp fatty acids, omega-3/ or fatty acids, essential/ or Dietary Fats, Unsaturated/ or linolenic acids/ or exp fish oils/ or (n 3 fatty acid$ or omega 3).tw. or docosahexa?noic.tw,hw,rw. or eicosapenta?noic.tw,hw,rw. or alpha linolenic.tw,hw,rw. or (linolenate or cervonic or timnodonic).tw,hw,rw. or menhaden oil$.tw,hw,rw. or (mediterranean adj diet$).tw. or ((flax or flaxseed or flax seed or linseed or rape seed or rapeseed or canola or soy or soybean or walnut or mustard seed) adj2 oil$).tw. or (walnut$ or butternut$ or soybean$ or pumpkin seed$).tw. or (fish adj2 oil$).tw. or (cod liver oil$ or marine oil$ or marine fat$).tw. or (salmon or mackerel or herring or tuna or halibut or seal or seaweed or anchov$).tw. or (fish consumption or fish intake or (fish adj2 diet$)).tw. or diet$ fatty acid$.tw. or borage oil$.tw.
37	exp premature labor/ or exp prematurity/ or exp low birth weight/ or preterm$.tw. or premie$.tw. or low gestational age.tw. or low birthweight.tw. or exp low birth weight/ or small for gestational age.tw. or smallfordate$.tw. or exp small for date infant/
38	32 and 33 and 34 and 35 and 36 and 37

MEDLINE

MEDLINE terms for cognitive ability, study population, premature infants, systematic review and infant nutrition (see above) plus:

("Fatty Acids, Unsaturated"[Mesh] OR "Arachidonic Acids"[Mesh] OR "Docosahexaenoic Acids"[Mesh] OR "LCPUFA"[TIAB] OR "PUFA"[TIAB] OR "Borage

Oil"[TIAB] OR "Fish Oil"[TIAB] OR "Arachidonic Acids"[TIAB] OR "Docosahexaenoic Acids"[TIAB] OR "fatty acid"[TIAB] OR "omega 3"[TIAB])

LCPUFA-enriched infant formula for term infants
Embase

#	Searches
1	exp Meta Analysis/
2	((meta adj analy$) or metaanalys$).tw.
3	(systematic adj (review$1 or overview$1)).tw.
4	or/1-3
5	cancerlit.ab.
6	cochrane.ab.
7	embase.ab.
8	(psychlit or psyclit).ab.
9	(psychinfo or psycinfo).ab.
10	(cinahl or cinhal).ab.
11	science citation index.ab.
12	bids.ab.
13	or/5-12
14	reference lists.ab.
15	bibliograph$.ab.
16	hand-search$.ab.
17	manual search$.ab.
18	relevant journals.ab.
19	or/14-18
20	data extraction.ab.
21	selection criteria.ab.
22	20 or 21
23	review.pt.
24	22 and 23
25	letter.pt.
26	editorial.pt.
27	animal/
28	human/
29	27 not (27 and 28)
30	or/25-26,29
31	4 or 13 or 19 or 24
32	31 not 30
33	infant nutrition.tw. or exp infant nutrition/ or nutritional support.tw. or exp nutritional support/ or infant$food.tw. or infant$diet.tw. or baby$food.tw. or baby$diet.tw. or exp baby food/ or exp artificial milk/
34	exp cognition assessment/ or Cognition.tw. or exp cognition/ or exp social cognition/ or IQ.tw. or intelligence.tw. or exp Wechsler adult intelligence scale/ or exp intelligence quotient/ or exp Wechsler intelligence scale/ or exp "Wechsler preschool and primary scale of intelligence"/ or exp intelligence/ or exp Wechsler intelligence scale for children/ or exp intelligence test/ or exp emotional intelligence/ or Intelligence.mp. or exp Stanford-Binet Intelligence Scale/ or (neurodevelopment or cognitive).tw. or child$development.tw. or exp child development/ or exp neuropsychological test/ or exp child development/ or exp "Bayley Scales of Infant Development"/ or exp mental development/ or bayley.tw. or exp reading/ or exp

	school/ or exp learning disorder/ or exp educational status/ or exp education/ or exp achievement/ or exp academic achievement/ or school attainment.tw.
35	(child$ or infant$ or baby or babies or adolescent$ or teenager$).tw.
36	exp fatty acids, omega-3/ or fatty acids, essential/ or Dietary Fats, Unsaturated/ or linolenic acids/ or exp fish oils/ or (n 3 fatty acid$ or omega 3).tw. or docosahexa?noic.tw,hw,rw. or eicosapenta?noic.tw,hw,rw. or alpha linolenic.tw,hw,rw. or (linolenate or cervonic or timnodonic).tw,hw,rw. or menhaden oil$.tw,hw,rw. or (mediterranean adj diet$).tw. or ((flax or flaxseed or flax seed or linseed or rape seed or rapeseed or canola or soy or soybean or walnut or mustard seed) adj2 oil$).tw. or (walnut$ or butternut$ or soybean$ or pumpkin seed$).tw. or (fish adj2 oil$).tw. or (cod liver oil$ or marine oil$ or marine fat$).tw. or (salmon or mackerel or herring or tuna or halibut or seal or seaweed or anchov$).tw. or (fish consumption or fish intake or (fish adj2 diet$)).tw. or diet$ fatty acid$.tw. or borage oil$.tw.
37	32 and 33 and 34 and 35 and 36

MEDLINE terms for cognitive ability, study population, systematic review and infant nutrition (see above) plus:

("Fatty Acids, Unsaturated"[Mesh] OR "Arachidonic Acids"[Mesh] OR "Docosahexaenoic Acids"[Mesh] OR "LCPUFA"[TIAB] OR "PUFA"[TIAB] OR "Borage Oil"[TIAB] OR "Fish Oil"[TIAB] OR "Arachidonic Acids"[TIAB] OR "Docosahexaenoic Acids"[TIAB] OR "fatty acid"[TIAB] OR "omega 3"[TIAB])

Iron-fortified infant formula

Embase

#	Searches
1	exp Meta Analysis/
2	((meta adj analy$) or metaanalys$).tw.
3	(systematic adj (review$1 or overview$1)).tw.
4	or/1-3
5	cancerlit.ab.
6	cochrane.ab.
7	embase.ab.
8	(psychlit or psyclit).ab.
9	(psychinfo or psycinfo).ab.
10	(cinahl or cinhal).ab.
11	science citation index.ab.
12	bids.ab.
13	or/5-12
14	reference lists.ab.
15	bibliograph$.ab.
16	hand-search$.ab.
17	manual search$.ab.
18	relevant journals.ab.
19	or/14-18
20	data extraction.ab.
21	selection criteria.ab.
22	20 or 21
23	review.pt.
24	22 and 23

25	letter.pt.
26	editorial.pt.
27	animal/
28	human/
29	27 not (27 and 28)
30	or/25-26,29
31	4 or 13 or 19 or 24
32	31 not 30
33	infant nutrition.tw. or exp infant nutrition/ or nutritional support.tw. or exp nutritional support/ or infant$food.tw. or infant$diet.tw. or baby$food.tw. or baby$diet.tw. or exp baby food/ or exp artificial milk/
34	exp cognition assessment/ or Cognition.tw. or exp cognition/ or exp social cognition/ or IQ.tw. or intelligence.tw. or exp Wechsler adult intelligence scale/ or exp intelligence quotient/ or exp Wechsler intelligence scale/ or exp "Wechsler preschool and primary scale of intelligence"/ or exp intelligence/ or exp Wechsler intelligence scale for children/ or exp intelligence test/ or exp emotional intelligence/ or Intelligence.mp. or exp Stanford-Binet Intelligence Scale/ or (neurodevelopment or cognitive).tw. or child$development.tw. or exp child development/ or exp neuropsychological test/ or exp child development/ or exp "Bayley Scales of Infant Development"/ or exp mental development/ or bayley.tw. or exp reading/ or exp school/ or exp learning disorder/ or exp educational status/ or exp education/ or exp achievement/ or exp academic achievement/ or school attainment.tw.
35	(child$ or infant$ or baby or babies or adolescent$ or teenager$).tw.
36	exp iron metabolism/ or exp iron/ or exp iron blood level/ or exp iron deficiency/ or exp iron derivative/ or exp iron deficiency anemia/ or iron complex/ or exp iron intake/ or exp iron depletion/ or iron.tw. or ferritin$.tw. or exp ferritin blood level/ or exp ferritin/ or Hemoglobin$.tw.
37	32 and 33 and 34 and 35 and 36

MEDLINE

MEDLINE terms for cognitive ability, systematic review and infant nutrition (see above) plus:

("iron"[MeSH Terms] OR "Anemia, Iron-Deficiency"[Mesh] OR "Ferritins"[Mesh] OR "Iron, Dietary"[Mesh] OR "Hemoglobins"[Mesh] OR "iron"[TIAB] OR "Iron*Deficiency"[TIAB] OR "Ferritins"[TIAB] OR "Hemoglobins"[TIAB])

F. Subgroup analyses

In the nutritional literature much focus is placed on who benefits the most from nutritional interventions. I did not present these results in the main body of the thesis because the trials were generally underpowered to detect interaction effects. Here, I present the effects stratified by group and by birth weight so that these estimates may be used in combination with external estimates to undertake meta-analyses on the effect of these interventions in the future.

Table A15 Effect on primary outcome in boys vs girls

Trial	Intervention vs standard in boys	Intervention vs standard in girls
	Estimate (95% CI)	Estimate (95% CI)
NEP-PD	-0.38 (-0.72, -0.04)	0.37 (0.05, 0.69)
NETSGA	-0.04 (-0.35, 0.27)	-0.15 (-0.45, 0.14)
LCPUFAP	0.09 (-0.30, 0.49)	-0.41 (-0.77, -0.04)
LCPUFAT	-0.16 (-0.43, 0.11)	-0.05 (-0.38, 0.27)
NUCLEO	0.62 (-0.15, 1.40)	1.35 (0.42, 2.27)
IRON	-0.15 (-0.41, 0.12)	-0.12 (-0.42, 0.18)
PALM	-0.18 (-0.54, 0.18)	0.12 (-0.26, 0.50)

Table A16 Effect on primary outcome by maternal smoking status during pregnancy

Trial	Mother did not smoke	Mother smoked
	Estimate (95% CI)	Estimate (95% CI)
NEP-PD	0.09 (-0.23, 0.40)	-0.15 (-0.54, 0.24)
NETSGA	-0.03 (-0.31, 0.26)	-0.19 (-0.52, 0.14)
LCPUFAP	-0.14 (-0.52, -0.23)	-0.11 (-0.48, 0.25)
LCPUFAT	-0.01 (-0.26, 0.24)	-0.36 (-0.76, 0.04)
NUCLEO	1.11 (0.42, 2.27)	0.44 (-0.67, 1.55)
IRON	-0.17 (-0.39, 0.05)	-0.06 (-0.50, 0.39)
PALM	0.06 (-0.38, 0.26)	-0.07 (-0.51, 0.38)

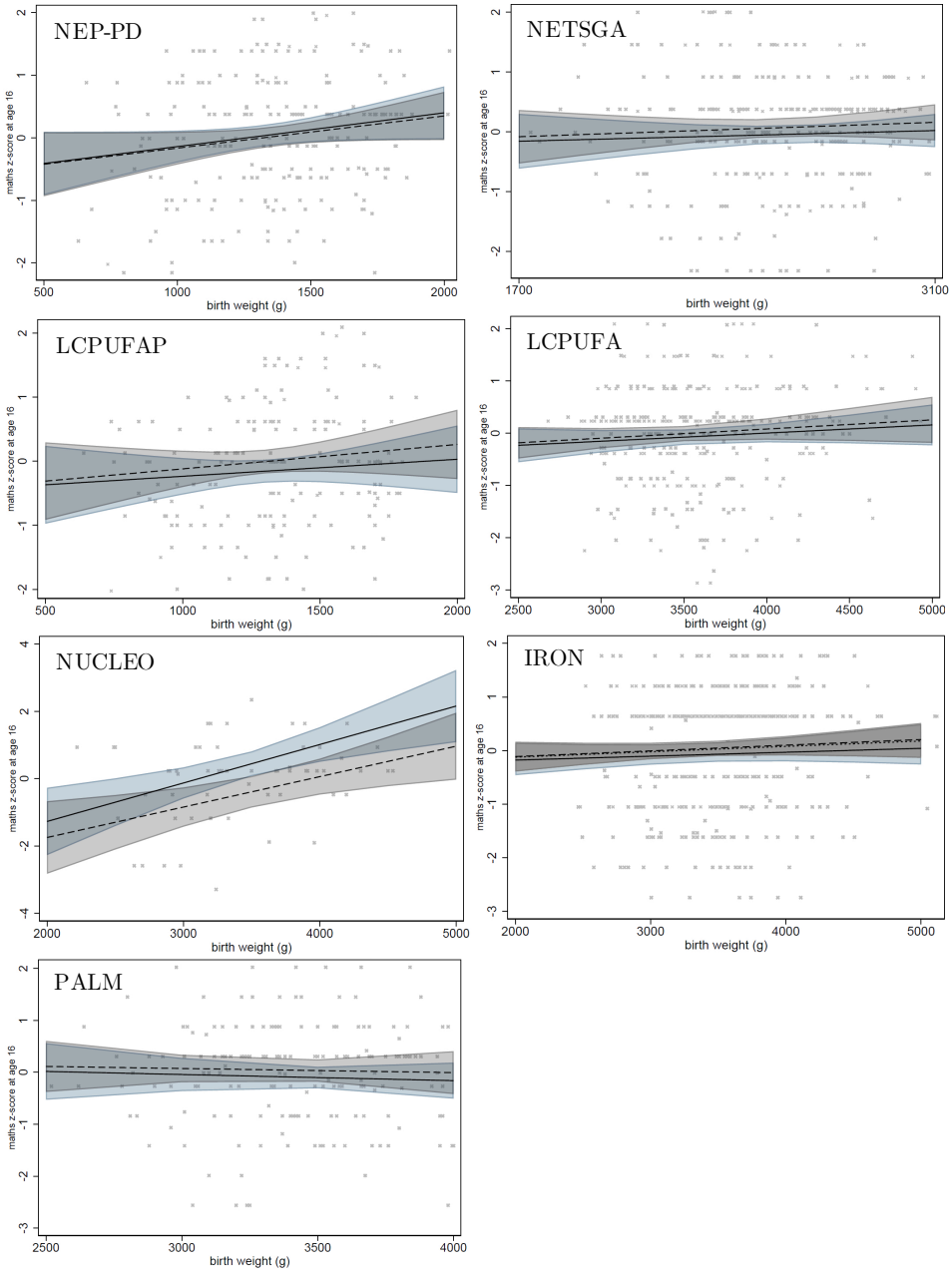


Fig. A1: Effect by birth weight

G. Performance of participants relative to the national average

Estimation of national GCSE Maths average grade and SD:

14-Nov-20	*Purpose:*	To calculate national standard SDs All years were combined in the blue column												
Data from		https://www.gov.uk/government/collections/statistics-gcses-key-stage-4 Table S5: GCSE Full Course results of pupils at the end of key stage 4 in all schools1, by subject and grade2												
			A*	A	B	C	D	E	F	G	U[4]	**N**		
		Points	8	7	6	5	4	3	2	1	0	**N**		mean grade is:
	Mathematics	Table provided these nrs in 1000s	164.3	313.6	448.5	735.5	344.2	205.5	152.6	87.8	44.2	**2,496,200.0**		$\bar{x}$
		N	164,300	313,600	448,500	735,500	344,200	205,500	152,600	87,800	44,200	**2,496,200**		4.913
		%	6.58%	12.56%	17.97%	29.46%	13.79%	8.23%	6.11%	3.52%	1.77%	**100.00%**		
		value minus mean	3.09	2.09	1.09	0.09	-0.91	-1.91	-2.91	-3.91	-4.91			
		value minus mean squared	9.53	4.35	1.18	0.01	0.83	3.66	8.49	15.31	24.14			
	$\sum \lvert x-\bar{x}\rvert^2$	sum of value minus mean squared	8212205.208											
		sum of value minus mean squared / N	3.29											
		SD	1.8138											

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

Performance of trial participants in the intervention and control groups, was compared to the English national average in the 2008/09 to 2011/12 academic years. Participants in all but two trials (IRON both groups and PALM control) were estimated to have performed significantly worse at the 5% level compared to the national average. In this estimation, participants in the LCPUFAP trial were furthest from the national average and participants in the IRON trial were closest to the national average.

Table A17 Performance in GCSE Maths relative to the combined national average 2008/09-11/12

	Modified			Standard		
Maths SD-score age 16 years standardised by national average	Mean	SE	p	Mean	SE	p
NEP-PD: MI adjusted (N_I=113, N_C=116)	-0.35	0.10	0.001	-0.38	0.09	<0.001
NETSGA: MI adjusted (N_I=152, N_C=147)	-0.40	0.09	<0.001	-0.29	0.09	0.001
LCPUFAP: MI adjusted (N_I=92, N_C=100)	-0.69	0.12	<0.001	-0.49	0.11	<0.001
LCPUFAT: MI adjusted (N_I=155, N_C=154)	-0.29	0.08	<0.001	-0.17	0.08	0.027
IRON: MI adjusted (N_I=162, N_C=165)	-0.12	0.08	0.106	-0.01	0.07	0.929
PALM: MI adjusted (N_I=103, N_C=100)	-0.27	0.10	0.010	-0.18	0.10	0.062

These observed differences in school performance could be explained by parental or infant attributes: the trial participants constitute a selected group because their carers had already opted to feed formula in the trial period, and some trials included children vulnerable to difficulties with cognitive development (LCPUFAP, NEP-PD, NETSGA). The observation that intervention groups were overall further away from the national average than control groups suggests that none of the interventions succeeded in moving cognitive outcomes closer to that of average healthy children.

H. Additional data on the NUCLEO trial

Nucleotides have been added to infant formulas on the basis that they are present in human milk and that as structural components of ribonucleic acid (RNA) and DNA, they are involved in the synthesis of proteins, lipids and carbohydrates, as well as in the transfer of energy. Previous research found no evidence that the addition of nucleotides to infant formula translates into any functional benefits for immune function, gastrointestinal function or growth. Three RCTs measured head growth, two of which found increased head growth in children fed formula with added nucleotides. Increased head growth has been associated with higher cognitive function later in life, but no study has so far explored the direct effect on cognitive function through standardised tests or school outcomes (Chapter 6). At the age when trial participants could be linked to the National Pupil Database, the majority of participants in the NUCLEO trial were not old enough to sit the GCSE exams at age 16 years, which was my primary outcome. I therefore excluded this trial from the analysis chapter. There were no significant differences between modified and standard formula groups in the secondary outcomes. However, an analysis of the small subgroup that had recorded to GCSE exam scores showed an extreme benefit of nucleotides on GCSE grades. I consider this likely to be a chance finding given that a) higher scores for children fed formula with nucleotides are not seen for earlier measures, b) that the sample size is very small and that c) imputing GCSE outcomes for the whole sample on the basis of all observed participant characteristics significantly attenuated this benefit, suggesting selection bias. As the identifiers for this study have been retained, it would be useful to re-assess GCSE outcomes once the NUCLEO cohort has passed through the whole educational trajectory. This would clarify whether the benefits observed for the small subgroup, who received the intervention and were old enough for the GCSEs at the time of linkage, holds for the whole trial population. So far, evidence cannot be used to confirm the absence or presence of an effect.

Table A18 Characteristics of the linked sample vs the original sample in the NUCLEO trial

NUCLEO	Original sample	NPD-linked sample
N	196	176
Male	113 (58%)	102 (58%)
Birth weight, grams	3457 563)	3467 (565)
Gestational age weeks	39.3 (1.4)	39.4 (1.4)
Mother's age, years	27.1 (5.6)	27.2 (5.4)
Mother smoked during pregnancy	66 (34%)	58 (33%)
Mother has degree	16 (8%)	14 (8%)

Table A19 Characteristics of the linked sample in the NUCLEO trial, by group

NUCLEO	Modified	Standard
Linked/randomised	90/99	86/97
Birth weight (grams)	3453 (2210-4720)	3482 (2170-5360)
Gestational age (weeks)	39.5 (37-42)	39.2 (37-42)
Mother's age (years)	27 (16-38)	27 (16-40)
Infant sex		
Male	55 (61%)	47 (55%)
Female	36 (39%)	39 (45%)
Mother smoked during pregnancy		
No	65 (73%)	51 (60%)
Yes	24 (27%)	34 (40%)
Missing	1	1
Mother has degree		
No	83 (92%)	77 (92%)
Yes	7 (8%)	7 (8%)
Missing	0	2

Table A20 Effects of NUCLEO formula modification on primary outcomes

NUCLEO		
Primary outcome: within-trial standardised GCSE Maths grade	Standardised mean difference	95% CI
Primary analysis		
MI adjusted (N_P=99, N_S=97)	0.34	0.06, 0.63
Sensitivity analyses		
MI unadjusted (N_P=99, N_S=97)	0.30	0.02, 0.58
Complete-case adjusted (N_P=29, N_S=25)	0.68	0.19, 1.18
Complete-case unadjusted (N_P=29, N_S=25)	0.62	0.08, 1.15

Table A21 Effects of NUCLEO formula modification on secondary outcomes

NUCLEO N_I=99, N_C=97		
Within-trial standardised grades:	SMD	95% CI
GCSE English (age 16)	0.20	-0.04, 0.45
KS2 Maths (age 11)	0.17	-0.12, 0.46
KS2 English (age 11)	0.00	-0.28, 0.28
Other secondary outcomes:	Odds ratio	95% CI
Ever qualified for special educational needs	0.50	0.25, 1.01
5+ GCSE grades ≥C	4.15	1.52, 11.32

I. Summary of current infant formula recommendations

Table A22 Summary of current infant formula recommendations according to the European Food Safety Authority

Formula Modification	Population	Rationale for formula modification	Concentration in human milk	Range currently recommended for specific population
Nutrient enrichment after discharge	Preterm infants	To meet increased nutrient demand and facilitate healthy (cognitive) development.	Energy: 65 kcal/100 mL; Protein: 1.2-10 g/100 kcal	Energy max: 110 kcal/kg/day; Protein: 3.6-4.1 g/100 kcal for bwt <1000g and 3.2-3.6 g/100 kcal for bwt 1000-1800 g.
Nutrient enrichment	SGA term infants	Support catch-up growth (failure to catch up has previously been associated with lower cognitive ability)		
LCPUFA fortification	Terms	Because it is in breastmilk. Potential benefit on cognition and vision.	DHA: 17 mg/100kcal (Brenna 2007)	DHA: 20-50 mg/100kcal
LCPUFA fortification	Preterm infants	Because it is in breastmilk and preterm infants less able to synthesise Potential benefit on cognition and vision.		DHA: 11-27 mg/100 kcal
Iron fortification of follow-on formula	Terms	To prevent iron deficiency and iron-deficiency anaemia. Supplementation is hypothesised to lead to potential benefit on cognition.	0.35 mg/L	0.45 mg/100 kcal
Sn-2 palmitate fortification	Terms	To mimic the configuration of palmitate in breast milk and achieve softer stools	About 70% of human milk palmitic acid is esterified in the sn- 2 position of triacylglycerols	No target but minimise levels of glycerol-based process contaminants in infant formulas.

J. Manuscript 1 (Systematic Review)

Removed in deposited thesis to comply with third-party copyright.

K. Manuscript 2 (Protocol)

Removed in deposited thesis to comply with third-party copyright.

L. Manuscript 3 (Main analysis)

Removed in deposited thesis to comply with third-party copyright.

Appendix References

BOYD, A., GOLDING, J., MACLEOD, J., LAWLOR, D. A., FRASER, A., HENDERSON, J., MOLLOY, L., NESS, A., RING, S. & DAVEY SMITH, G. 2013. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol, 42, 111-27.

EFSA PANEL ON DIETETIC PRODUCTS, N. & ALLERGIES 2014. Scientific Opinion on the essential composition of infant and follow-on formulae. EFSA Journal, 12, 3760.

HAGGER-JOHNSON, G., HARRON, K., FLEMING, T., GILBERT, R., GOLDSTEIN, H., LANDY, R. & PARSLOW, R. C. 2015. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open, 5, e008118.

HORTA, B. L., LORET DE MOLA, C. & VICTORA, C. G. 2015. Breastfeeding and intelligence: a systematic review and meta-analysis. Acta Paediatr, 104, 14-9.

OFFICE FOR NATIONAL STATISTICS. 2020. Dataset: Provisional long-term international migration estimates [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/migrationstatisticsquarterlyreportprovisionallongterminternationalmigrationltimestimates [Accessed 15 Oct 2020].

WOLFE, I., MACFARLANE, A., DONKIN, A., MARMOT, M. & VINER, R. 2014. Why children die: death in infants, children, and young people in the UK—Part A. London: Royal College of Paediatrics and Child Health.