CHYLD

Children with Hypoglycaemia and their Later Development

Mid-childhood Outcomes Study

STATISTICAL ANALYSIS PLAN

Version 1.5

Date:

22/6/2020

STATISTICAL ANALYSIS PLAN APPROVAL SHEET

Study: CHYLD

Title: CHYLD Mid-childhood Outcomes Study

Author(s): Chris McKinlay, Rajesh Shah

Approved by:

Jane Harding, Principal Investigator

CONTENTS

Table of Contents

1 PREFACE

The purpose of this Statistical Analysis Plan is to provide an overview of the intended statistical analyses that will be performed on data from the CHYLD Mid-childhood Outcomes Study. This document is intended to stand alone from the protocol and adhere to the main points in the analysis summary specified in the protocol. However, it is envisaged that the Statistical Analysis Plan (SAP) can undergo revision outside of the protocol.

The following documents have been reviewed in preparation of this SAP:

- The CHYLD Mid-childhood Outcomes Study protocol (version 1 – 20/10/2016)
- CHYLD Mid-childhood Outcomes Study data collection forms
- Statistical analysis plan for the CHYLD 4.5-year follow-up (version 1 – 31/08/2015)
- Primary report of CHYLD 4.5-year follow-up[1]
- Primary report of CHYLD 2-year follow-up[2]
- Primary reports of BABIES[3,4] and Sugar Babies studies[5]
- CHYLD STROBE diagram

2 SCOPE

The scope of this SAP is intended to cover ONLY those main analyses described in the protocol. For additional research questions not detailed within this SAP and/or questions requiring further exploratory analyses please refer to additional addenda. A separate SAP will be developed to examine the effect of neonatal hypoglycaemia on measures of brain volume and structure using Magnetic Resonance Imaging (MRI).

3 INVESTIGATORS

CHYLD Study Steering Group:

- Jane Harding (Principal Investigator)
- Jane Alsweiler
- Gavin Brown
- Geoff Chase
- Deborah Harris
- Christopher McKinlay
- Ben Thompson
- Trecia Wouldes

4 STUDY OBJECTIVES

4.1 Aim

The primary aims of the CHYLD Mid-Childhood Outcomes Study are to determine at 9 to 10 years of age the effect of neonatal hypoglycaemia, including frequency, severity, and glycaemic response to treatment, on:

i. Neurocognitive function:
    a. Academic achievement (reading, mathematics)
    b. Executive function (attention, memory and inhibition)
    c. Visual-motor function (fine motor skills, visual processing, visual-motor integration)
    d. Psychosocial adaptation (social, emotional, behavioural)
ii. General health
iii. Brain structure, including volume and tractography

4.2 Hypothesis

Our hypothesis is that exposure to neonatal hypoglycaemia will be associated with worse neurocognitive function, poorer general health and/or impaired brain structure at 9 to 10 years of age, with highest risk of poor outcome in children exposed to severe, recurrent, or clinically undetected (interstitial episodes only) hypoglycaemia.

5 STUDY DESIGN

CHYLD is a prospective longitudinal cohort study of moderate to late preterm and term babies born from 32 weeks' gestation who were at risk of or required treatment for neonatal hypoglycaemia. The CHYLD Study includes children recruited to two neonatal studies: the Babies and Blood Sugars Influence on EEG Study (BABIES)[3] and the Sugar Babies Study,[5] both performed at Waikato Women's Hospital, Hamilton, New Zealand. The combined CHYLD cohort comprises 614 babies (2 babies participated in both studies). The aim of the CHYLD Study is to explore the relationship between neonatal hypoglycaemia, including the severity and frequency, and neurodevelopmental outcomes.

5.1 BABIES Study

The aim of the BABIES study was to establish the relationship between blood glucose concentrations (BGCs), alternative cerebral fuels and brain function, as measured by a cot-side continuous electroencephalography (EEG) monitor, in newborn babies at risk of hypoglycaemia.[3,4] A total of 102 babies were recruited between December 2006 and February 2009.

Babies were eligible for this study if they were admitted to the Waikato Hospital Newborn Intensive Care Unit (NICU) at ≥32 weeks' gestation and had one or more of the following risk factors for or required treatment for neonatal hypoglycaemia:

- Small babies (birth weight <10th percentile or <2500g)
- Large babies (birth weight >90th percentile or >4500g)
- Babies of diabetic mothers
- Babies who were stressed, e.g., those suffering from birth asphyxia, sepsis, haemolytic disease of the newborn, respiratory distress, congenital heart disease
- Babies not feeding well

Babies who had serious congenital malformations, terminal conditions, abnormalities of the scalp or skin lesions which prevented continuous glucose monitoring (CGM) or application of the cot-side EEG monitor, were excluded.

In addition to the standard care for the management of hypoglycaemia, study babies received additional monitoring including:

- Additional blood sampling (2-4 blood samples)
- Continuous cot-side EEG monitoring
- CGM

All babies had routine blood samples as part of their clinical care. If a baby remained hypoglycaemic, up to two additional study blood samples were taken during a period of hypoglycaemia and at least four hours apart. The study stopped, including additional monitoring, when routine testing of BGC was discontinued, usually when a baby had 3 consecutive blood glucose measurements above 2.6 mmol/L.

5.2 Sugar Babies Study

The Sugar Babies study was a two-arm parallel, randomised, placebo-controlled, double-blind trial to determine the effectiveness of treatment with 40% dextrose gel in reversing neonatal hypoglycaemia in the first 48 hours after birth. The primary hypotheses were that:

1. 40% dextrose gel is more effective than feeding alone
2. Intermittent blood glucose monitoring does not detect all episodes of low glucose concentrations

Babies were eligible for study if born at ≥35 weeks' gestation and less than 48 hours old, and identified as being at risk of hypoglycaemia using the same criteria as above. The majority of babies were recruited prior to birth. Once enrolled, if hypoglycaemia occurred (<2.6 mmol/L), babies were randomised to receive either 40% dextrose gel (BioMed NZ Limited) or a placebo vehicle gel (Carboxymethyl cellulose 2%). Randomisation was stratified by risk factor (baby of a diabetic mother, small, large or other). A total of 514 babies were recruited (242 randomised and 272 not randomised) between November 2008 and December 2010. Five babies were randomised in error, leaving 237 for analysis in the randomised group (118 in the dextrose group and 119 in the placebo group).[5,6]

The primary outcome was treatment failure, defined as a BGC of <2.6mmol/L 30 minutes after the second of two treatment attempts. Secondary outcomes included:

- The time taken to achieve an interstitial glucose concertation (IGC) above 2.6 mmol/L for >1 hour

- Incidence of recurrent hypoglycaemia (blood or interstitial glucose concentration <2.6 mmol/L) after initial successful treatment (defined as blood or interstitial glucose concentration > 2.6 mmol/L for >1 hour after initial treatment)
- Total duration of IGC < 2.6 mmol/L
- Incidence of admission to the NICU
- Frequency and total volume of formula administered in the first 48 hours
- Frequency and total volume of expressed breast milk administered in the first 48 hours
- Total dose of dextrose gel administered
- Incidence and total dose of intravenous dextrose administered in the first 48 hours
- Rate of full breast feeding at two weeks of age
- Mothers' experience of having a baby in the Sugar Babies Study

All babies were managed according to the NICU clinical guidelines at Waikato hospital. BGC was measured by heel-prick sampling at one hour of age, then before feeds two to four hourly for 12 hours. Hypoglycaemia was defined as a BGC <2.6mmol/L. In babies randomised to gel treatment, BGC was measured 30 minutes after gel administration and, if the baby remained hypoglycaemic or hypoglycaemia recurred later, treatment was repeated with another syringe from the allocated pack. Up to six doses of study gel could be given over 48 hours. Additional measures could include supplementary feeds and adding Polycose (4g/100ml) to the feed. In an unwell baby or a baby not responding to these measures, intravenous 10% dextrose (2ml/kg) was given over ten minutes followed by intravenous dextrose at 60 or 90 ml/kg/day (4-6 mg/kg/min). If the intravenous dextrose rate increased beyond 9 mg/kg/min additional investigations were initiated. Babies receiving intravenous dextrose had BGCs measured 4 hourly for 12 hours, and then less frequently as clinically indicated.

All BGCs were measured on a blood gas analyser (Radiometer, ABL800Flex, Copenhagen) using the glucose oxidase method (reading range 0.0 to 60 mmol/L, coefficient of variation 2.1%). Where possible, a continuous glucose monitor sensor (CGMS® system gold ™ Medtronic, MiniMed, Northridge,CA, USA) was inserted into the lateral aspect of the thigh as soon as possible after recruitment, and remained in place for at least 48 h and up to 7 d. The data collected from the CGMS were downloaded at the end of the study period, and were therefore not available to clinicians and could not influence the clinical management of hypoglycaemia in these babies.

5.3 Two Year Follow-up Study

The 2-year follow-up study occurred at 24 months' corrected age ± 1 month and included only children born from 35 weeks' gestation due to restrictions on funding and other logistical reasons.[2] A total of 528 eligible babies were contacted at 2 years, of whom 404 completed follow up, with 14 (3.5%) and 390 (96.5%) babies from the BABIES and Sugar Babies studies respectively. Fifty-one babies were from twin pregnancy (25 pairs of twins and 1 single baby).

Data collected at the 2 year follow-up study included standardised measures of growth, neurological function (presence of cerebral palsy, tone and reflexes), vision, visual processing (global motion perception), cognitive, language and motor development (Bayley-III), executive function, emotional and behavioural status (assessment tasks, Bayley-III parent

questionnaires, BRIEF-P), general health and family environment.

5.4 Four-and-a-half Year Follow-up Study

All children in the CHYLD Study were invited to a follow-up assessment at 4.5 years, except for 7 children whose caregiver(s) previously had indicated that they did not want to be contacted again, and 3 children known to have died. Thus, a total of 604 children were eligible for the 4.5 year study, born between May 2006 and December 2010 (Figure). Children were assessed at 54 months' corrected age (± 2 months).

The 4.5-year assessment consisted of standardised measures of general cognitive function (WPPSI-III), visual-motor integration (BBVMI, 6th Edition), social-emotional and adaptive behaviour (Child Behaviour Checklist and parent questionnaire) executive function (assessment tasks), vision and visual processing (global motion perception), auditory processing, neurological examination including gross and fine motor (MABC-2), general health and family environment.

5.5 CHYLD Mid-childhood Outcomes Study

Follow-up of CHYLD cohort at 9 to 10 years of age was planned because of the significant associations seen at 4.5 years between neonatal hypoglycaemia and executive dysfunction and impaired visual-motor integration (VMI), and the higher than average rates of neurodevelopmental impairment.[1] It was expected that executive dysfunction and impaired visual-motor integration would likely contribute to poorer academic performance in affected children.

Children were assessed at 9 to 10 years' corrected age by suitably trained assessors who were blinded to the neonatal glycaemic history of the child. This age was chosen so that children could be seen in school Years 5 and 6 (usually 9 to 10 years of age) prior to intermediate school (Years 7 and 8), when educational influences and opportunities begin to change substantially. Assessments targeted following key domains:

i. Academic achievement: Assessment Tools for Teaching and Learning (asTTle)[7] and teacher questionnaire (TQ)

ii. Executive function: Cambridge Neuropsychological Test Automated Battery (CANTAB)[8], Behavior Rating Inventory of Executive Function (BRIEF)[9]-parent questionnaire (PQ) and TQ

iii. Visual-motor function: Beery-Buktenica Test of VMI (BBVMI)-6th edition[10], Movement Assessment Battery for Children- 2nd edition (MABC-2)[11], laptop based tests of motion coherence and form coherence

iv. Psychosocial adaptation: Strength and Difficulties Questionnaire (SDQ)[12]-PQ and TQ, Autism Spectrum Quotient (ASQ)[13,14]-PQ

v. General health: Child Health Questionnaire (CHQ)[15]-PQ

vi. Family and environment: CHYLD questionnaire-PQ

vii. Structural Magnetic Resonance Imaging (MRI): various brain sequences on MRI

6 OUTCOME MEASURES

The following outcome variables measured at 9-10 years' corrected age will be evaluated:

6.1 Primary Outcome

The primary outcome of the CHYLD Mid-childhood Outcome Study is low educational achievement, defined as an e-asTTle score below or well below the normative curriculum level in Reading Comprehension/Pānui or Mathematics/Pāngarau.

6.2 Secondary Outcomes

- Academic achievement:
    - e-asTTle z-score for year and term of school in Reading Comprehension/Pānui *
    - e-asTTle z-score for year and term of school in Mathematics/Pāngarau*
    - Low reading achievement: e-asTTle score below or well below the normative curriculum level in Reading Comprehension/Pānui
    - Low mathematics achievement: e-asTTle score below or well below the normative curriculum level in Mathematics/Pāngarau
    - Teacher report (global judgement) of low academic achievement in reading, relative to peers (much worse or worse than peers) and expected curriculum level (well below or below expected level)
    - Teacher report (global judgement) of low academic achievement in mathematics relative to peers (much worse or worse than peers) and expected curriculum level (well below or below expected level)
    - Teacher report (global judgement) of low academic achievement in writing relative to peers (much worse or worse than peers) and expected curriculum level (well below or below expected level)
    - Need for additional learning support or older than expected for year level
    - Need for additional learning support or older than expected for year level or low educational achievement, defined as an e-asTTle score below or well below the normative curriculum level in Reading Comprehension/Pānui or Mathematics/Pāngarau*
- Executive function (see below for details and definitions):
    - Attention: mean AST response latency (median), switching and congruent
    - Visual memory: total PAL errors (adjusted)[#] and mean first attempt memory score
    - Working memory/planning:* mean number of OTS problems solved on first choice and mean latency (median) to the first choice; total SWM between errors (sum of trials of four, six and eight tokens)[#] and mean strategy score
    - Inhibition: mean SST reaction time, and proportion with failed stop reaction time less than go reaction time, and failed stop reaction time increases

- Visual-motor function:
    - MABC-2 fine motor subtest scale score and proportion of children at risk (≤15th percentile)* and significant (≤5th percentile) motor difficulty
    - Motion coherence threshold
    - Form coherence threshold
    - Difference between form and motion coherence thresholds
    - BBVMI-6: motor coordination, visual perception, VMI standard scores and proportion of children with VMI impairment (<85)*
- Psychosocial adaptation:
    - SDQ-P and SDQ-T Total Difficulties Scores
    - Proportion of children with borderline or abnormal Total Difficulties Scores on either SDQ questionnaire (SDQ-P ≥14 or SDQ-T ≥12)*
    - SDQ-P and SDQ-T Prosocial Behaviour Scores
    - Proportion of children with borderline or abnormal Prosocial Behaviour Scores on either SDQ questionnaire (SDQ-P or SDQ-T subscale ≤5)
    - BRIEF-PQ and BRIEF-TQ T-scores for Behavioural Regulation, Metacognition and Global Executive Composite; proportion of children with Global Executive Composite in the clinical range (>65)
    - AQ-Child total score* and proportion in the clinical range (>75)
- General health
    - CHQ physical functioning and psychosocial scaled scores

*Critical secondary outcomes for secondary analyses. #Exposure effect presented as count ratio.

6.3 Variable definitions

The following definitions relate specifically to the dependent variables required for the analyses specified above.

6.3.1 Academic achievement

Academic achievement was assessed using a standardised curriculum-based online achievement test of reading comprehension and mathematics in English or Māori, the Assessment Tools for Teaching and Learning (e-asTTle). The e-asTTle software system provides a computer-assisted item bank (calibrated with item response theory and against national norms) allowing the design and administration of standardised tests of reading comprehension and mathematics related to the New Zealand Curriculum and Te Marautanga o Aotearoa (New Zealand Māori-medium Curriculum).[7] It was developed at the University of Auckland and deployed to all schools by the Ministry of Education. It is designed for students in Years 5 to 10 (approximately age 9 to 15 years) with normative data for Years 4 to 12 inclusive, and it is aligned to New Zealand Curriculum Levels 2 to 6.

For English-medium students, a computer-adaptive version was administered, where

possible, starting with a mixture of Level 2 and Level 3 test items and related material. Test difficulty was adapted using a computer adaptive sequential test mechanism that adjusts the difficulty of the test dependent on individual success in the previous section[16]. In each of the domains of Reading Comprehension and Mathematics, two strands were tested. Strands were selected to contain objects that are related to executive function and visual-spatial perception.

Domain	Level	Strands	Time
Reading Comprehension Adaptive	2 and 3 onwards	Purposes & Audience Structure	28 min
Mathematics Adaptive	2 and 3 onwards	Position & Orientation Shape	28 min

For Māori-medium students, a fixed form e-asTTle test was used with three levels of difficulty. The difficulty level was selected based on the teacher's evaluation of the student's current achievement (Easy for those below or well below Year Standard; Medium if at Year Standard; Hard if above Year Standard).

Domain	Level	Strands	Time
Pānui	Easy: most Level 2, few Level 3 Medium: many Level 2, many Level 3 Hard: few Level 2, most Level 3	Understanding Interpretation & Evaluation	28 min
Pāngarau	Easy: most Level 2, few Level 3 Medium: many Level 2, many Level 3 Hard: few Level 2, most Level 3	Geometry Algebra	28 min

The e-asTTle tests were limited to 28 minutes per domain, followed by six general aptitude questions on interest and self-efficacy, which took an additional 2 minutes to complete. When online testing was not available, a paper version was used from an earlier desktop version of the system (asTTle V4), appropriate for the child's school year and term. The results of the paper test were entered into the desktop computer version of asTTle for calculation of scores.

The e-asTTle computer-assisted test bank of about 10,000 items was normed on a nationally representative sample of over 150,000 children in Years 4 to 12. All items were calibrated for difficulty and to each other using one-parameter item response theory. During testing, a probabilistic score is derived, representing the difficulty level at which the child would likely get 50% of items correct, which is transformed into a scale score. Scale scores are reported separately by subject, ranging from 1300 to 1700 (standard deviation [SD]=100, standard error of mean [SEM]=22 points), for each year and term of schooling. Based on normative cut-scores set using a variety of standard setting techniques, the scale score is used to define the student curriculum level.[17] Curriculum levels are reported by subject ranging from below Level 2 to above Level 6, with sub-levels of Basic, Proficient, and Advanced.[18]

In keeping with the reporting against National Standards that was in place at the commencement of the study, achievement was rated as being "At the curriculum level" if achievement was within the range of the levels seen in the normative data for students in the three terms either side of the tested child's school year and term; "Above the curriculum level" meant the student was performing equal to students enrolled in year levels four or

more terms above the tested child's school year and term; "Below the curriculum level" meant the student was performing equal to students enrolled in year levels four to seven terms below the child's school year and term; and "Well Below the curriculum" meant that the student was performing equal to students enrolled eight or more terms below the child's school year and term.

Need for additional learning support was as reported by teacher or parent. Older than expected for year level was defined as the child being one or more years older than 95% of children in their year level on 1 July of the year in which they were assessed ie being 9 years or older in year 4, 10 years or older in year 5, or 11 years or older in year 6 (Data from Ministry of Education, New Zealand, years 1996 to 2019).

6.3.2 Executive function

Cognitive elements of executive function were assessed using items from the tablet-based CANTAB system, focusing on attention (Attention Switching Task), visual memory (Paired Associate Learning), planning/working memory (One-touch Stockings of Cambridge, Spatial Working Memory,) and inhibition (Stop Signal Task).[8]

The Attention Switching Task (AST) assessed the ability of a child to switch their attention between the direction of an arrow that appears on the screen and the location of an arrow on the screen in presence of distracting information that was given during testing. The outcome measure was median response latency (switching and congruent).

The One Touch Stocking of Cambridge (OTS) task involved rearranging colored balls in the right sequence on the bottom half of the screen in two to six moves to make it look like the top half. The outcome measures were the number of problems solved on first choice and median latency to the first choice.

The Spatial Working Memory (SWM) task involved finding a blue token on the screen by eliminating boxes that also appeared on the screen. The two outcome measures were total between errors (number of times the subject revisited a box in which a token had previously been found, calculated for trials of four, six and eight tokens) and strategy score.

The Paired Associate Learning (PAL) task involved correctly locating patterns that were previously shown on the screen hidden in the boxes. The outcome measures were total errors (adjusted) and the first attempt memory score (number of correct box choices made on the first attempt during assessment problems).

The Stop Signal Task (SST) involved responding to an arrow stimulus, by touching the side of the screen where the arrow was pointing. The participant needed to inhibit the response if the audio tone was switched on. The main outcome measure was the stop-signal reaction time (length of time between the go stimulus and the stop stimulus at which the subject was able to successfully inhibit their response on 50% of the trials); two categorical outcomes were also assessed (failed stop reaction time less than go reaction time, and mean failed stop reaction time increases).

6.3.3 Visual-motor function

Fine motor skills were assessed using the Movement Assessment Battery for Children, second edition, age band 2 (MABC-2).[11] MABC-2 age band 2 tests children from 7 to 10 years for a series of fine and gross motor tasks grouped in three categories: Manual Dexterity, Aiming and Catching, and Balance. The fine motor subtest (Manual Dexterity) was used, which

included three tasks: placing pegs on a pegboard, threading a lace and a drawing trail. The outcome measure was the fine motor component standard score (normative mean 10, SD 3).

Visual processing was measured using laptop-based tests of motion coherence and form coherence, which represent global motion processing.[19] The outcome measures were motion and form coherence thresholds as percentage, and the difference between these as a measure of the differential functioning of the dorsal visual stream (motion coherence) and ventral visual stream (form coherence).

Visual-motor integration was assessed by the Beery-Buktenica Developmental Test of Visual-Motor Integration, 6th edition (BBVMI-6).[10] BBVMI-6 can be used for ages 2 to 100 years and consists of 3 tests: fine motor coordination, visual perception and visual-motor integration (VMI). The VMI task involves copying geometric drawings of increasing complexity, visual perception task involves discriminating geometric figures, and motor coordination task requires drawing figures in between lines. The outcome measures were standard scores for motor coordination, visual perception and VMI (normative mean 100, SD 15).

6.3.4 Parent questionnaires

Parents completed questionnaires about their child's emotional and behavioural development using the SDQ,[12] BRIEF,[9] and ASQ.[13,14] In addition, functional health and wellbeing was assessed using the CHQ.[15]

SDQ is a brief questionnaire for 4- to 17-year old children that assesses emotional and behavioural difficulties. The 25 items (questions) are divided into five scales, four of which (emotional symptoms, conduct problems, hyperactivity/inattention, peer problems) comprise the Total Difficulties Score (range 0-40). The fifth scale, prosocial behavior, represents the strengths of a child. Borderline and abnormal thresholds have been defined for the scales and the Total Difficulties Score. We used the Australian one-sided version for 4- to 10-year old children (without impact score).

The BRIEF questionnaire is used in children aged 5 to 18 years and assesses behavioural manifestations of executive function. It consists of 86 items in eight non-overlapping clinical scales (inhibit, shift, emotional control, initiate, working memory, plan/organize, organisation of materials, monitor). These theoretically and statistically derived scales form two indexes: a) Behavioral Regulation (three scales) and b) Metacognition (five scales), as well as a Global Executive Composite score, which takes into account all of the clinical scales and represents the child's overall executive function. There are also two validity scales to measure Negativity and Inconsistency of responses (≥7 indicates a high degree of inconsistency). T-scores are provided for scales, indices and the Global Executive Composite (normative mean 50, SD 10). Scores >65 are considered clinically significant.

ASQ-Child is a 50-item questionnaire that assesses autistic traits in children from 4 to 11 years. It has five subscales: social skills, attention switching, attention to detail, communication and imagination. If five or fewer answers are missing, these can be substituted for the mean item score. A total score >75 (range 0-150) is indicative of an autistic spectrum disorder.

CHQ assesses general health and health-related quality of life, including 14 physical and psychosocial concepts. We used the parent short form with 28 items. Outcome measures were physical functioning and psychosocial scaled scores (normative mean 50, SD 10).

6.3.5 Teacher questionnaire

The child's teacher was asked to complete a structured questionnaire about the child's year level, school term, how well they knew the child, need for additional learning support, and whether the child was being instructed in English or Māori. Teachers were asked to rate the child's academic achievement against the expected curriculum level for school year and term (well below, below, at, or above). Teachers also completed teacher versions of the SDQ and BRIEF questionnaires, which had similar items and structure to the parent forms. The teacher completing the questionnaire needed to have known the child for at least one month.

7 GENERAL STATISTICAL METHODS

7.1 Data Management

Data were collected using a combination of online tests (CANTAB), computerised assessments (visual perception, e-asTTle, MRI analysis), and hard copy questionnaires. The collected data were stored in an electronic database (REDCap®) on secure servers and access was controlled by unique user ID and password. Strict confidentiality was maintained by using study ID numbers for every child on all documents and database records. Appropriate range and logic checks were used to check for data entry errors. A data monitor reviewed all electronic case record forms (eCRF) for completeness. Once data queries were resolved, the eCRFs were locked. Hard copy questionnaires were scanned and added to electronic archive and then stored in a locked cabinet.

7.2 Analysis Population

A total of 479 out of 587 eligible participants in the CHYLD Mid-childhood Outcomes Study with outcome data will be included in the final analysis and considered as the overall analysis population (i.e., base denominator) for the primary outcome. For secondary outcomes, the denominator will be those with data available.

A small number of children have suffered events which are thought to be unrelated to neonatal hypoglycaemia, but which may possibly influence later outcomes. Their relevant clinical history has been reviewed by a medical review panel (JH, JA, CM, NB) and the medical condition has been classified as unrelated to the study outcome or potentially related. All children will be included in the primary analyses, but a sensitivity analysis will be performed excluding those children who have had a postnatal insult that has been adjudicated as unrelated to neonatal hypoglycaemia and potentially influencing the study outcome. At all times, the results will be clearly flagged with the number (and reasons) for any exclusion.

7.3 Missing Data

Multiple imputations are generally recommended on missing data if they are missing at random, which is a reasonable assumption in most statistical models. However, when the missing data are likely to be related to behavioural or developmental problems, as in this study, this approach can no longer represent the whole analysis population and may produce biased parameter estimates. Therefore, we do not propose to perform any imputation of missing outcome data in the primary analysis.

If testing was attempted but could not be completed due to underlying neurological or developmental disorder, then children were assigned a score of >3 SD below the mean.

7.4 Data Transformation

Since the sample mean of a continuous outcome variable is closely approximated by the Normal distribution in large samples (central limit theorem) it is anticipated that the CHYLD Mid-childhood Outcomes Study sample size (n=479) is sufficiently large that transformation will not be needed for continuous data, except for extreme distributions, in which case data may be categorised or transformed to normality as appropriate

7.5 Clustering

There are 35 pairs of twins in this study for whom the data may be treated as clustered within each family. Since the numbers are limited compared to the total cohort, primary analysis will not be adjusted for the clustering effect of children from multiple pregnancy.

Another potential cluster effect arises from combining the participants' data from two different studies (BABIES, Sugar Babies). While those in the same study may be more similar than those from different studies, major differences in prognostic factors between studies are likely to relate to reasons for being at risk of hypoglycaemia, which will be adjusted for in all regression analyses. Therefore, further adjustment for clustering of children in studies will not be performed.

7.6 Statistical Significance

Statistical tests will be two-sided with a significance level of 5% for the primary outcome No adjustment will be made for multiple comparisons on secondary outcomes, of which results will be interpreted with caution. The number of formal comparisons will be recorded and reported in any subsequent publications.

7.7 Potential Confounders

The following confounding variables are considered important and will be adjusted for in all regression analyses:

- New Zealand Deprivation Index (ordinal scale 1-10)[20]
- Child's sex
- Primary risk factor for neonatal hypoglycaemia, prioritised as follows: infant of a diabetic mother (IDM), preterm, small, large and other.

The NZDPI is produced by Statistics New Zealand and combines nine socioeconomic variables from the national census into a decile scale from 1 to 10, with 10 indicating the highest deprivation level. The NZDPI is available for small geographic clusters of approximately 80 households. NZPDI will be calculated using the child's primary residential address at follow-up.

7.8 Descriptive Summary of Study Cohort

Descriptive information will be provided on all study data collected from the total cohort (N=479), at birth and 9-10 years' corrected age.

Continuous variables will be summarised as numbers of observed and missing values, and mean (standard deviation), or median (inter-quartile range). Categorical variables will be described as frequencies and percentages.

The neonatal and maternal characteristics of the CHYLD Mid-childhood Outcomes Study participants (N=479) will be compared against those who were eligible for follow-up but who did not participate (N=108). This is to provide evidence of external validity of this follow-up cohort. Participants and non-participants will be compared by Chi-square or Fisher's exact test for categorical variables; and two-sample t-test or Wilcoxon rank-sum test for continuous variables, as appropriate.

7.9 Continuous Glucose Monitoring

CGM data were previously downloaded with CGMS Solutions software (version 3.0C) and recalibrated to BGCs to optimise accuracy at low concentrations.[21] Recalibrated traces commence at the first blood glucose calibration. Recalibrated data will be used in all analyses.

7.10 Exposure Definitions

The following neonatal glycaemic exposures in the first week after birth were previously defined for the CHYLD Study:[2]

- Hypoglycaemic episode: ≥1 consecutive BGC <2.6 mmol/L (mild ≥2.0 mmol/L to <2.6 mmol/L; severe <2.0 mmol/L)
- Interstitial episode: IGC <2.6 mmol/L for ≥10 minutes
- Hypoglycaemic events: sum of nonconcurrent hypoglycaemic and interstitial episodes more than 20 minutes apart
- Minimum, maximum and mean BGC and IGC in the first 12 and 48 h
- Percentage of BGC outside the central band of 3 to 4 mmol/L in the first 48 h
- Percentage of time with IGC outside the central band of 3 to 4 mmol/L in the first 48 h
- Negative interstitial increment: area above the interstitial glucose concentration curve and below a given threshold

8 Primary Analysis

In the primary analysis, children with and without hypoglycaemia, defined as ≥1 hypoglycaemic event in the first week after birth, will be compared for the prespecified primary and secondary outcomes using generalised linear models with binomial, poisson or normal distributions, and the relevant link function as appropriate, adjusted for the predefined potential confounders of New Zealand Deprivation Index (NZDPI),[20] sex, and primary risk factor for neonatal hypoglycaemia. In the event that models fail to converge because there are too few children per outcome in a particular primary risk factor for hypoglycaemia group, risk factors maybe collapsed into a larger "other" category. Exposure effects will be presented as adjusted risk ratio (RR), mean difference (MD) or count ratio (CR), as appropriate, with 95% confidence interval (CI). Secondarily to adjusted risk ratios, odds ratios (OR), with 95% confidence interval and appropriate measures of effect size will be

tabulated for each outcome.

9 Secondary Analyses

The following pre-specified secondary analyses will be performed for the primary and critical secondary outcomes in the entire cohort, and then excluding children who a) experienced a postnatal insult that has been adjudicated as unrelated to neonatal hypoglycaemia and potentially influencing the study outcome OR were not educated using New Zealand national curricula; b) did not have any hypoglycaemic episodes and did not have neonatal CGM to potentially detect low IGC, and c) were born at <35 weeks' gestation

- Comparison of children with and without hypoglycaemic episodes in the first week after birth
- Comparison of children with no hypoglycaemic events to those with different severities of hypoglycaemia: a) mild hypoglycaemia, defined as 1 or 2 hypoglycaemic events ≥2.0 to 2.5 mmol/L; b) severe or recurrent hypoglycaemia, defined as any hypoglycaemic event <2.0 mmol/L or ≥3 hypoglycaemic events
- Comparison of children with no hypoglycaemic events to those whose hypoglycaemia was or was not detected (and therefore did or did not provide opportunity for treatment): a) detected hypoglycaemia, defined as ≥1 hypoglycaemic episodes; b) undetected hypoglycaemia, defined as ≥1 hypoglycaemic events but no hypoglycaemic episodes
- Comparison of children with no hypoglycaemic events to those with different frequencies of hypoglycaemia: 1-2 events; or ≥3 events
- Comparison of children with no hypoglycaemic events to those with different degrees of hypoglycaemia: mild hypoglycaemic events only; or ≥1 severe event.

10 Exploratory Analyses

It is possible that children with overall poorer educational achievement may start school later or progress more slowly through year levels and thus be older than average for their year level, but perform appropriately for that year level. In order to explore whether hypoglycaemia is related to delayed school progression, we will visualise the relationship between asTTle z scores and corrected age at assessment to evaluate whether this relationship differs in children with and without hypoglycaemia (any event). Generalised linear models will be fitted to each of the asTTle z scores with an interaction term between corrected age at assessement and hypoglycaemia. The Post-fitting for Linear Models (PLM) procedure of SAS will be used to visualise this relationship plotting first and higher order associations with 95% confidence bands. This analysis will be performed on the entire cohort, separately for Māori and English language medium assessments, and then stratified by primary risk factor for hypoglycaemia.

11 Study Power

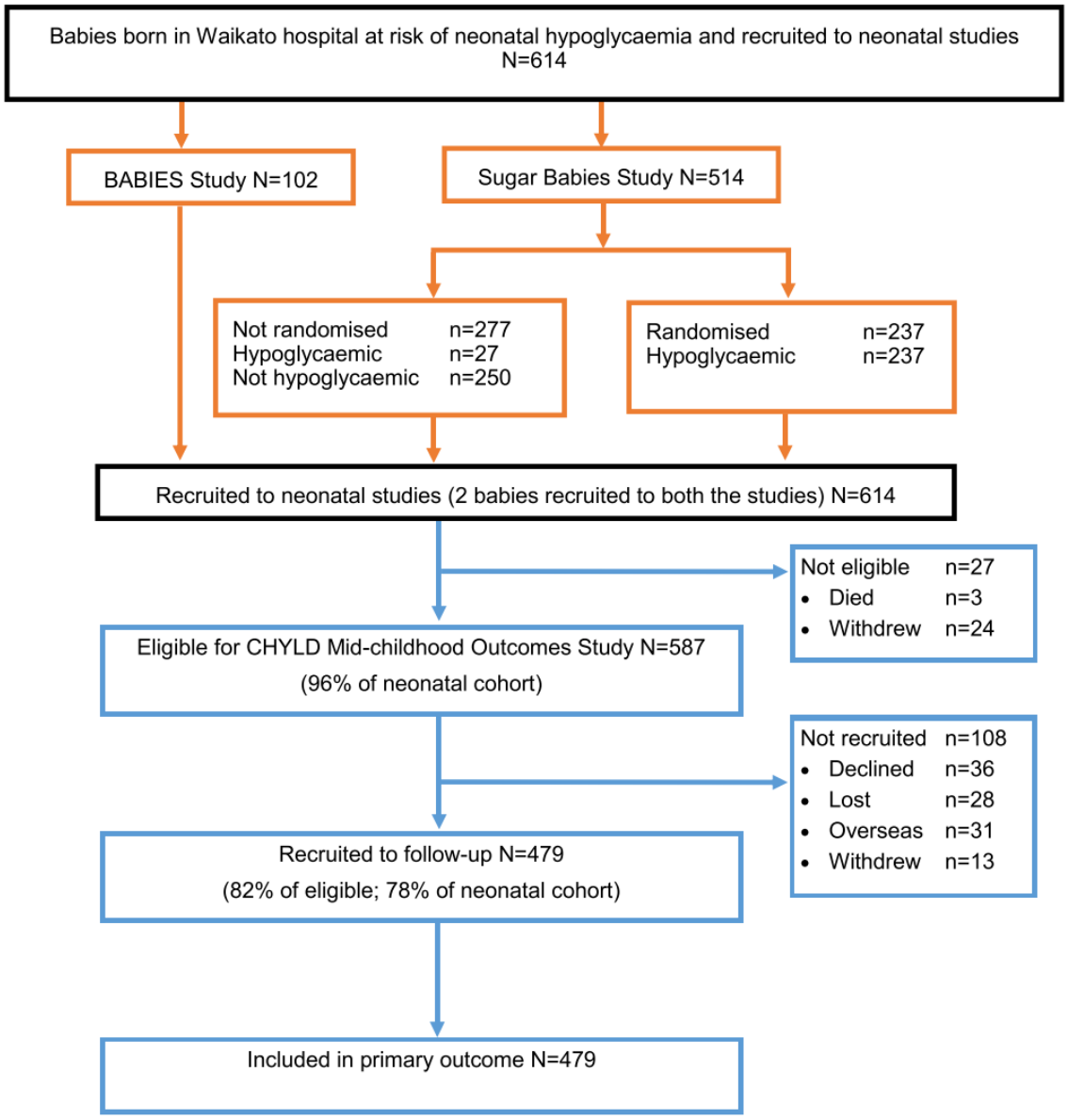
Based on previous data from this cohort we estimate that the incidence of the primary outcome in controls would be 20%. Assuming 80% follow-up we estimate that the study has 90% power to detect an increase in incidence of the primary outcome from 20% in controls to 33% (RR 1.60) in children exposed to hypoglycaemia (two tailed α=0.05, assuming half were exposed to hypoglycaemia). For continuous outcomes, we estimate the study would have 90% power to detect differences of 0.3 of a standard deviation (SD).

12 References

1. McKinlay CD, Alsweiler JM, Anstice NS, et al. Association of neonatal glycemia with neurodevelopmental outcomes at 4.5 years. JAMA Pediatr. 2017;171(10):972-83.
2. McKinlay CD, Alsweiler JM, Ansell JM, et al. Neonatal glycemia and neurodevelopmental outcomes at 2 years. N Engl J Med. 2015;373(16):1507-18.
3. Harris DL, Battin MR, Weston PJ, et al. Continuous glucose monitoring in newborn babies at risk of hypoglycemia. J Pediatr. 2010;157(2):198-202. e1.
4. Harris DL, Weston PJ, Williams CE, et al. Cot-side electroencephalography monitoring is not clinically useful in the detection of mild neonatal hypoglycemia. J Pediatr. 2011;159(5):755-60. e1.
5. Harris DL, Weston PJ, Signal M, et al. Dextrose gel for neonatal hypoglycaemia (the Sugar Babies Study): a randomised, double-blind, placebo-controlled trial. Lancet. 2013;382(9910):2077-83.
6. Harris DL, Weston PJ, Harding JE. Incidence of neonatal hypoglycemia in babies identified as at risk. J Pediatr. 2012;161(5):787-91.
7. Hattie JA, Brown GT, Keegan PJ. A national teacher-managed, curriculum-based assessment system. Int J Learn. 2003;10.
8. Luciana M. Practitioner review: computerized assessment of neuropsychological function in children: clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). J Child Psychol Psychiatry. 2003;44(5):649-63.
9. Gioia G, Isquith P, Guy S, et al. BRIEF—Behavior Rating Inventory of Executive Function: professional manual: . Odessa, FL: Psychological Assessment Resources; 2000.
10.Beery KE, Beery NA. The Beery-Buktenica Developmental Test of Visual-Motor Integration. 6th ed. Bloomington: MN: Pearson; 2010.
11.Henderson SE, Sugden DA, Barnett AL. Movement Assessment Battery for Children-2 (Second Edition). London, UK: The Psychological Corporation; 2007.
12.Goodman R. The Strengths and Difficulties Questionnaire. J Child Psychol Psychiatry. 1997;38(5):581-6.
13.Auyeung B, Baron-Cohen S, Wheelwright S, et al. The Autism Spectrum Quotient: Children's version (AQ-Child). J Autism Dev Disord. 2008;38(7):1230-40.
14.Zhang L, Sun Y, Chen F, et al. Psychometric properties of the Autism-Spectrum Quotient in both clinical and non-clinical samples: Chinese version for mainland China. BMC Psychiatry. 2016;16(1):213.
15.Waters E, Salmon L, Wake M. The parent-form Child Health Questionnaire in Australia: comparison of reliability, validity, structure, and norms. J Pediatr Psycho. 2000;25(6):381-91.
16.Luecht R, Brumfield T, Breithaupt K. A testlet assembly design for adaptive multistage tests. Appl Meas Educ. 2006;19(3):189-202.
17.Hattie J, Brown GT, Keegan PJ, et al. Assessment Tools for Teaching and Learning (asTTle) manual (Version 4, 2005). Wellington: University of Auckland/Ministry of Education/Learning Media; 2004.
18.Education Service Desk. Measurement error in e-asTTle, e-asTTle Help: Reports. Wellington: Ministry of Education2020 [cited 2019 October 9]; Available from: https://easttlehelp.vln.school.nz/category/reports.
19.Chakraborty A, Anstice NS, Jacobs RJ, et al. Global motion perception is independent from contrast sensitivity for coherent motion direction discrimination and visual acuity in 4.5-year-old children. Vision Res. 2015;115:83-91.
20.Atkinson J, Salmond C, Crampton P. NZDep2013 Index of Deprivation. In: Health DoP, editor. Wellington: University of Otago; 2014.
21.Signal M, Le Compte A, Harris DL, et al. Impact of retrospective calibration algorithms on hypoglycemia detection in newborn infants using continuous glucose monitoring. Diabetes Technol Ther. 2012;14(10):883-90.

13 Appendices

13.1 CHYLD STROBE diagram



13.2 Table of characteristics of participants and non-participants of the CHYLD Mid-childhood Outcome Study

Characteristics	Total eligible for follow-up N=587	Participants N=479	Non-participants N=108	P
Maternal				
Age at entry—year				
BMI in early pregnancy—kg/m^2				
Diabetes in pregnancy Gestational Pre-gestational				
Smoking in pregnancy				
Alcohol use in pregnancy				
Highest education level Schooling incomplete High school ≥3 y Technical or trade University				
Neonatal				
Primary risk factor for hypoglycaemia IDM Preterm Small Large Other				
Sex—female				
Twins				
Gestational length—week				
Birthweight—g				
Birthweight z-score				
Apgar <7 at 5 min				
Admitted to NICU				
Prioritised ethnicity at entry Māori Pacific Other NZ European				
Socio-economic deprivation at entry Most deprived Less deprived				
Neonatal glycaemia				
CGM				

No hypoglycaemia				
Mild hypoglycaemia				
Severe or recurrent hypoglycaemia				
Undetected hypoglycaemia				
CHYLD follow-up				
Enrolled in BABIES				
Enrolled in Sugar Babies Study				
Randomised in Sugar Babies Study				
Seen at 2 years				
Seen at 4.5 years				

Data are number (%) and mean (SD). BMI, Body mass index; CGM, Continuous glucose monitoring; IDM, Infant of diabetic mother; NICU, Neonatal intensive care unit; NZ, New Zealand. P value is for comparison between participants and non-participants. Most deprived defined as NZDPI 8-10. Mild, severe, and undetected hypoglycaemia are as defined above (section 9).

13.3 Table of characteristics of CHYLD Mid-childhood Outcome Study participants with or without hypoglycaemia

Characteristics	Total N=479	N	Hypoglycaemia* N=	N	No Hypoglycaemia N=	N	P
Maternal							
Age at entry—year							
BMI in early pregnancy—kg/m^2							
Diabetes in pregnancy Gestational Pre-gestational							
Smoking in pregnancy							
Alcohol use in pregnancy							
Highest education level Schooling incomplete High school ≥3 y Technical or trade University							
Neonatal							
Primary risk factor for hypoglycaemia IDM Preterm Small Large Other							
Sex—female							
Twins							
Gestational length—week							
Birthweight—g							
Birthweight z-score							
Apgar <7 at 5 min							
Admitted to NICU							
Prioritised ethnicity at entry Māori Pacific Other NZ European							
Socio-economic deprivation at entry							

Most deprived Less deprived							
BGC mean, minimum, maximum—mmol/L <12 h 12 to <24 h 24 to <48 h ≥48 h							
Percentage of BGC outside central band of 3-4 mmol/L in first 48 h							
CGM							
IGC mean, minimum, maximum—mmol/L <12 h 12 to <24 h 24 to <48 h ≥48 h							
Percentage of time IGC outside central band of 3-4 mmol/L in first 48 h							
Feeding in the first week No enteral feeds Breast milk only Formula milk only Breast and formula milk							
Intravenous dextrose							
Mid-childhood							
Age—years							
Prioritised ethnicity Māori Pacific Other NZ European							
Main language spoken at home is English							
Te reo Māori spoken at home							
Socio-economic deprivation Most deprived Less deprived							
Inadequate family resources**							
Year at school							

Year 4 Year 5 Year 6							
Te Marautanga o Aotearoa (Māori-medium Curriculum)							
Geographical location Northland region Auckland region Waikato region Eastern North Island Central/Lower North Island South Island							
Location of assessment School Home Liggins Institute Other							

Data are number (%) and mean (SD). BGC, blood glucose concentration; IDM, Infant of diabetic mother; IGC, interstitial glucose concentration; NICU, Neonatal intensive care unit; NZ, New Zealand. *Hypoglycaemia defined as ≥1 hypoglycaemic event. Most deprived defined as NZDPI 8-10. **Family resources score <34. P value is for comparison between those with hypoglycaemia and no hypoglycaemia. Mild, severe or recurrent, and undetected hypoglycaemia are as defined above (section 9).