SYMBA Study
STATISTICAL ANALYSIS PLAN

Study Title: Dietary modulation of maternal gut flora with oligosaccharides in pregnancy as a novel allergy prevention strategy: a randomised controlled trial [SYMBA Study]

Trial registration	Australian and New Zealand Clinical Trials Registry ACTRN12615001075572
SAP version	1.0, 07/07/2023
SAP revision history	-
Funding	National Health & Medical Research Council (NHMRC) (Project Grant ID1099480) and Telethon-Perth Children's Hospital Research Fund
SAP Authors	Dr Alana Cuthbert Trial Statistician SAHMRI Women & Kids, South Australian Health & Medical Research Institute Dr Debra Palmer Telethon Kids Institute Principal Investigator

	Signature	SAP version	Date
Alana Cuthbert		1.0	13/07/2023
Debra Palmer		1.0	11/07/2023

Table of Contents

1 PREFACE

This Statistical Analysis Plan (SAP) describes the planned analyses and reporting for the SYMBA randomised controlled trial. The following documents were reviewed in preparation of this SAP:

- Protocol publication: DJ Palmer, J Keelan, J Garssen, K Simmer, MC Jenmalm, R Srinivasjois, D Silva, SL Prescott. Study protocol for a randomised controlled trial investigating the effects of maternal prebiotic fibre dietary supplementation from mid-pregnancy to six months' post-partum on child allergic disease outcomes. Nutrients 2022, 14(13), 2753. https://doi.org/10.3390/nu14132753
- SYMBA Study case report form (CRF) (version 1, 6th January 2016);
- SYMBA Study protocol (version 1, 24th February 2016).

Any deviations from the planned analyses detailed in this SAP will be clearly documented with reasons in a post-analysis version of the SAP. Any post-hoc analyses which are not identified in this SAP but are completed to support planned study analyses will also be clearly identified.

2 INTRODUCTION

2.1 Background and rationale

Infant allergy is the most common early manifestation of an increasing propensity for inflammation and immune dysregulation in modern environments. Refined low-fibre diets are a major risk for inflammatory diseases through adverse effects on the composition and function of gut microbiota. This has focused attention on the potential of prebiotic dietary fibres to favourably change gut microbiota, for local and systemic anti-inflammatory effects. In pregnancy, the immunomodulatory effects of prebiotics may also have benefits for the developing foetal immune system and provide a potential dietary strategy to reduce the risk of allergic disease. Hence this randomised controlled trial will investigate the effects of maternal prebiotics supplementation on child allergic disease outcomes.

2.2 Objectives

To determine whether the incidence of eczema in infancy can be reduced by maternal prebiotics supplementation during pregnancy and lactation.

3 STUDY METHODS

3.1 Trial design

This is a single-centre, parallel, two-arm (1:1 allocation), double-blinded, superiority, randomised controlled trial. Pregnant women whose infants have a first-degree relative (mother, father, or sibling) with a history of medically diagnosed allergic disease (asthma, allergic rhinitis, eczema and/or food allergy) will be randomised to either a prebiotics intervention group or maltodextrin placebo control group. Women will be recruited at 18-21 weeks' gestation and are instructed to consume study powder from the time of randomisation until 6 months' postnatal infant age. The primary outcome is infant medically diagnosed eczema by 1 year of age.

3.2 Randomisation

Pregnant women were assigned to a prebiotics intervention group or maltodextrin placebo control group using a secure web-based randomisation service. The randomisation service allocated group assignments according to a computer-generated randomisation schedule, produced by an independent statistician not involved with trial participants or data analysis. Randomisation was stratified by maternal allergy status (yes/no - history of medically diagnosed asthma, allergic rhinitis, eczema and/or food allergy) and maternal pre-pregnancy body mass index (BMI < 25 or ≥25) using randomly permuted, size-balanced blocks.

3.3 Sample size

To detect a reduction in infant medically diagnosed eczema from 30% in the control group to 20% in the treatment group (absolute reduction 10%, relative reduction 33%) with 80% power and two-sided alpha of 0.05, 293 women per group were required. Allowing for a 10% dropout, a total sample size of 652 women (326 per group) was the sample size target aim for this trial.

3.4 Framework

All comparisons will be undertaken assuming a standard superiority hypothesis testing framework.

3.5 Timing of statistical analysis and unblinding

The database will be locked for analysis once data collection and cleaning are complete and the final version of this SAP has been approved. Following the database lock, blinded treatment codes will be made available to the trial statistician and analysis of the listed outcomes will be performed blinded to treatment group. Results of these analyses will be made available to the Trial Steering Committee members, with the blinding broken following a review of results.

3.6 Timing of outcome assessments

The 1 year of age clinical outcome assessment should occur when infants are between 12 and 18 months of age. Outcome data collected outside this window will still be included in the main intention to treat analyses but excluded from a per-protocol analysis of the primary outcome (see Section 4.3).

4 STATISTICAL PRINCIPLES

4.1 Confidence intervals and p values

For each outcome variable, a 95% confidence interval will be reported to express uncertainty about the estimated treatment effect. The statistical significance of the estimated treatment effect will be assessed at the 0.05 level using a two-sided comparative test, unless otherwise specified.

In describing the effectiveness of the intervention, multiple hypothesis tests will be performed due to multiple secondary outcomes and per-protocol analyses. No multiplicity adjustment will be made for the number of secondary and per-protocol analyses, as these are of less importance than the overall intention to treat (ITT) analysis of the primary outcome. In the absence of a formal procedure for controlling the type-I error rate, less emphasis will be placed on the results of secondary/per-protocol analyses.

4.2 Adherence and protocol deviations

Adherence to the intervention involves:

1. maternal consumption of the allocated study powder, and
2. maternal avoidance of non-study powder prebiotic drinks or supplements.

Participants are asked to return unused study powder at the end of the intervention period. All unused study powder will be weighed in order to record compliance to study powder consumption. In addition, at each appointment throughout the intervention period participants are asked to report how many days per week they have consumed the powder since the last appointment, on average.

For each randomised group, summary data will be presented for:

4.2.1 Powder consumption during the intervention period

This will be measured by the percentage of days during the intervention period each woman consumed a full dose of study powder, calculated as:

$$100 \times \frac{\text{Number of days a full dose of study powder was consumed}}{\text{Total number of days in intervention period}}$$

The number of days each woman consumed a full dose of study powder (numerator) is determined by dividing the calculated powder consumption weight (in grams) by

- 14.2 g/day for the women in the prebiotics (intervention) group, or
- 8.7 g/day for the women in the maltodextrin (control) group

For example, if a woman in the intervention (prebiotic) group consumed 4088g of powder throughout the study then the number of days a full dose of study powder was consumed would be 4088/14.2 = 288 days.

The total number of days in the intervention period (denominator) is calculated as the number of days between the randomisation date and six months post-partum.

The average percentage of days women consumed a full dose of study powder will be presented as a mean and standard deviation, or median and quartiles, as appropriate. Women who did not return the study powder will be excluded from the numeric summary and considered 'non-compliant' for the per-protocol analysis described in Section 4.3. Because this calculation requires knowledge of treatment group (due to different powder weights for intervention and placebo), this data will be summarised after the primary analysis is complete and the treatment groups have been unblinded.

4.2.2 Powder consumption during pregnancy

This will be measured by the percentage of days each woman consumed a full dose of study powder between randomisation and the 36 weeks' gestation appointment. Compliance to study powder over this period will be assessed using women's self-reported compliance at the 24, 28, 32, and 36-week appointments. This information will be used to approximate the average number of days study powder was consumed between randomisation and the 36-week appointment. This summary will only be calculated for women who were still pregnant at 36 weeks' gestation and reported compliance at the 36-week appointment. If compliance data from the 24, 28 and/or 32-week appointments is missing, then average compliance will be calculated using the available appointments. The denominator is the number of days between randomisation and the 36-week appointment. The average percentage of days women consumed a full dose of study powder will be presented as a mean and standard deviation, or median and quartiles, as appropriate.

4.2.3 Avoidance of non-study prebiotics

This will be reported as the proportion of women who did not consume non-study powder prebiotic drinks or supplements between randomisation and 6 months' post-partum. Participants who attended the 6-month appointment will be used as the denominator for this proportion.

4.2.4 Other protocol deviations

Frequencies and percentages will also be presented separately by randomised group for the following protocol deviations:

- Ineligible participant randomised;
- Randomised in the wrong stratum;
- Given the wrong study powder type according to randomisation;
- Never commenced study powder use
- Ceased study powder during pregnancy
- Ceased study powder between birth and 6 months' post-natal age
- 12-month clinical outcome assessment not occurring between 12 and 18 months of age
- Withdrawal from study
- Loss to follow-up

The reasons for not commencing the study powder will be described: Forgot; Changed mind, does not want to take the study powder; Health professional advice; Illness; Other.

The reasons for ceasing the study powder from randomisation to 36 weeks' gestation will be described: Does not want to take study powder anymore; Health professional advice; Perceived adverse reaction; Other.

The reasons for ceasing the study powder use between birth and 6 months' post-natal age will be described: Does not want to take study powder anymore; Health professional advice; Perceived adverse reaction; Other.

4.3 Analysis populations

For the primary outcome and secondary clinical outcomes (detailed in Section 6.1), the planned analyses will be performed using an intention to treat (ITT) approach. The ITT population will include all randomised woman-infant pairs, analysed as randomised, irrespective of eligibility or compliance with the protocol. Infant deaths occurring prior to the outcome measurement will be excluded from the ITT approach (see Section 6.2.6.5 for more detail),

For the primary outcome only, analysis will also be performed using a per-protocol approach. The per-protocol population will consist of all woman-infant pairs where:

- Women consumed their allocated study powder on at least 75% of days during the intervention period, from randomisation to 6 months' post-natal (as described in Section 4.2.1)
- Women avoided non-study powder prebiotic drinks and supplements during the intervention period (Section 4.2.3)
- No other protocol deviations were reported (as described in Section 4.2.4)

As outlined in Section 4.2.1, because assessing study powder compliance requires knowledge of the unblinded treatment group assignment, this per-protocol analysis will be performed after the blind has been broken (as described in Section 3.5).

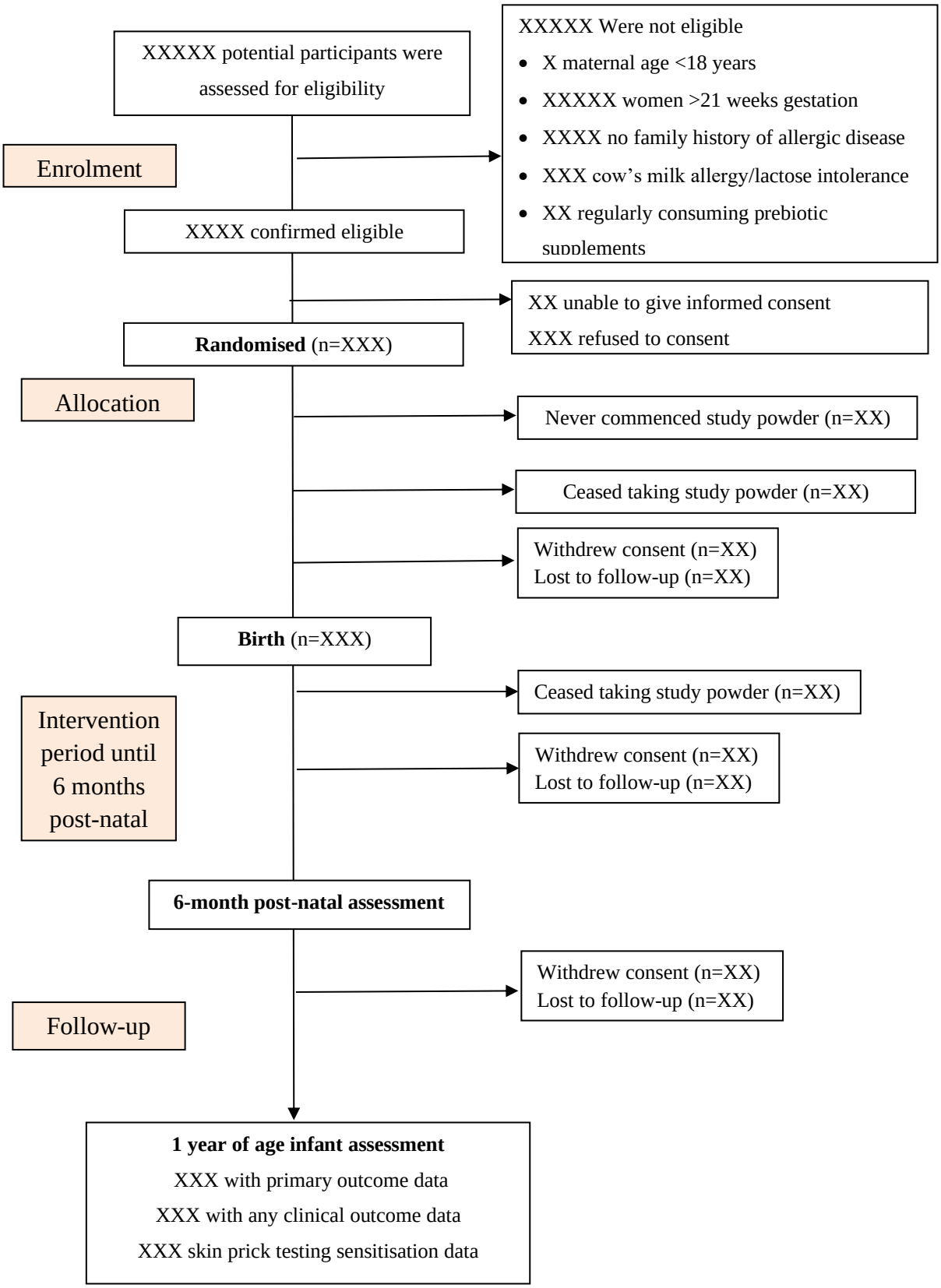
The per-protocol analysis described above will also be repeated for the pregnancy period with the requirement that women consumed 75% of their study powder and avoided non-study prebiotics from randomisation to 36 weeks' gestation (last clinic visit/phone call prior to birth), rather than the entire intervention period (Section 4.2.2).

5 TRIAL POPULATION

5.1 Screening data, eligibility, recruitment, and withdrawal/follow-up

The CONSORT flow diagram below will be completed to document numbers screened and randomised and the flow of participants through the trial (Figure 1).

Figure 1: CONSORT flow diagram to be completed for study



5.2 Baseline characteristics

A descriptive comparison of the randomised groups will be conducted on the baseline characteristics presented in the table below (Table 1).

Table 1: Baseline characteristics to be included in descriptive comparison

Baseline characteristic	Categories
Pre-pregnancy BMI (kg/m^2)	< 25 ≥ 25
Maternal history of allergic disease	Yes No
Paternal history of allergic disease	Yes No
Sibling history of allergic disease	Yes No Not applicable
Maternal age (years)	-
Maternal ethnicity	Caucasian Black African North African/Middle Eastern/Mediterranean Aboriginal / Torres Strait Islander East Asian (China, Japan, Korea) South / South East Asian Maori / Pacific Islander Other
Maternal school education - completed secondary school	Yes No
Maternal further education	No further study Certificate/diploma Degree Higher degree
Ever previously used prebiotic supplements	Yes No
Ever previously used probiotic supplements	Yes No
Gestation at randomisation (weeks)	-
Infant sex	Female Male

Means and standard deviations, or medians and quartiles will be reported for continuous variables. Frequencies and percentages will be reported for categorical variables. The clinical importance of any observed imbalances will be noted.

5.3 Post-randomisation characteristics

A descriptive comparison of the randomised groups will be conducted on the post randomisation characteristics presented in the table below (Table 2).

Table 2: Post-randomisation characteristics to be summarised descriptively

Post-randomisation characteristic	**Variable type**	**Summary measure**	**Derivation**
Maternal probiotic drinks or supplements consumption during intervention period	Categorical Yes No	N (%)	Yes if mother answered "Yes" to "Have you consumed any probiotic drinks or supplements" at any of the 9 appointments between 24 weeks GA and 6 months' post-partum
Maternal antibiotics taken during intervention period	Categorical Yes No	N (%)	Yes if mother answered "Yes" to "Have you taken any antibiotics" at any of the 9 appointments between 24 weeks' GA and 6 months' post-partum
Infant birth mode of delivery	Categorical Vaginal Caesarean section	N (%)	Vaginal birth includes spontaneous, instrumental vacuum extraction, and instrumental forceps. Caesarean section includes elective and non-elective caesarean.
Infant probiotic supplement consumption during intervention period	Categorical Yes No	N (%)	Yes if parent answered "Yes" to "Has your baby taken any probiotic supplements" at any of the 5 appointments between 1 month and 6 months' post-partum
Infant antibiotics taken during intervention period	Categorical Yes No	N (%)	Yes if parent answered "Yes" to "Has your baby taken any antibiotics" at any of the 5 appointments between 1 month and 6 months' post-partum
Ever breastfed	Categorical Yes No	N (%)	Yes if parent has answered "Yes" to "Has your baby been fed any breast milk" at any of the 6 appointments

			between 1 month and 12 months' post-partum
Breastfeeding duration (months)	Time-to-event	Proportion of women breastfeeding at 6 months (95 % CI) Proportion of women breastfeeding at 12 months (95 % CI) Median duration (95 % CI)	Time from infant DOB to breastfeeding cessation will be censored at the time of the 12-month appointment.
Infant given any infant formula up until 6 months of age	Categorical Yes No	N (%)	Yes if parent has answered "Yes" to "Has your baby been fed any infant formula" at any of the 5 appointments between 1 month and 6 months' post-partum
Infant given any prebiotic containing formula up until 6 months of age	Categorical Yes No	N (%)	Yes if the infant is given formula during the intervention period (as defined above) and the most-used formula brand reported by parent at one or more appointments contains prebiotics
Infant given any probiotic containing formula up until 6 months of age	Categorical Yes No	N (%)	As above, if the most-used formula brand reported at one or more appointment contains probiotics.
Age at introduction to infant formula (months) for infants given formula	Continuous	Median (Q1, Q3)	Reported only for infants who were introduced to formula during the intervention period. From answer to question "What was your baby's age when they were first given any infant formula?" If an answer is given at multiple appointments, then the value from the earliest available appointment will be used.
Infant age at time of 12-month appointment (in months)	Continuous	Median (Q1, Q3)	Calculated as the date of the 12-month appointment minus infant DOB (with no correction for GA).
Age at introduction to solid foods in months	Time-to-event	Median (95 % CI)	Time from infant DOB to introduction of solid food will be censored at the time of the 12-month appointment.
Age of introduction to egg in months	Time-to-event	Median (95 % CI)	As above

Age of introduction to cow's milk in months	Time-to-event	Median (95 % CI)	As above
Age of introduction to wheat in months	Time-to-event	Median (95 % CI)	As above
Age of introduction to fish in months	Time-to-event	Median (95 % CI)	As above
Age of introduction to peanut in months	Time-to-event	Median (95 % CI)	As above
Age of introduction to cashew nut in months	Time-to-event	Median (95 % CI)	As above

Means and standard deviations, or medians and values of the upper and lower quartiles will be reported for continuous variables. Frequencies and percentages will be reported for categorical variables.

Time-to-event variables will be censored at the time of the 12-month appointment and summarised using the Kaplan Meier estimate of survivorship (for proportion of women breastfeeding at 6/12 months) and/or the median and 95% confidence interval. The median and 95% confidence interval will be defined by drawing a horizontal line at 0.5 on the plot of the Kaplan Meier survival curve and its 95% confidence bands. The intersection of the line with the lower CI band will define the lower limit for the median's interval, and similarly for the upper band. If any of the intersections are not a point (e.g. the survival curve is exactly equal to 0.5 over an interval) then the centre of the intersection interval will be taken as the median. If the line does not intersect the survival curve, then the median will be undefined. However, it is anticipated that at least 50% of participants will have had the event of interest by the 12-month appointment (ceased breast-feeding or been introduced to each food).

It is anticipated that some participants will miss appointments throughout the study, resulting in missing data for these time points. For post-randomisation characteristics that are calculated using data from multiple appointments (e.g. Yes if parent has answered "Yes" at any of *N* appointments), the response for each participant will be calculated using all available data from their non-missing appointments. A post-randomisation characteristic for a participant will only be considered missing if data from *all* appointments required for its calculation are missing. A more formal imputation approach will not be considered here as this is a descriptive summary only.\

6 ANALYSIS

6.1 Trial outcomes

The primary and secondary outcomes for the trial are summarised in the table below:

Table 3: Definitions of primary and secondary trial outcomes

#	Outcome description	Outcome type	Derivation of outcome
Primary outcome			
1	Infant medically diagnosed eczema by 1 year of age	Binary	Parent has answered "Yes" to the question ": Has a medical doctor diagnosed your baby as having eczema/atopic dermatitis?" at one or more appointments (3-4, 6, or 12-month) This outcome will be treated as missing if the response for the 12-month appointment is missing, irrespective of responses given at earlier time points. If eczema is reported at the 3-4 or 6 month appointment but the 12-month response is missing, then the outcome will still be treated as missing in order to avoid introducing bias. If the response from the 3-4 or 6-month appointments is missing but the 12-month appointment is not, then the response is not considered missing, as missing information about eczema diagnosis over earlier time periods will be collected at the 12-month appointment.
Secondary outcomes			
2	Parent reported eczema symptoms by 1 year of age	Binary	Parent has answered "Yes" to the question "Has your baby shown signs of dry, red, itchy, scaly skin (eczema)?" at one or more appointments (3-4, 6, or 12 month). As per the primary outcome, this outcome will only be considered missing if data from the 12-month appointment is missing.
3	Eczema treatment with prescription steroid cream by 1 year of age	Binary	Parent has answered "Yes" to the question "Has your baby shown signs of dry, red, itchy, scaly skin (eczema)?", then answered "Yes" to the follow-up question "Is your baby currently on any treatments for eczema/atopic dermatitis?" and then selected prescription steroid cream from the list of treatments, at one or more appointments (3-4, 6, and/or 12 month appointment). As per the primary outcome, this outcome will only be considered missing if data from the 12-month appointment is missing.
4	Infant sensitisation to food and/or environmental allergens at 1 year of age	Binary	Positive skin prick test (mean weal diameter ≥3mm larger than the size of the negative control weal) to at least one of 9 food/environmental allergens at the 12-month appointment. Tested allergens include: egg, cow's milk, wheat, fish (tuna), peanut, cashew nut, ryegrass pollen, cat, and house dust mite (Dermatophagoides pteronyssinus).

5 - 13	Infant sensitisation to each of the following allergens at 1 year of age: Egg Cow's milk Wheat Fish (tuna) Peanut Cashew nut Ryegrass pollen Cat House dust mite	Binary	As described above, but with binary outcome for each individual allergen. Positive skin prick test (mean weal diameter ≥3mm larger than the size of the negative control weal) to [allergen] at the 12-month appointment.
14	Infant IgE-mediated food allergy by 1 year of age	Binary	Infant has IgE mediated food allergy to one or more of the following: • Egg • Cow's milk • Wheat • Fish (tuna) • Peanut • Cashew nut An infant is classified as having an IgE-mediated food allergy for a particular food if the following criteria are fulfilled: At the 6-month and/or 12-month appointment: • Parent reports that the food has been introduced (directly ingested) • Parent has selected "Yes within 2 hours" to the question "Did your child have a reaction to this food?" • The symptoms listed include at least one of the following: hives, swelling of face or body, wheeze or stridor, cough, vomiting, floppy unresponsive baby (corresponding to symptom numbers 2, 3, 4, 5, 6, or 9 on the CRF) AND at the 12-month appointment: • Infant has positive skin prick test (SPT) for the same food (as described above). This response for each food will be treated as missing if: There is no SPT result available, or SPT is positive but the child has not been introduced to the food at the time of the 12-month appointment (or response to reaction question is missing)

			If the SPT is negative and the child has not been introduced to the food (or response to the reaction question is otherwise missing), then the response for this food will be recorded as "Not allergic" rather than missing. If the response for any of the 6 individual foods is missing, then the outcome will be treated as missing.
15	Infant recurrent wheeze by 1 year of age	Binary	An infant is classified as having recurrent wheeze by 1 year of age if: At one or more appointments (3-4, 6, or 12-month), parent has answered "Yes" to the question "Has your child wheezed (ever/since the last appointment)", and The total number of wheezing episodes reported is >1. The total number of wheezing episodes will be calculated by summing responses to the question "Since birth/the last appointment, how many times (episodes) has your child wheezed" across the 3-4, 6 and 12-month appointments. If number of wheezes is not reported at an appointment, then it is assumed to be 0. This outcome is only treated as missing if the response at 12-month is missing, as earlier wheezing episodes will be captured at the 12-month appointment if the 3-4 or 6-month appointments were missed (as per primary outcome).
16	Maternal gestational diabetes	Binary	
17	Maternal pre-eclampsia	Binary	
18	Maternal weight gain (kg) per week from randomisation to 36 weeks' gestation	Continuous	This is calculated as $$\frac{[maternal\ weight\ at\ 36\ weeks] - [maternal\ weight\ at\ randomisation]}{([date\ of\ 36\ week\ appointment] - [randomisation\ date]) \div 7}$$ This analysis will exclude women who were no longer pregnant at 36 weeks of gestation
19	Maternal BMI at 3-4 months post-partum	Continuous	$$\frac{maternal\ weight\ at\ 3-4\ month\ appointment\ (kg)}{\left(height\ at\ randomisation\ (m)\right)^2}$$
20	Maternal BMI at 6 months post-partum	Continuous	$$\frac{maternal\ weight\ at\ 6\ month\ appointment\ (kg)}{\left(height\ at\ randomisation\ (m)\right)^2}$$
21	Infant weight at birth (g)	Continuous	
22	Infant length at birth (cm)	Continuous	

23	Infant head circumference at birth (cm)	Continuous	
24	Infant gestational age at birth (weeks)	Continuous	Gestational age at birth was recorded in weeks and days. This will be converted to weeks: $([GA\ in\ weeks] \times 7 + [GA\ in\ days]) \div 7$. If the number of days is not provided it will be assumed to be 0.
25	Infant preterm birth <37 weeks	Binary	Outcome is "Yes" if GA at birth in weeks (described above) is <37
26	Infant preterm birth <34 weeks	Binary	Outcome is "Yes" if GA at birth in weeks (described above) is <34

For infants with parent-reported eczema at the 3-4 month, 6 month, or 12-month appointments, eczema severity will be summarised descriptively using the objective SCORAD score recorded at each time point.

For infants with parent-reported eczema, the age at first appearance of eczema symptoms will also be summarised descriptively, calculated using parent's answer to the question "At what age did the eczema symptoms first occur?" This question is asked at the 3-4, 6, and 12 month appointments. If more than one answer is given, the answer from the earliest appointment will be used.

Both SCORAD score and age at first appearance of eczema symptoms will be treated as continuous variables and summarised using mean and standard deviation, or median and quartiles, as appropriate.

6.2 Analysis methods

6.2.1 Overall analysis approach

The primary outcome and all binary secondary outcomes will be analysed using log binomial regression models, with the effect of treatment described as a relative risk with a 95% confidence interval. Should any of the models fail to converge, a known problem with log binomial regression, a log Poisson model using generalised estimating equations (independence working correlation structure assumed) will be used for analysis(1). If the number of infants experiencing an outcome is considered too small for a regression model to be sensible (less than 5 events in either randomised group) then a Fisher exact test will be performed instead, regardless of convergence.

Continuous secondary endpoints will be analysed using linear regression models, with the effect of treatment described as a difference in mean outcomes with 95% confidence interval. Data transformations are not planned to correct for departures from normality, since the sample size is sufficient for the central limit theorem to apply(2).

6.2.2 Covariate adjustment

Analyses will be adjusted for the two stratification variables: maternal history of allergic disease (yes/no - history of medically diagnosed asthma, allergic rhinitis, eczema and/or food allergy) and maternal pre-pregnancy BMI. Maternal history of allergic disease will be modelled as a binary categorical variable and

BMI will be modelled linearly as a continuous variable. For outcomes related to infant allergic disease (outcomes 1 – 15 in Table 3) adjustment will also be made for infant sex as this is considered an important prognostic variable for infant allergic disease outcomes(3,4). All adjustment variables will be treated as fixed effects in the analysis models.

If the adjusted log Poisson model fails to converge for any of the outcomes, adjustment variables will be removed sequentially (infant sex, then maternal pre-pregnancy BMI, then maternal history of allergic disease) until convergence is achieved. Adjusted analyses will not be considered for binary outcomes analysed using a Fisher exact test.

6.2.3 Methods for addressing outlying values

Outliers will be queried during data collection and the statistical analysis. Unless confirmed as a data entry error, outliers will not be excluded from any analyses.

6.2.4 Methods for handling missing data

Missing data will be summarised descriptively by treatment group for all baseline characteristics (Table 1), post-randomisation characteristics (Table 2), and outcome variables (Table 3). Composite binary outcomes will be treated as missing if any of the individual components of the outcome are missing, even if the composite can be derived from the observed components, unless otherwise specified in Table 3. Variables not ascertained due to infant death will not be treated as missing, as such data are not meaningful for analysis. Instead, these data will be treated as undefined and excluded from analyses (see Section 6.2.6.5).

To address missing data in the primary outcome and secondary clinical outcomes, multiple imputation performed under a missing at random assumption will be used to create 100 complete datasets for analysis. Multiple imputation will be used even if only a small percentage of data are missing. Use of 100 imputations ensures that the loss of power compared to full information maximum likelihood methods is minimal(5), which is important in the context of a confirmatory clinical trial. Imputation will be performed separately by treatment group using fully conditional specification, also known as chained equations. The conditional logistic imputation models for the incomplete outcomes will include covariates pre-specified for adjustment (Section 6.2.2). Additional auxiliary variables associated with the incomplete outcomes will also be added to the imputation model as appropriate to improve the prediction of missing values and the plausibility of the missing at random assumption.

If there are less than 20 events in either arm for an outcome, then multiple imputation will not be performed as imputation models may be overfit. In this case, a complete case analysis will be performed instead.

For composite binary outcomes (numbered 4 and 14 in Table 3), the composite outcome will be calculated using available data and multiple imputation will be used to impute the composite outcome (calculated prior to imputation), as opposed to multiply imputing each individual component and then calculating the composite.

For the primary outcome, the sensitivity of results to the missing at random assumption will be explored by considering missing not at random mechanisms. Using pattern mixture models, the odds of infant medically diagnosed eczema will be assumed to be between half and twice as high in infants with missing data compared to infants with observed data. These differences will be applied to control group infants only, treatment group infants only, and infants in both groups.

For the primary outcome and secondary clinical outcomes, a complete case analysis will also be performed on the unimputed data for comparison with imputed results. However, these analyses will not be used to inform conclusions about the effect of treatment.

6.2.5 Clustering

If a pregnant woman gave birth to multiple infants, then only the first-born infant was included in the trial. Therefore, statistical methods to account for partial clustering of multiple infants per mother are not required for this trial.

6.2.6 Estimands

The estimand is a precise description of the treatment effect to be estimated from a trial, consisting of five attributes: the treatment conditions, target population, outcome variable, population-level summary, and handling of intercurrent events. While some attributes apply to all primary and secondary outcomes in this trial, the target population and intercurrent events vary between outcomes depending on the timing of the assessment and whether the outcome was measured in the mother or infant. For the purpose of defining the estimands, the trial outcomes can be grouped as follows (using the reference numbers from Table 3).

#	Population and time of outcome assessment
1 - 15	Infant allergic outcomes. Outcome determined in the infant at the 12-month appointment.
16 – 17	Maternal pre-eclampsia and diabetes.

	Outcome determined in the mother at end of pregnancy
18	Maternal weight gain during pregnancy. Outcome determined in the mother at 36 weeks' gestation
19 - 20	Maternal BMI. Outcome determined in the mother at 3-4 months' and 6 months' post-natal
21 – 26	Infant birth outcomes (anthropometrics and gestational age). Outcome determined in the infant at the time of birth

6.2.6.1 Treatment conditions

Consumption of 14.2g per day of prebiotic powder (intervention) versus 8.7g per day of maltodextrin powder (placebo) throughout the intervention period (randomisation until 6 months' postpartum). Participants in both groups are asked not to consume any other prebiotic products during the intervention period.

For secondary outcomes that are assessed before the end of the intervention period (16 – 17 and 21 – 26), the treatment conditions are then consumption of the intervention or placebo powder from randomisation until the time of outcome assessment.

6.2.6.2 Target population

Pregnant women meeting the inclusion criteria for the study are the target population for all trial outcomes:

Pregnant women who are:

- aged ≥18 years
- currently at <21 weeks' gestation
- not currently consuming prebiotic supplements more than two times per week

and whose infant will

- have one or more first-degree relative with a history of medically diagnosed allergic disease.

The following additional specifications to the target population apply depending on the outcome:

Outcomes 1 – 15: Additional requirement that infant is alive at 1 year of age.

Outcome 18: Additional requirement that woman is still pregnant at 36 weeks' gestation

Outcome 21 – 26: Additional requirement that pregnancy results in a live birth

6.2.6.3 Outcome measure of interest

The outcome measures are defined in Table 3.

6.2.6.4 Population-level summary

The population level summary is the relative risk for binary outcomes and difference in mean outcome for continuous outcomes (Section 6.2.1).

6.2.6.5 Intercurrent events

Possible intercurrent events, the trial outcomes they apply to, and the strategies for handling them are outlined below in Table 4.

Table 4: Strategies for handling intercurrent events for trial outcomes

Intercurrent event	**Outcomes applicable**	**Strategy**
Treatment discontinuation for any reason during the intervention period (randomisation to 6 months' postpartum). Protocol violation (consumed non-study prebiotics during intervention period)	All trial outcomes	In the event of treatment discontinuation or protocol violation, a treatment policy strategy will be used. That is, participants will be analysed in the treatment group to which they were assigned, regardless of how much study powder they consumed or whether they consumed non-study prebiotics. This corresponds to the ITT policy.
Infant death prior to outcome assessment (either during pregnancy or after birth)	1 - 15, 18, 21 - 26	In the case of infant death, the outcomes will be treated as undefined and excluded from analyses. Such a "survivors analysis" has been recommended in settings where the intervention is considered biologically unlikely to impact on the risk of mortality(6) as is the case here. If there is any evidence to suggest that the risk of infant death is influenced by the intervention ($p<0.20$ according to a Fisher exact test), we will also consider the effect of the intervention on the composite of death and infant medically diagnosed eczema (i.e., the primary outcome) in a post-hoc analysis.
Maternal death prior to outcome assessment	16 - 26	Maternal death will be handled using the "survivors analysis" strategy defined above.
Preterm birth (woman is no longer pregnant at 36 weeks' gestation)	18	Women who are no longer pregnant at 36 weeks' gestation will be excluded from the analysis of maternal weight-gain during pregnancy. If the treatment is found to have an effect on pre-term birth, then the results of this analysis will be interpreted with caution.

6.3 Harms

The number and percentage of mother-infant pairs experiencing an adverse event (AE) will be summarised descriptively for each treatment group (irrespective of eligibility or compliance with the protocol). The

number of serious adverse events (SAEs) will be compared between treatment groups using Fisher exact tests. The following AEs will be evaluated:

- Maternal admission to intensive care unit (SAE)
- Maternal death (SAE)
- Infant admission to intensive care unit (SAE)
- Infant death, including stillbirths (SAE)
- Any SAE
- Maternal hospitalisation > 24 hours, excluding admission for baby's birth (AE)
- Infant hospitalisation > 24 hours, excluding admission for baby's birth (AE)
- Maternal symptoms experienced during the intervention period (total number of days)
    - Loose watery stools
    - Blood-stained stools
    - Constipation
    - Vomiting
    - Excessive bloating
    - Abdominal pain
- Infant symptoms experienced during the intervention period (total number of days)
    - Blood-stained stools
    - Loose watery stools
    - Vomiting
    - Eczema
    - Hives
    - Generalised Skin Rash
    - Irritable (more than usual)
    - Floppy Unresponsive

6.4 Analysis Software

All analyses will be performed using R version 4.2.2 or later.

6.5 Analyses not included in SAP

The following secondary outcomes are outlined in the trial protocol but will be analysed in separate projects rather than the primary trial analysis. Therefore, they have not been included in this SAP:

- Maternal and infant transepidermal water loss (TEWL) measurements - Maternal and infant skin barrier permeability at 36 weeks (maternal only), 3–4 and 6 months post-natal, and at 1 year of age (infant only).

- Maternal hypertension
- Infant birth body composition.
- Infant weight gain and growth during infancy.
- Secondary laboratory outcomes:
  - Maternal immune responses at 20, 28 and 36 weeks of gestation.
  - Maternal gut microflora colonisation patterns and stool short-chain fatty acid levels and composition during the pregnancy and lactation intervention period.
  - Infant gut microflora colonisation patterns and stool short-chain fatty acid profiles at 2, 3–4, 6 and 12 months of age.
  - Infant immune responses at birth, 3–4, 6 and 12 months of age.
  - Breast milk composition and immune components at 2, 3-4, 6 and 12 months of age.

7 REFERENCES

1. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702–6.

2. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health. 2002;23:151–69.

3. Nurmatov U, Nwaru BI, Devereux G, Sheikh A. Confounding and effect modification in studies of diet and childhood asthma and allergies. Allergy. 2012 Aug;67(8):1041–59.

4. Pali-Schöll I, Jensen-Jarolim E. Gender aspects in food allergy. Curr Opin Allergy Clin Immunol. 2019 Jun;19(3):249–55.

5. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007 Sep;8(3):206–13.

6. Colantuoni E, Scharfstein DO, Wang C, Hashem MD, Leroux A, Needham DM, et al. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. BMJ. 2018 Jan 3;360:j5748.