

**PrEggNut Study**  
**STATISTICAL ANALYSIS PLAN**

**Study Title: Maternal diet rich in eggs and peanuts to reduce food allergies: a randomised controlled trial [PrEggNut Study]**

**Trial registration** Australian and New Zealand Clinical Trials Registry  
ACTRN12618000937213

**SAP version** 1.0, 28/07/2021

**SAP revision history** -

**Funding** National Health & Medical Research Council (NHMRC)  
Project Grant ID 1147576

**SAP Authors**

**Dr Thomas Sullivan**  
Chief Investigator and Trial Biostatistician  
SAHMRI Women & Kids, South Australian Health &  
Medical Research Institute

**Dr Debra Palmer**  
Principal Investigator and Chair of the PrEggNut Study  
Steering Committee  
Telethon Kids Institute

**Prof Maria Makrides**  
Chief Investigator  
SAHMRI Women & Kids, South Australian Health &  
Medical Research Institute

	Signature	SAP version	Date
Thomas Sullivan	.....		
Debra Palmer	.....		
Maria Makrides	.....		

Version 1.0

28/07/2021

PrEggNut Study	Statistical Analysis Plan	Page 2
Table of Contents		
1	PREFACE.....	4
2	INTRODUCTION.....	4
2.1	Background and rationale.....	4
2.2	Objectives.....	5
3	STUDY METHODS.....	5
3.1	Trial design.....	5
3.2	Randomisation.....	5
3.3	Sample size.....	5
3.4	Framework.....	6
3.5	Statistical interim analyses and stopping guidance.....	6
3.6	Timing of final analysis and unblinding.....	6
3.7	Timing of outcome assessments.....	6
4	STATISTICAL PRINCIPLES.....	6
4.1	Confidence intervals and p values.....	6
4.2	Adherence and protocol deviations.....	7
4.3	Analysis populations.....	8
5	TRIAL POPULATION.....	9
5.1	Screening data, eligibility, recruitment and withdrawal/follow-up.....	9
5.2	Baseline characteristics.....	10
5.3	Post-randomisation characteristics.....	12
6	ANALYSIS.....	13
6.1	Primary outcome.....	13
6.1.1	IgE-mediated egg and/or peanut allergy at 12 months of age.....	13
6.2	Secondary clinical outcomes.....	13
6.2.1	Sensitisation to egg.....	13
6.2.2	Sensitisation to peanut.....	13
6.2.3	IgE-mediated egg allergy.....	13
6.2.4	IgE-mediated peanut allergy.....	14
6.2.5	Infant eczema.....	14
6.3	Exploratory laboratory outcomes.....	14
6.3.1	Maternal egg specific IgG4 concentrations.....	14
6.3.2	Maternal peanut specific IgG4 concentrations.....	14
6.3.3	Infant egg specific IgG4 concentrations.....	14
6.3.4	Infant peanut specific IgG4 concentrations.....	14
6.3.5	Infant egg specific IgE concentrations.....	15
6.3.6	Infant peanut specific IgE concentrations.....	15

PrEggNut Study	Statistical Analysis Plan	Page 3
6.4	Analysis methods.....	15
6.4.1	Overall analysis approach.....	15
6.4.2	Covariate adjustment .....	15
6.4.3	Planned subgroup analyses.....	16
6.4.4	Methods for addressing outlying values .....	16
6.4.5	Methods for handling missing data.....	17
6.4.6	Methods for addressing infant deaths .....	18
6.5	Harms .....	18
6.6	Analysis Software.....	18
7	REFERENCES .....	19

## 1 PREFACE

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

- PrEggNut case report form (CRF) (20<sup>th</sup> March 2019);
- PrEggNut NHMRC project grant application (APP1147576);
- PrEggNut trial protocol (version 2, 11<sup>th</sup> June 2019).

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

Australia has the highest reported prevalence of food allergy in children in the world. Recent randomised controlled trials have shown that regularly including traditionally allergenic foods, such as egg and peanut, with solid foods from mid-late infancy reduces the risk of developing egg and peanut allergies in some infants. Australian studies, led by Chief Investigators Palmer, Makrides, Prescott, Campbell and Gold, have made a key contribution to clinical practice internationally in this field. Our findings also demonstrate that a significant proportion of infants have allergic reactions, including anaphylaxis, on first introduction of egg in solid foods as early as 4 months of age. We have recently shown that food-allergen (egg) specific immune responses can be established prior to infants eating any egg-allergen in solid foods. Furthermore, these responses were not altered by early introduction of egg in the infant diet. In other words, it is too late for many infants who are already allergic by the time these foods are introduced around 4-6 months of age.

Food allergens are first encountered before birth and can be detected in amniotic fluid, and in the immediate postnatal period in breastmilk. Intriguingly, the fetal immune system has been indicated to be particularly tolerogenic during this developmental period. Immune tolerance appears to be the default response to allergen encounter. Hence, early allergen exposure is a logical strategy for food allergy prevention. Accordingly, we have recently shown that higher maternal intakes of egg during early lactation can beneficially modify infant egg-specific immune responses. However, it is not known whether these changes are associated with a reduced likelihood of developing egg allergy in

infants. Hence this randomised controlled trial will investigate higher regular egg and peanut maternal dietary intakes during pregnancy and lactation as a strategy for infant food allergy prevention.

## 2.2 Objectives

To determine whether the incidence of food allergies in infancy can be reduced by a maternal diet rich in eggs and peanuts during pregnancy and lactation.

## 3 STUDY METHODS

### 3.1 Trial design

Multi-centre, parallel, two-arm (1:1 allocation), researcher blinded (outcome data assessors, statistical analyst and investigators), randomised trial. Women with a singleton pregnancy who are planning to breastfeed for at least 4 months will be randomised to follow either a high (at least 6 eggs and 60 peanuts per week) or standard (no more than 3 eggs and 30 peanuts per week) egg and peanut intake diet from 22 weeks gestation until 4 months postnatal infant age. The primary outcome is food challenge proven IgE-mediated egg and/or peanut allergy in the infants at 12 months of age.

### 3.2 Randomisation

Pregnant women are assigned to a high egg and peanut diet (treatment group) or standard egg and peanut diet (control group) using a secure web-based randomisation service. The randomisation service allocates group assignments according to a computer-generated randomisation schedule, produced by an independent statistician using `ralloc.ado` version 3.7.6 in Stata version 15.1. Randomisation was stratified by city (Adelaide, Melbourne, Perth, Sydney) and by first or subsequent born child to the mother using randomly permuted blocks of varying sizes.

### 3.3 Sample size

To detect a reduction in egg and/or peanut allergy from 16% in the control group to 11.2% in the treatment group (absolute reduction 4.8%, relative reduction 30%) with 85% power and two-sided alpha of 0.049 at the final analysis, 961 women per group were required (with a continuity correction applied). Conservatively assuming 10% loss to follow-up, this led to a sample size estimate of 1068 women per group, or 2136 women total. Further details on the assumptions involved in the sample size calculations are provided in the PrEggNut trial protocol (Version 2, Section 5.7).

### 3.4 Framework

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

### 3.5 Statistical interim analyses and stopping guidance

An independent statistician will conduct a single interim analysis of the primary outcome once 961 infants (50% of the planned sample size, accounting for loss to follow-up) have primary outcome data available. The analysis will be performed according to the statistical methods described in Section 6. An independent Data Monitoring Committee will review the results of this analysis and employ O'Brien-Fleming stopping criteria (1) to maintain the overall alpha for the primary outcome across interim and final analyses at 0.05. A two-sided p-value of less than 0.0031 at the interim analysis (symmetric stopping boundary of  $Z = \pm 2.96259$ ), indicating a large difference between treatment groups, will be taken to provide statistical evidence in support of early stopping. A p-value of 0.0490 ( $Z = \pm 1.96857$ ) will be used to indicate statistical significance at the final analysis.

### 3.6 Timing of final 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.

### 3.7 Timing of outcome assessments

The 12-month clinical outcome assessment should occur when infants are between 12 and 15 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, subgroup analyses and sensitivity analyses for the primary outcome. No multiplicity adjustment will be made for the number of secondary 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 analyses.

## 4.2 Adherence and protocol deviations

Adherence to the intervention involves (a) following the recommended intake of egg and peanuts to 4 months postnatal age, and (b) continuing breastfeeding to 4 months postnatal age. Women will be considered to have adhered to diet recommendations if they reported meeting recommendations in the last week (at least 6 eggs and 60 peanuts for the treatment group, no more than 3 eggs and 30 peanuts for the control group) across at least 75% of their scheduled pre-natal and post-natal assessments. Depending on timing of birth, pre-natal assessments are scheduled at 26, 30, 34 and 38 weeks gestation, while post-natal assessments are scheduled at 1, 2, 3 and 4 months after delivery. Women with missing data on egg and peanut intake at a scheduled appointment will be considered to have not met diet recommendations at that specific appointment. Women will be considered to have adhered to the breastfeeding recommendations if they continue any breastfeeding, including use of expressed breast milk, until at least 4 months of age. Women with missing data on breastfeeding duration will be considered to have not met breastfeeding recommendations.

For each randomised group, frequencies and percentages will be presented for:

- Overall adherence (following dietary consumption and breastfeeding recommendations);
- Adherence to breastfeeding recommendations;
- Adherence to both egg and peanut dietary consumption recommendations;
- Adherence to egg consumption recommendations;
- Adherence to peanut consumption recommendations.

The participants included in the ITT analysis dataset (see Section 4.3) will be used as the denominator in the calculation of percentages. The number of eggs and peanuts consumed in the last week at each scheduled assessment for each randomised group will also be described using medians and inter-quartile ranges.

Frequencies and percentages (of the ITT analysis dataset) will also be presented separately by randomised group for the following protocol deviations:

- Ineligible participant randomised;
- Randomised in the wrong stratum;
- Given the wrong dietary recommendations according to randomisation;
- 12-month clinical outcome assessment not occurring between 12 and 15 months of age;
- Withdrawal from study;
- Loss to follow-up.

### 4.3 Analysis populations

For the primary outcome and secondary clinical outcomes (detailed in Sections 6.1 and 6.2), the planned analyses will be performed using an ITT approach. Excluding infant deaths occurring prior to outcome measurement (see Section 6.4.6 for more detail), the ITT population will include all randomised women-infant pairs, analysed as randomised, irrespective of eligibility or compliance with the protocol. For the primary outcome only, a sensitivity analysis will also be performed using a per-protocol approach. The per-protocol population will consist of all women-infant pairs that:

- Followed dietary consumption and breastfeeding recommendations (as described in Section 4.2);
- Did not report a protocol deviation (as described in Section 4.2);
- Provided data on the primary outcome at 12 months.

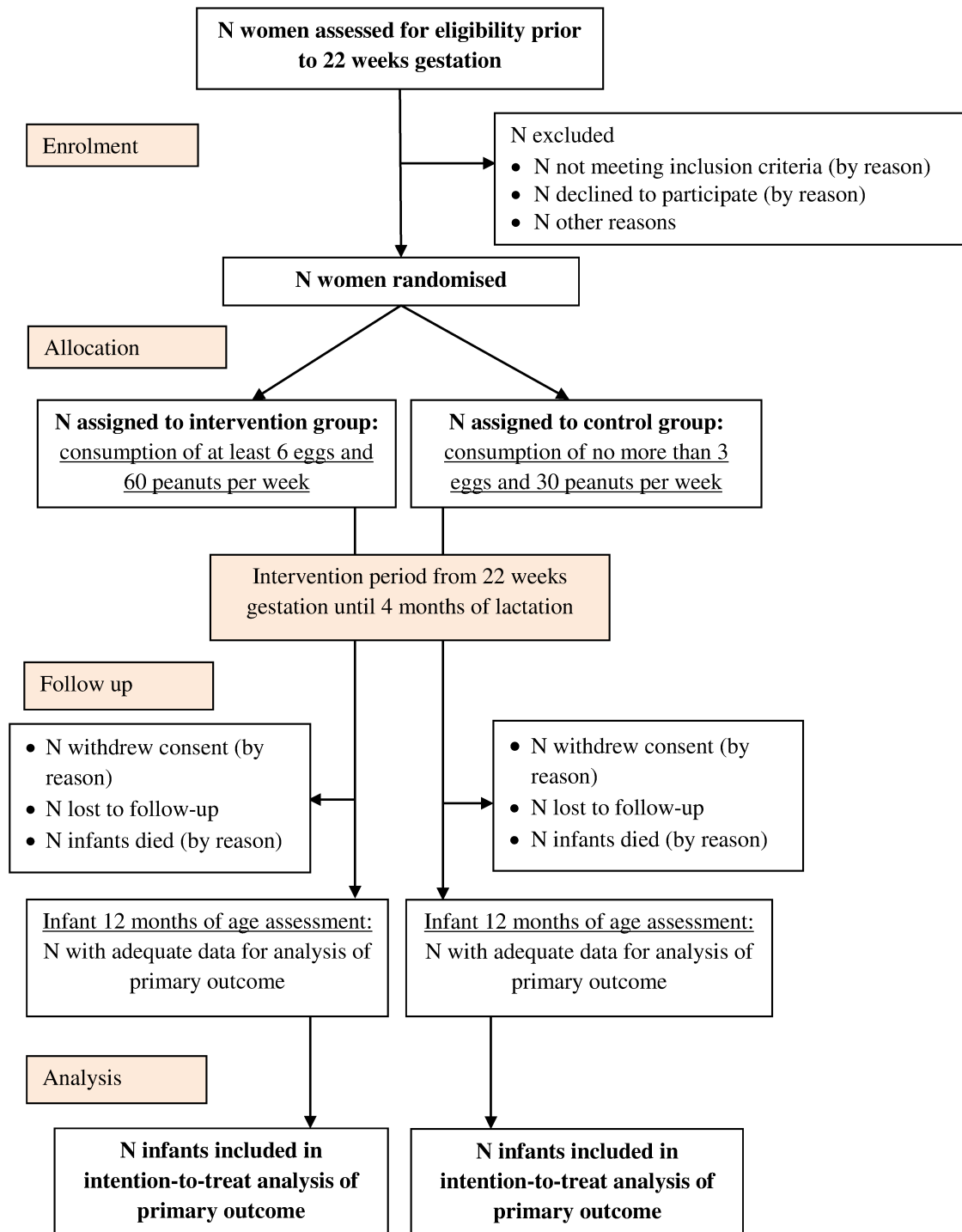
For exploratory laboratory outcomes (Section 6.3) collected in Perth and Sydney, the analysis population will consist of all women or infants that provided a blood sample at the relevant time point, analysed according to their randomised group.



## 5 TRIAL POPULATION

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

The following CONSORT flow diagram will be completed to document numbers screened and randomised and the flow of participants through the trial.



The number of screened women and exclusions from the trial (not meeting the protocol-defined inclusion criteria, declined to participate, withdrew consent and loss to follow-up) will also be reported separately for each city.

## 5.2 Baseline characteristics

A descriptive comparison of the randomised groups will be conducted on the baseline characteristics presented in the following table.

Baseline characteristic	Categories
City	Adelaide Melbourne Perth Sydney
Maternal age in years	-
Maternal ethnicity	Caucasian Aboriginal/Torres Strait Islander Maori/Pacific Islander African Middle Eastern Mediterranean East Asian South/South East Asian Other
Maternal school education - completion of secondary school	Yes/No
Maternal further education	No further study Certificate/diploma Degree Higher degree
Maternal smoking	Yes/No
Other smoking in the household	Yes/No
Pet dog(s)	Yes/No
Pet cat(s)	Yes/No

Pet chicken(s)	Yes/No
Parity	First born child Subsequent born child
Gestation at randomisation in weeks	-
Pre-pregnancy weight in kg	-
Current weight in kg	-
Maternal BMI in kg/m <sup>2</sup>	-
Maternal history of allergic disease	Yes/No
Paternal history of allergic disease	Yes/No
Sibling history of allergic disease	Yes No Not applicable
Maternal history of food allergy	Yes/No
Paternal history of food allergy	Yes/No
Sibling history of food allergy	Yes No Not applicable
Maternal peanut intake per week	-
Total household peanut intake per week	-
Maternal egg intake per week	-
Total household egg intake per week	-
Maternal egg specific-IgG4 levels (Perth and Sydney women only)	-
Maternal peanut specific-IgG4 levels (Perth and Sydney women only)	-
Infant sex	Female Male

Means and standard deviations, or medians and interquartile ranges 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 following table:

Post-randomisation characteristic	Categories
Infant birth weight in kg	-
Infant gestational age at birth in weeks	-
Preterm birth < 37 weeks	Yes/No
Early preterm birth < 34 weeks	Yes/No
Infant birth mode of delivery	Vaginal Caesarean section
Ever breastfed	Yes/No
Breastfeeding during the whole intervention period until infant is 4 months of age	Yes/No
Breastfeeding up to 8 months of age	Yes/No
Breastfeeding up to 12 months of age	Yes/No
Breastfeeding duration in months*	-
Maternal mastitis during the intervention period	Yes/No
Maternal weight change during intervention period in kg	-
Infant given any infant formula during intervention period	Yes/No
Age at introduction to infant formula in months in infants given formula	-
Age at introduction to solid foods in months*	-
Age of introduction to egg in months*	-
Age of introduction to peanut in months*	-
Infant weight at 4 months of age in kg	-

\* Time to event variable

Means and standard deviations, or medians and interquartile ranges will be reported for continuous and time to event variables. Frequencies and percentages will be reported for categorical variables.

Differences between study groups will be tested using t-tests for continuous variables, Fisher exact tests for categorical variables and log-rank tests for time to event variables.

## 6 ANALYSIS

### 6.1 Primary outcome

#### 6.1.1 IgE-mediated egg and/or peanut allergy at 12 months of age

Binary outcome based on measurements on the infant at 12 months of age. IgE-mediated egg and/or peanut allergy is defined as an allergic reaction to egg with sensitisation to egg and/or an allergic reaction to peanut with sensitisation to peanut. Sensitisation is defined as a 3mm or greater weal size based on a skin prick test (SPT). Infants with a positive SPT who do not proceed with the corresponding food challenge due to previous anaphylaxis or allergic reaction will be considered positive for IgE-mediated egg and/or peanut allergy; otherwise, the allergic reaction will be established according to an oral food challenge.

Note: Sensitisation is defined the same way throughout Section 6.2.

### 6.2 Secondary clinical outcomes

#### 6.2.1 Sensitisation to egg

Binary outcome based on a positive SPT to egg.

#### 6.2.2 Sensitisation to peanut

Binary outcome based on a positive SPT to peanut.

#### 6.2.3 IgE-mediated egg allergy

Binary outcome based on an allergic reaction to egg with sensitisation to egg. Infants with a positive SPT for egg who do not proceed with the egg food challenge due to previous anaphylaxis or allergic reaction to egg will be considered positive for IgE-mediated egg allergy; otherwise, the allergic reaction will be established according to an oral egg challenge.

### 6.2.4 IgE-mediated peanut allergy

Binary outcome based on an allergic reaction to peanut with sensitisation to peanut. Infants with a positive SPT for peanut who do not proceed with the peanut food challenge due to previous anaphylaxis or allergic reaction to peanut will be considered positive for IgE-mediated peanut allergy; otherwise, the allergic reaction will be established according to an oral peanut challenge.

### 6.2.5 Infant eczema

Binary outcomes based on a medical diagnosis of infant eczema by 4 and 12 months of age.

## 6.3 Exploratory laboratory outcomes

### 6.3.1 Maternal egg specific IgG4 concentrations

Continuous outcomes based on egg specific IgG4 levels (mgA/L) in maternal blood samples at 34-38 weeks gestation and 4 months postnatal age (Perth and Sydney women only). Egg-specific IgG4 serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.07 mgA/L.

### 6.3.2 Maternal peanut specific IgG4 concentrations

Continuous outcomes based on peanut specific IgG4 levels (mgA/L) in maternal blood samples at 34-38 weeks gestation and 4 months postnatal age (Perth and Sydney women only). Peanut-specific IgG4 serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.07 mgA/L.

### 6.3.3 Infant egg specific IgG4 concentrations

Continuous outcomes based on egg specific IgG4 levels (mgA/L) in infant blood samples at 4 and 12 months of age (Perth and Sydney infants only). Egg-specific IgG4 serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.07 mgA/L.

### 6.3.4 Infant peanut specific IgG4 concentrations

Continuous outcomes based on peanut specific IgG4 levels (mgA/L) in infant blood samples at 4 and 12 months of age (Perth and Sydney infants only). Peanut-specific IgG4 serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.07 mgA/L.

### 6.3.5 Infant egg specific IgE concentrations

Continuous outcomes based on egg specific IgE levels (kUA/L) in infant blood samples at 4 and 12 months of age (Perth and Sydney infants only). Egg-specific IgE serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.1 kUA/L.

### 6.3.6 Infant peanut specific IgE concentrations

Continuous outcomes based on peanut specific IgE levels (kUA/L) in infant blood samples at 4 and 12 months of age (Perth and Sydney infants only). Peanut-specific IgE serum antibody concentrations will be measured using the ImmunoCAP 250 system, which has a lower limit of detection of 0.1 kUA/L.

## 6.4 Analysis methods

### 6.4.1 Overall analysis approach

The primary outcome and all secondary clinical outcomes (see Sections 6.1 and 6.2) 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 (2). 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, regardless of convergence, a Fisher exact test will be performed instead.

To account for censoring due to the lower detection limits of the measuring equipment, the exploratory laboratory outcomes (Section 6.3) will be analysed using tobit regression models. Based on previous experience with these measures (3), the IgG4 and IgE concentrations will be log transformed prior to analysis to satisfy an assumption of homogeneous error variance. The effect of treatment on the latent (uncensored) IgG4 and IgE concentrations will be described as a ratio of means with a 95% confidence interval. Should the assumption of homogeneous error variance on the log scale be deemed unreasonable, based on a scatterplot of model residuals versus fitted values, other transformations will be explored and justified as appropriate.

### 6.4.2 Covariate adjustment

Given recommendations to adjust for variables used to stratify the randomisation when estimating treatment effects (4), analyses will be adjusted for city (Adelaide, Melbourne, Perth, Sydney) and

birth order (first or subsequent born child to the mother). Adjustment will also be made for infant sex and maternal history of allergic disease (defined as a medical diagnosis of asthma, eczema, hayfever or food allergy), both considered important prognostic variables for infant allergic disease outcomes (5, 6). All adjustment variables will be treated as fixed effects in the analysis models. For each outcome, both unadjusted and adjusted analyses will be performed, with the adjusted analyses used to draw final conclusions about the effect of treatment.

If adjusted models fail to converge for any of the outcomes, adjustment variables will be removed sequentially (infant sex, then maternal history of allergic disease, then city, then birth order) until convergence is achieved. Adjusted analyses will not be considered for binary outcomes analysed using a Fisher exact test (see Section 6.4.1).

#### 6.4.3 Planned subgroup analyses

For the primary outcome only, analyses will be performed to test for evidence of effect modification by (1) birth order (first or subsequent born child to the mother), (2) baseline total household egg intake, (3) baseline total household peanut intake, and (4) socio-economic status, determined using the home post-code of the mother and the index of relative socio-economic advantage and disadvantage ([www.abs.gov.au/websitedbs/censushome.nsf/home/seifa](http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa)). Effect modification by birth order will be assessed by including this subgroup variable as well as its interaction with treatment group into the log binomial regression model for the primary outcome (with the covariates detailed in Section 6.4.2 also included in the model). Effect modification by total household egg intake, total household peanut intake and socio-economic status will be assessed in a similar fashion in separate log binomial models, but with these variables treated as continuous rather than categorical in the analysis. To account for potential non-linear effects, two-term fractional polynomials will be fitted using the “mfp” command in Stata v16 (or later) using default settings (7). For each potential effect modifier, the p-value for the interaction term with treatment group will be reported. Independent of the statistical significance of the interaction p-value, estimates of the treatment effect with 95% confidence intervals will be reported for each birth order subgroup or in treatment effect plots for total household egg intake, total household peanut intake and socio-economic status.

#### 6.4.4 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.



#### 6.4.5 Methods for handling missing data

Missing data will be summarised descriptively by treatment group for all baseline characteristics (Section 5.2), post-randomisation characteristics (Section 5.3), outcome variables (Sections 6.1 to 6.3), covariates for adjustment (Section 6.4.2) and covariates for subgroup analyses (Section 6.4.3). Composite binary outcomes (6.1.1, 6.2.3 and 6.2.4) 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. 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 6.4.6).

To address missing data on 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, 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 (8), 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 6.4.2) and for conducting subgroup analyses (Section 6.4.3). 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. For composite binary outcomes (6.1.1, 6.2.3 and 6.2.4), conditional logistic imputation models will be defined for each component and the composite outcome calculated following imputation.

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 IgE mediated egg and/or peanut allergy will be assumed to be between half and twice as high in children with missing data compared to children with observed data. These differences will be applied to control group children only, treatment group children only and children in both treatment 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.

Considering the exploratory nature of the laboratory outcomes and the expectation of little missing data among the analysis population of women and infants providing a blood sample at each time point, analyses of these outcomes will be restricted to participants with complete outcome data.

#### 6.4.6 Methods for addressing infant deaths

It is anticipated that a small proportion of infants will die during the study, either during pregnancy or following birth, preventing the later measurement of outcome variables. In these cases, 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 (9), as is the case here with peanut and egg consumption during pregnancy and lactation. 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 IgE-mediated egg and/or peanut allergy at 12 months of age (i.e., the primary outcome) in a post-hoc analysis.

### 6.5 Harms

The number and percentage of mother-infant pairs experiencing an adverse event (AE) will be reported for each treatment group (irrespective of eligibility or compliance with the protocol) and compared across groups using Fisher exact tests. The following AEs will be evaluated:

- Maternal admission to intensive care unit (serious adverse event, 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)
- Infant anaphylaxis to egg or peanut (AE)

### 6.6 Analysis Software

All analyses will be performed using Stata v16 or later (College Station, TX: StataCorp LP).

## 7 REFERENCES

1. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549-56.
2. Zou G. A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*. 2004;159(7):702-6.
3. Palmer DJ, Sullivan TR, Gold MS, Prescott SL, Makrides M. Randomized controlled trial of early regular egg intake to prevent egg allergy. *The Journal of Allergy and Clinical Immunology*. 2017;139(5):1600-7.e2.
4. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine*. 2012;31(4):328-40.
5. Nurmatov U, Nwaru BI, Devereux G, Sheikh A. Confounding and effect modification in studies of diet and childhood asthma and allergies. *Allergy*. 2012;67(8):1041-59.
6. Pali-Scholl I, Jensen-Jarolim E. Gender aspects in food allergy. *Current opinion in Allergy and Clinical Immunology*. 2019;19(3):249-55.
7. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in Medicine*. 2004;23(16):2509-25.
8. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*. 2007;8(3):206-13.
9. 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;360:j5748.