Web Appendix A: Methods for handling missing data and results of sensitivity analyses

Of the 706 infants in the study, 681 (96.5%) attended their one year medical review and 666 (94.3%) had skin prick test results. Consequently, there was a small amount of missing data for outcome variables assessed at one year. In order to minimise bias due to missing data in the estimation of treatment effects, multiple imputation was used to create 50 complete datasets for analysis. Imputation was performed sequentially (baseline variables, post randomisation variables, outcomes) using the parametric regression method for continuous and count variables and the logistic regression method for binary and ordinal variables (Rubin, 1987). For each outcome, all variables pre-specified as predictors in the analysis model (i.e. treatment group, centre, parity, infant sex and maternal history of allergic disease) were included in the imputation model. Additional baseline variables and post randomisation variables potentially predictive of the outcomes were also added to the imputation models. Each imputation model was examined based on the estimated regression coefficients and the imputed values generated. Where appropriate, variables were omitted from the imputation models to avoid multicollinearity problems and to improve the overall quality of the model. All data were imputed using the mi procedure in SAS version 9.2. Following separate analyses of the 50 imputed datasets using standard methods for complete data, single estimates with valid standard errors were computed using the mi_analyze procedure in SAS version 9.2 according to Rubin's rules (Rubin, 1987).

Tables A1, A2 and A3 summarise the amount of missing data and the type of imputation model used for key baseline, key post randomisation and outcome variables respectively.

Table A1: Amount of missing data and imputation approach for key baseline variables

Baseline variable	*Variable type*	*N (%) missing*	*Imputation model* [a]
Treatment group	Binary	0 (0.0)	-
Parity zero	Binary	0 (0.0)	-
Infant sex male	Binary	0 (0.0)	-
Enrolment centre	Binary	0 (0.0)	-
Mother's age at trial entry (years)	Continuous	0 (0.0)	-
Mother Caucasian	Binary	0 (0.0)	-
Mother completed secondary education	Binary	0 (0.0)	-
Mother employed	Binary	0 (0.0)	-
Maternal smoking during pregnancy	Binary	0 (0.0)	-
Maternal smoking prior to pregnancy	Binary	0 (0.0)	-
Maternal history of allergic disease	Binary	0 (0.0)	-
Paternal history of allergic disease	Binary	17 (2.4)	Logistic regression

[a] ***Missing data were imputed using other key baseline variables.***

Table A2: Amount of missing data and imputation approach for key post randomisation variables

Post randomisation variable	*Variable type*	*N (%) missing*	*Imputation model* [a]
Infant born by caesarean delivery	Binary	0 (0.0)	-
Gestational age (weeks)	Continuous	0 (0.0)	-
Infant weight at birth (kg)	Continuous	0 (0.0)	-
Infant recumbent length at birth (cm)	Continuous	1 (0.1)	Linear regression
Infant head circumference at birth (cm)	Continuous	3 (0.4)	Linear regression
Other children in home at 6 months	Binary	11 (1.6)	Logistic regression
Infant breast fed in first 6 months	Binary	11 (1.6)	Logistic regression
Avoided foods during breastfeeding in first 6 months	Binary	11 (1.6)	Logistic regression
Infant fed formula in first 6 months	Binary	11 (1.6)	Logistic regression
Fish introduced to infant	Binary	9 (1.3)	Logistic regression
Nuts introduced to infant	Binary	12 (1.7)	Logistic regression
Egg introduced to infant	Binary	6 (0.9)	Logistic regression
Dog or cat around home	Binary	7 (1.0)	Logistic regression
Smoker in the house	Binary	13 (1.8)	Logistic regression
Free standing gas heater without chimney in home	Binary	15 (2.1)	Logistic regression
House dust mite protector or cover on mattress	Binary	7 (1.0)	Logistic regression
Regular contact with other children	Binary	6 (0.9)	Logistic regression

[a] ***Missing data were imputed using other key post randomisation variables and key baseline variables.***

Table A3: Amount of missing data and imputation approach for outcomes

Outcome variable [a]	*Variable type*	*N (%) missing*	*Imputation model* [b]
Allergic disease with sensitisation	Binary	40 (5.7)	Logistic regression
Allergic disease without sensitisation	Binary	25 (3.5)	Logistic regression
Sensitisation (incl. sensitisation by extract)	Binary	40 (5.7)	Logistic regression [c]
Sensitisation without allergic disease	Binary	37 (5.2)	Logistic regression
Eczema	Binary	22 (3.1)	Logistic regression
Eczema with sensitisation	Binary	25 (3.5)	Logistic regression
Parent reported doctor diagnosis of eczema	Binary	13 (1.8)	Logistic regression
Objective SCORAD [d]	Continuous	3 (2.8)	Linear regression
Confirmed food allergy with sensitisation	Binary	40 (5.7)	Logistic regression
Suspected food allergy with sensitisation	Binary	40 (5.7)	Logistic regression
Confirmed egg allergy	Binary	40 (5.7)	Logistic regression
Suspected egg allergy	Binary	40 (5.7)	Logistic regression
Confirmed or suspected egg allergy	Binary	40 (5.7)	Logistic regression
Frequent or persistent wheeze or asthma with sensitisation	Binary	20 (2.8)	Not imputed due to rarity of outcome
Allergic rhinitis with sensitisation	Binary	20 (2.8)	Not imputed due to rarity of outcome
Antibiotics use	Binary	12 (1.7)	Logistic regression
Ear antibiotics use	Binary	14 (2.0)	Logistic regression
Respiratory tract infections	Binary	23 (3.3)	Logistic regression
Admission to hospital	Binary	15 (2.1)	Logistic regression
Number of hospitalisations	Count	15 (2.1)	Linear regression
Visited doctor	Binary	8 (1.1)	Logistic regression
Number of doctor visits	Count	16 (2.3)	Linear regression
Serious adverse event	Binary	0 (0.0)	-

[a] Missing data for composite outcome measures (e.g. eczema with sensitisation) created by first imputing for individual components of outcome (e.g. sensitisation to individual extracts and eczema).

[b] Missing data were imputed using other outcomes, key post randomisation variables and key baseline variables.

[c] Sensitisation to individual extracts only imputed for egg and peanut. Sensitisation to other extracts was too rare for the imputation model to be reasonable.

[d] SCORAD only administered to the 108 infants with a current diagnosis of eczema at the one year medical review.

As illustrated in Table A3, the percentage of missing data was 5.7% or lower for all outcomes assessed at one year. Reasons for missing outcome data included death (1 infant), study withdrawal (3), loss to follow-up (21) and refusal or inability to complete the skin prick test (15). Overall the missing at random assumption imposed by the multiple imputation approach appeared reasonable for these data.

In addition to the primary imputed analysis, sensitivity analyses were performed on the original data (i.e. complete case analysis) and on imputed data created using different seeds and using different imputation models. All approaches produced similar results. Comparisons of the primary imputed results with complete case results are provided in Tables A4, A5 and A6 below.

Table A4: Comparison of imputed results with complete case analysis for binary outcomes

Outcome [a]	*Adjusted relative risk from imputed model (95% CI)* [b]	*Adjusted relative risk from complete case analysis (95% CI)* [b]
Allergic disease with sensitisation	0.70 (0.45 to 1.09)	0.69 (0.44 to 1.07)
Allergic disease without sensitisation	1.10 (0.79 to 1.55)	1.09 (0.77 to 1.54)
Sensitisation	0.75 (0.53 to 1.04)	0.75 (0.54 to 1.05)
Egg sensitisation	0.62 (0.41 to 0.93)	0.62 (0.41 to 0.93)
Peanut sensitisation	0.63 (0.34 to 1.19)	0.64 (0.34 to 1.20)
Sensitisation without allergic disease	0.83 (0.47 to 1.47)	0.87 (0.49 to 1.55)
Eczema	0.91 (0.70 to 1.17)	0.90 (0.70 to 1.17)
Eczema with sensitisation	0.64 (0.40 to 1.03)	0.63 (0.39 to 1.02)
Parent reported doctor diagnosis of eczema	0.93 (0.74 to 1.18)	0.93 (0.74 to 1.18)
Confirmed food allergy with sensitisation	0.96 (0.41 to 2.25)	0.93 (0.39 to 2.18)
Suspected food allergy with sensitisation	0.69 (0.36 to 1.31)	0.69 (0.36 to 1.31)
Confirmed egg allergy	0.79 (0.27 to 2.30)	0.78 (0.27 to 2.28)
Suspected egg allergy	0.47 (0.21 to 1.02)	0.46 (0.21 to 1.00)
Confirmed or suspected egg allergy	0.56 (0.30 to 1.04)	0.55 (0.29 to 1.02)

Outcome [a]	*Adjusted relative risk from imputed model (95% CI)* [b]	*Adjusted relative risk from complete case analysis (95% CI)* [b]
Antibiotics use	1.04 (0.91 to 1.18)	1.05 (0.92 to 1.19)
Ear antibiotics use	1.10 (0.86 to 1.41)	1.11 (0.87 to 1.42)
Respiratory tract infections	0.91 (0.67 to 1.23)	0.90 (0.66 to 1.23)
Admission to hospital	0.82 (0.61 to 1.11)	0.83 (0.62 to 1.12)
Visited doctor	1.00 (0.95 to 1.04)	0.99 (0.95 to 1.04)

[a] ***Results presented for only those binary outcomes that were imputed.***

[b] ***Relative risk n-3 LCPUFA vs control adjusted for centre, parity, infant sex and maternal history of allergic disease.***

Table A5: Comparison of imputed results with complete case analysis for continuous outcomes

Outcome	*Adjusted mean difference from imputed model (95% CI)* [b]	*Adjusted mean difference from complete case analysis (95% CI)* [b]
Objective SCORAD [a]	0.21 (-0.03 to 0.44)	0.21 (-0.03 to 0.45)

[a] ***Objective SCORAD scores were log transformed to better satisfy the assumptions of the linear regression model.***

[b] ***Mean difference n-3 LCPUFA vs control adjusted for centre, parity, infant sex and maternal history of allergic disease.***

Table A6: Comparison of imputed results with complete case analysis for count outcomes

Outcome	*Adjusted ratio of means from imputed model (95% CI)* [a]	*Adjusted ratio of means from complete case analysis (95% CI)* [a]
Number of hospitalisations	0.73 (0.51 to 1.04)	0.73 (0.51 to 1.05)
Number of doctor visits	1.06 (0.93 to 1.22)	1.07 (0.93 to 1.22)

[a] ***Ratio of means n-3 LCPUFA vs control adjusted for centre, parity, infant sex and maternal history of allergic disease.***

Web Appendix A reference list

Rubin D. *Multiple imputation for nonresponse in surveys.* New York: Wiley, 1987.