

Statistical Analysis Plan

Trial N°: DGC201911

ClinicalTrials.gov ID: NCT04599946

Effects of infant feeding with goat milk formula or cow milk formula on atopic dermatitis

Goat Infant Formula Feeding and Eczema: The GiraFFE study

Author: Laura Schlosser

Reviewed by: Dr. Veit Grote, MD MSc

Date: 20.02.2025

Version: Version log

N°: 1

Sponsor: Dairy Goat Co-operative (NZ) Ltd
18 Gallagher Drive, PO Box 1398 Hamilton 3240,
New Zealand

SIGNATURE PAGE

AUTHOR

Laura Schlosser

Signature 

Date 20.02.2025

REVIEWER

Dr. Veit Grote, MD MSc

Signature 

Date 24.02.2024

Chief Principal Investigator

Prof. Dr. med. Berthold Koletzko

Signature 

Date 24.02.2025

Sponsor

Dr. Elisabeth Weichselbaum

Signature 

Date 20/02/2025

INHALT

1	Introduction.....	7
1.1	Background and rationale.....	7
1.2	Study Objectives.....	7
1.2.1	Primary Objective.....	7
1.2.2	Secondary Objectives.....	7
1.3	Safety parameters.....	8
1.4	Modifications in the SAP in relation to the Current Protocol V1.1.....	9
1.5	Revision History for SAP Amendments.....	10
2	Investigational Plan.....	10
2.1	Trial Design.....	10
2.2	Study assessments and procedures.....	10
2.3	Eligibility.....	10
2.4	Randomization.....	13
2.5	Sample size.....	13
2.6	Statistical Hypothesis.....	13
3	Analysis population.....	14
3.1	The Full Analysis Set.....	14
3.2	The Per Protocol Data Set.....	14
4	Analysis variables.....	14
4.1	Demographic and Baseline Characteristics.....	14
4.2	Outcome variables.....	16
4.2.1	Primary outcome – atopic dermatitis.....	16
4.2.2	Secondary outcome variables.....	18
4.3	Other variables.....	20
4.3.1	Interruption and discontinuation of study formula consumption.....	20
4.3.2	Solid food intake before 4 months of age.....	21
4.3.3	Formula feeding before start of study formula.....	21
4.3.4	Cow’s milk allergy and intolerance.....	22
4.3.5	Filaggrin null mutations.....	22
4.4	Safety variables.....	22
4.4.1	Adverse events and serious adverse events.....	22
4.4.2	Laboratory Safety Variables.....	24
4.4.3	Anthropometry.....	24
5	Statistical Methods.....	24
5.1	General statistical methods.....	24
5.2	Confidence intervals and p-values.....	25

5.3	Demographic and Baseline Characteristics	25
5.4	Subject Disposition and Discontinuation	25
5.5	Withdrawal / follow-up	26
5.6	Protocol deviations	26
5.7	Subject's compliance.....	28
5.8	Observation time for incidence calculations.....	28
6	Analyses of Outcome Variables	29
6.1	Primary and secondary AD endpoints.....	29
6.2	Safety parameters	30
6.2.1	Adverse events	30
6.2.2	Blood parameters and anthropometry	31
6.2.3	Cow milk protein allergy (CMPA).....	31
6.2.4	Comparison of PP and ITT.....	32
7	Analysis methods	32
7.1	Missing data	33
7.2	General statistical principle for other secondary outcomes	33
7.3	Subgroup analysis	33
7.3.1	Early versus late randomization.....	33
7.3.2	Weeks of formula feeding before start of study formula.....	33
7.3.3	Study country.....	34
7.3.4	Parental atopic dermatitis and Filaggrin gene variants.....	34
8	Statistical interim analyses and stopping guidance	34
9	Unblinding.....	34
10	Statistical software.....	34
11	References.....	35

TABLES IN TEXT

Table 1: Summary table of secondary outcomes8

Table 2: Summary table of safety parameters9

Table 3: Summary table of time points for endpoint measures 12

Table 4: Summary table of patient's baseline characteristics 15

Table 5: AD-outcomes with respective definitions up to 12 months of age 17

Table 6: Drop-out reasons per intervention group 25

Table 7: Summary table of major and minor protocol deviations 27

Table 8: Descriptive table of new AD cases by visit..... 29

Table 9: List of reported AEs and SAEs 30

Table 10: Descriptive table for safety measures 31

Table 11: Comparison of CMPA among the intervention groups and diagnostic certainty measures for the diagnosis..... 32

Table 12: Incidence rate ratio for AD up to 12 months of age 32

ABBREVIATIONS

AD	Atopic Dermatitis
AD_any	AD_UKWP or AD_Dr (study personnel diagnosed, parental reported, or doctor diagnosed AD)
AD_Dr	Doctor diagnosed AD
AD_primary	Study personnel diagnosed AD, defined as meeting the UK Working Party Diagnostic Criteria for AD
AD_reported	AD diagnosis via telephone interview reported by parents fulfilling UK Working Party Diagnostic Criteria for AD
AD_UKWP	Study personnel diagnosed or parental reported (telephone interview) AD, defined as meeting the UK Working Party Diagnostic Criteria for AD
AE	Adverse Event
BISQ	Brief Infant Sleep Questionnaire
BMI	Body-Mass-Index
ES	Spain
HLGT	High Level Group Term
IGSQ	Infant Gastrointestinal Symptom Questionnaire
ITQOL	Infant Toddler Quality Of Life Questionnaire
ITT	Intention-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
PD	Protocol Deviation
PL	Poland
POEM	Patient Orientated Eczema Measure
PP	Per-Protocol
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan

SCORAD	SCORing Atopic Dermatitis
SOC	System Organ Class
SOPs	Standard Operating Procedures
UKWP	United Kingdom Working Party

1 Introduction

1.1 Background and rationale

This statistical analysis plan (SAP) details the proposed method when analysing the main results from the GIRAFFE study (ClinicalTrials.gov ID: NCT04599946). This analysis plan is meant to supplement the study protocol V1.1 from 25-06-2020 along with the published manuscript (Ferry et al. 2023). If differences occur between analyses described in the SAP and the current protocol, those found in this SAP will assume primacy. Any deviations from this plan will be described in the version log of the SAP. Emphasis of the SAP is the analysis of the primary outcome and principles of the analysis of secondary atopic dermatitis outcomes. Further secondary or exploratory analysis of data (e.g., blood biomarkers, metabolome, and gut microbiome) will be added in following versions.

This SAP is a dynamic document; modifications may continue to be made. Any changes will be documented in a version log.

1.2 Study Objectives

1.2.1 Primary Objective

The primary objective of this trial is the determination of the relative risk of developing Atopic dermatitis (AD) during the first 12 months of age in infants fed a study formula based on whole goat milk compared to infants fed a study formula based on cow milk protein. AD has to be diagnosed by study personnel at visits according to United Kingdom Working Party (UKWP) Diagnostic Criteria.

1.2.2 Secondary Objectives

While other AD diagnostic criteria are regarded as less accurate and treated as secondary outcomes, use of the UKWP Diagnostic Criteria only at fixed time points might diminish the sensitivity of the AD diagnosis. This is highlighting the importance of secondary endpoints such as parental reported diagnosis of AD, defined as meeting the UKWP Diagnostic Criteria for AD, or doctor diagnosed AD. Also, other time points than 12 months of age, which are primarily determined by the end of the intervention, are considered as secondary outcomes. Table 1 shows secondary endpoints of AD, AD severity, AD treatment and other atopic diseases with respective time points. Furthermore, general well-being defined by gastrointestinal problems, sleep and quality of life are considered.

Table 1: Summary table of secondary outcomes

Secondary outcomes	Time frame (age)
Study personnel diagnosed AD, defined as meeting the UKWP Diagnostic Criteria for AD (AD_primary)	up to 24 and 60 months
Study personnel diagnosed or parental reported (telephone interview) AD, defined as meeting the UKWP Diagnostic Criteria for AD (AD_UKWP)	up to 12, 24 and 60 months
Doctor diagnosed AD (AD_Dr)	up to 12, 24 and 60 months
Any AD (AD_UKWP or AD_Dr) (AD_any)	up to 12, 24 and 60 months
Age at first diagnosed AD_UKWP	up to 12, 24 and 60 months
Age at first diagnosed AD_any	up to 12, 24 and 60 months
AD severity expressed with the SCORAD in AD_UKWP ^a	up to 12, 24 and 60 months
AD severity expressed with POEM in AD_UKWP ^a	up to 12, 24 and 60 months
Use of eczema-related medication or skin care in AD_any ^b	up to 12, 24 and 60 months
Parental report of a doctor diagnosed diagnosis of food allergy of the child	12, 24 and 60 months
Parental reported doctor diagnosed hay fever, asthma and asthma-related diseases (wheezing and allergic rhinitis) of the child	up to 60 months
Parental report of gastrointestinal symptoms (Infant Gastrointestinal Symptom Questionnaire –IGSQ 13) and sleep (Brief Infant Sleep Questionnaire –BISQ)	at 4, 6 and 12 months
Parental reported Quality of life in children using the Infant Toddler Quality Of Life Questionnaire (ITQOL-SF47)	at 4, 12, 24 and 60 months

^a AD severity in children will be assessed using the highest SCORAD/POEM-value, as applicable, ever recorded.

^b Any topical anti-inflammatory therapy [binary outcome: Yes / No] or systemic therapy for AD.

1.3 Safety parameters

Safety parameters encompass the listed range of assessments to monitor children's well-being throughout the study. Key safety parameters include Adverse Events (AEs), Serious Adverse Events (SAEs), used medication, full blood count and anthropometric measures.

Table 2: Summary table of safety parameters

Safety parameters	Time frame (age)
Risk of AEs and SAEs	Anytime during the intervention phase (until the visit at 12 months of age) and up to 30 days after the last study product consumption
Use of medications	see AEs/SAEs in 4.4.1
Full blood count	at 4 and 12 months of age
Anthropometric measures (weight-for-age-z-scores, length-for-age-z-scores and BMI-for-age z-scores)	at 4, 6, 12 months of age

1.4 Modifications in the SAP in relation to the Current Protocol V1.1

The protocol is defining cumulative incidence as the main analysis unit of AD. However, due to the nature of the study with several fixed time points used for AD assessment before the primary endpoint at 12 months and later endpoints at 24 and 60 months of age, and, additionally, the inherent problem of drop-outs and, hence, the change of the population at risk to develop AD, we define cumulative incidence as cumulative incidence rates over specific periods. Details are described below in 6.1.

Anthropometry was mentioned as secondary outcome in the protocol. However, it was also used as the main safety parameter defined in the protocol and for the Data and Safety Monitoring Board meeting. It also defined the number of subjects of the interim safety analysis. Therefore, we included anthropometric data exclusively as a safety variable for this SAP.

We will not perform any centre subgroup analyses for any AD or related outcome.

In contrast to the protocol, we added cumulative incidence rate of AD_{any}, as a secondary endpoint. Furthermore, in the protocol, doctor diagnosed AD (AD_{Dr}) and parental reported AD (AD_{reported}) had been combined into “reported AD”. We separated these two categories for the analysis, first, as AD_{Dr} is based on a standard question used in many other studies. Second, AD_{reported} was combined with the study personnel diagnosed AD (AD_{primary}) in one group as a secondary outcome (AD_{UKWP}) as both are based on the same standard in contrast to AD_{Dr} having no standardization.

According to the protocol the primary statistical method to be used for the primary outcome analysis was a Generalized Estimating Equation Poisson model with a log link and robust standard error by sandwich estimators of variance. This model was chosen as it was recommended by Pedroza and Truong (Pedroza and Truong 2017) to be the best performing model in randomized, multicentre trials with a small number of centres. Re-evaluation of the manuscript, revealed that, first, the number of subjects in each centre in the simulation study

was smaller than in our study and, second, that with higher numbers of subjects (>50) the methodological advantage vanished. Therefore, we decided for a more flexible generalized linear mixed model using a Poisson distribution with centre as random effect.

$\log(E(Y_{ij}))=X_{ij}\beta+u_j$ - where $E(Y_{ij})=X_{ij}\beta$ represents the expected value (the average rate) of the count variable.

- Y_{ij} is the count variable for the i^{th} observation in the j^{th} group (e.g., centre).
- λ_{ij} is the expected value (the mean number of events) for the i^{th} observation in the j^{th} group.
- X_{ij} is a vector of predictors (fixed effects, e.g., study formula) for the i^{th} observation in the j^{th} group.
- β is a vector of coefficients for the fixed effects.
- u_j is the random effect for the j^{th} site, with $u_j \sim N(0, \sigma_u^2)$ (normally distributed with a mean of 0 and variance σ_u^2).
- The log link is used to transform the linear predictor $X_{ij}\beta+u_j$ into the Poisson distribution.

1.5 Revision History for SAP Amendments

None.

2 Investigational Plan

2.1 Trial Design

The study is designed as a two-arm, parallel, randomized, double-blind, controlled nutritional trial. Further details can be found in the study protocol (Ferry et al. 2023).

2.2 Study assessments and procedures

Collection of data within the trial for outcome assessments will be conducted via interviews and physical examinations during face-to-face study visits or via telephone calls (Table 3). All face-to-face study visits, physical examinations, bio sample collections and telephone calls will be performed by trained personnel. Detailed description of conduct of assessments is described in the protocol mentioned in 1.1.

2.3 Eligibility

Children can be enrolled in the study if they fulfil the following inclusion criteria:

- Having obtained written informed consent (signed and dated) of the child's parent(s)/caregiver(s), indicating that the child's parent(s)/caregiver(s) has/have been informed of all pertinent aspects of the study
- Infant was born full term (≥ 37 weeks +0 days and ≤ 41 weeks +6 days of gestation)
- Age at enrolment < 3 months of age (< 90 days)
- Infant birth weight ≥ 2.5 kg and ≤ 4.5 kg
- Infant is born from a singleton pregnancy

- Child's parent(s)/caregiver(s) is/are of legal age of consent
- The child's parent(s)/caregiver(s) have sufficient local language skills to understand the study information, the informed consent, and to comply with the study procedure
- The child's parent(s)/caregiver(s) is/are willing and deemed able to fulfil the requirements of the study protocol and procedures
- Mother has expressed the intention to partially (in combination with breastfeeding) or fully formula-feed

Infants with one or more of the following criteria are excluded from enrolment in the study:

- Diagnosed disorder considered to interfere with nutrition, growth or development of the immune system
- Participation of the child in any other interventional trial or participation of the mother in any intervention trial with child follow-up
- Infant has a doctor's diagnosis of AD or a severe widespread skin condition prior to randomization that would make the detection or assessment of AD difficult
- Infant has consumed an infant formula (on average at least 3-4 days a week, at least one bottle per day) for more than 4 weeks prior to enrolment
- Cow milk allergy or intolerance of the child
- Institutionalized infant

Table 3: Summary table of time points for endpoint measures

Age of infant (months)	≤3 (study entry)	1 (± 1 wk)	2 (± 1 wk)	3 (± 1 wk)	4 (± 1 wk)	6 (± 1 wk)	8 (± 1 wk)	10 (± 1 wk)	12 (± 2 wk)	18 (± 2 wk)	24 (± 2 wk)	36 (± 2 wk)	48 (± 2 wk)	60 (± 2 wk)
Visit (V)/ phone call (T)	V0	T0	T2	T3	V4	V6	T8	T10	V12	T18	V24	T36	T48	V60
Eligibility criteria, Informed consent	X													
Randomisation	X													
Baseline and background data	X								X		X			X
Formula Supply	X ^a				X ^a	X ^a	X ^a	X ^a						
Blood sample					X				X					X
Stool sample (Subsample)					X				X					X
AD (UK working party, modified)	X				X	X	X ^c	X ^c	X	X ^c	X	X ^c	X ^c	X
SCORAD ^b					X	X			X		X			X
POEM ^b ,					X	X	X	X	X	X	X	X	X	X
AD treatment/skin care	X				X	X	X	X	X	X	X	X	X	X
Allergy symptoms					X	X			X		X			X
Feeding/ Dietary intake					X	X	X	X	X		X			X
Brief Infant Sleep questionnaire (BISQ)	X				X	X			X					
Infant gastrointestinal symptoms Questionnaire (IGSQ)	X				X	X			X					
Infant Toddler Quality of life Questionnaire™ (ITQOL)					X				X		X			X
Anthropometry	X				X	X			X		X			X
Compliance		X	X	X	X	X	X	X	X					
Health assessment (including adverse events and medication)		X	X	X	X	X	X	X	X	X	X	X	X	X

^a performed by the logistic company; Frequency of shipments depends on the maximal acceptable amount per delivery

^b only if AD criteria fulfilled (UK working party or reported)

^c UKWP diagnosis criteria are answered by telephone interview

2.4 Randomization

Minimisation randomization (White and Freedman 1978, Scott et al. 2002) was performed with centre as the only strata and 4 randomization codes for the two study formulas. Imbalances in age at randomization and sex were minimized by using continuous minimization factors. We also used a random element to make assignment unpredictable with a maximum permissible difference between the treatment groups of +/- 6.

2.5 Sample size

The number of subjects to be studied depends mainly on (a) the incidence of AD in the population and (b) the effect size to be detected as significant.

a) Reported AD incidence estimates for young children in European countries vary, e.g., 13% for Spain (Draaisma et al. 2015) and 17% for Poland (Kamer et al. 2013).

b) Meta-analyses of effects of protein hydrolysate formulas point to a potential reduction of AD by 40-50% compared to standard formula (Alexander and Cabana 2010, Iskedjian et al. 2010). A systematic review of studies testing the corresponding effect of prebiotics in infant formulas found a risk reduction by 32%, although not statistically significant (Cuellar-Garcia et al. 2017). Specifically, an Australian trial indicated a risk reduction for AD incidence of 30% comparing goat milk based to cow milk-based formulas (Zhou et al. 2014).

We assume a cumulative incidence of AD at 15% in the first 12 months of age, based on the cited data, and a 30% risk reduction by goat milk formula. Setting significance level to 0.05 and statistical power to 80% a sample size of 861 infants per group is required using a two-sample proportions test (Pearson's chi-squared test).

Initially, a drop-out rate of 25% was assumed and it was aimed to recruit 1148 infants per group (in total 2296). As it was realized in the course of the study that the drop-out rate was lower than expected, a re-assessment was performed by the principal investigator and the sponsor, according to the protocol. It was decided in April 2023 to finish the recruitment process when at least 2100 children for both groups were randomized, assuming a drop-out rate of 18%.

2.6 Statistical Hypothesis

For the comparison of the goat milk infant formula intervention group to cow milk infant formula intervention group, the following hypothesis of the primary endpoint will be tested, where IR_{AD} is the incidence rate of participants with study personnel diagnosis of AD (AD_{primary}) up to 12 months of age.

Null hypothesis (H_0): $IR_{AD_cow} / IR_{AD_goat}$: The incidence rate ratio to develop AD in the first 12 months of age for infants fed cow milk or goat milk formula equals one.

3 Analysis population

3.1 The Full Analysis Set

The Full Analysis Set includes all children, where the informed parental consent was obtained, which were randomized and who have consumed at least one feed of study product. This will be considered the intention-to-treat population (ITT).

3.2 The Per Protocol Data Set

The Per-Protocol (PP) data set includes all children of the ITT, who are additionally fulfilling compliance criteria (see 5.7) and attend all face-to-face visits up to 12 months of age, unless AD_primary occurred before any missing visit.

All outcome and safety analyses will be performed in ITT. Furthermore, for visits at 24 and 60 months of age, AD_primary and AD_any outcome analyses will be also performed in the PP group.

4 Analysis variables

4.1 Demographic and Baseline Characteristics

The following demographic and baseline characteristics will be summarized:

Age at enrolment (days), Sex (Male, Female), Birth weight (grams), Recumbent length (centimetres), Mode of delivery (Caesarean section, Vaginal birth), Gestational age (weeks), Birth order (calculated by number of life births: 1st child, 2nd child, > 2nd child), Antibiotic use during delivery by the mother (Yes, No), Previous feeding with other formula (Yes, No), Parental demographic variables: Educational level (0=Early childhood education, 1=Primary education, 2=Lower secondary education, 3=Upper secondary education, 4=Post-secondary non-tertiary education, 5=Short-cycle tertiary education, 6=Bachelor's or equivalent, 7=Master's or equivalent, 8=Doctoral or equivalent with division into low [0-2], medium [3-4], high [5-8]), Household income (low = ES: < 1900€ / PL: < 900€, medium = ES: 1900€-3550€ / PL: 900€-1450€, high = ES: > 3550€ / PL: > 1450€), Country of birth (Both study country, At least one other county), Maternal age (years), Paternal age (years), Maternal Body-Mass-Index (BMI) (based on pre-pregnancy weight), Paternal BMI (BMI will use weight/height² formula), Parental history of allergic diseases (One parent AD, Both parents AD, One other, Both other), Any sibling(s) history of allergic disease (AD, Other), Maternal smoking during pregnancy (Yes, No), Maternal currently smoking habits (Yes, No), Paternal currently smoking habits (Yes, No).

Table 4: Summary table of patient's baseline characteristics

	Cow	Goat
	n (%) / Mean (SD)*	n (%) / Mean (SD)*
Age at enrolment (days)	XX	XX
Child's sex		
Male	XX	XX
Female	XX	XX
Birth weight (g)	XX	XX
Recumbent length (cm)	XX	XX
Mode of delivery		
Caesarean section	XX	XX
Vaginal Birth	XX	XX
Gestational age (weeks)	XX	XX
Birth order		
1st child	XX	XX
2nd child	XX	XX
> 2nd child	XX	XX
Antibiotic use during delivery	XX	XX
Feeding with other formula before inclusion	XX	XX
Family background Information		
Education level		
Mother		
Low	XX	XX
Medium	XX	XX
High	XX	XX
Father		
Low	XX	XX
Medium	XX	XX
High	XX	XX
Household Income		
Low	XX	XX
Medium	XX	XX
High	XX	XX
Country of birth		
Both study country	XX	XX
At least one other country	XX	XX
Family health Information		
	Cow	Goat
	n (%) / Mean (SD)*	n (%) / Mean (SD)*
Maternal age (years)	XX	XX
Paternal age (years)	XX	XX
Maternal BMI (kg/m²)	XX	XX
Paternal BMI (kg/m²)	XX	XX

Parental history of allergic diseases		
One parent AD	xx	xx
Both parents AD	xx	xx
One other ^a	xx	xx
Both other ^a	xx	xx
Any Sibling(s) history of allergic diseases		
AD	xx	xx
Other ^a	xx	xx
Maternal smoking during pregnancy	xx	xx
Mother current smoking	xx	xx
Father current smoking	xx	xx

* as appropriate

^a *Other* allergic disease is defined as asthma, allergic rhinitis, or food allergy.

4.2 Outcome variables

4.2.1 Primary outcome – atopic dermatitis

The UKWP Diagnostic criteria for atopic dermatitis (modified for children in the first year of life) are satisfied where there is:

- An itchy skin condition in the last 12 months

plus three or more of:

- History of involvement of cheeks or outer arms/legs
- History of general dry skin
- Family history of atopic diseases in the first degree relatives
- Visible dermatitis on the cheeks or outer arms/legs with absence of axillary involvement

Furthermore, the diagnosis has to be made during a face-to-face visit, i.e., by study personnel.

Table 5: AD-outcomes with respective definitions up to 12 months of age

Entry question				UKWP assessment in a face-to-face visit		
				Yes	No	
Since birth, has your child had an itchy skin condition – by itchy we mean scratching or rubbing the skin?	If the question was answered with "Yes", the following four questions were asked.	Has the skin condition ever affected the cheeks or outer arms and legs? (Yes/No)	Three or more questions were answered with "Yes": AD diagnosis according to UKWP criteria.	AD_primary	AD_reported	AD_any (AD_UKWP or AD_Dr)
		Has your child suffered from a dry skin in general since birth? (Yes/No)				
		Does anyone in your child's immediate family (i.e., mother, father, brother, sister) suffer from AD, hay fever, or asthma? (Yes/No)		AD_UKWP (AD_primary or AD_reported)		
		At visits: Inspection for visible flexural dermatitis on child's cheeks and outer arms and legs. (Yes/No) OR At phone calls: Does the child have a visible dermatitis in any of these body areas today? – on the cheeks and outer arms and legs. (Yes/No)				
Was your child ever diagnosed by a doctor to have atopic dermatitis? (Yes/No)		Yes		AD_Dr		

4.2.2 Secondary outcome variables

4.2.2.1 Atopic dermatitis

If the Atopic dermatitis diagnosis criteria according to UKWP are fulfilled at any time but assessment was not via a face-to-face visit, AD is considered a secondary outcome. Furthermore, AD that was diagnosed by a doctor is considered a secondary outcome. Lastly, any AD diagnosis is given to a subject if either UKWP AD (face-to-face or reported) or doctor diagnosed AD is fulfilled.

4.2.2.2 SCORing Atopic Dermatitis (SCORAD)

The SCORAD is a validated tool used in clinical research and clinical practice that was developed to standardize the evaluation of the extent and severity of AD (Stalder et al. 1993). The SCORAD is evaluated from the first diagnosis of AD (AD_UKWP) and at each subsequent visit. The extent of AD is assessed by the investigator as a percentage of each defined body area and reported as the sum of all areas, with a maximum score of 100% (assigned as “A” in the overall SCORAD calculation). The severity of 6 specific symptoms (erythema, oedema / papulation, excoriations, lichenification, oozing / crusts and dryness) of AD is assessed by the investigator using the following scale: none (0), mild (1), moderate (2), or severe (3) (for a maximum of 18 total points, assigned as “B” in the overall SCORAD calculation). Subjective assessment of itch and sleeplessness is recorded for each symptom by the patient or relative on a visual analogue scale, where 0 is no itch (or sleeplessness) and 10 is the worst imaginable itch (or sleeplessness), with a maximum possible score of 20. This parameter is assigned as “C” in the overall SCORAD calculation. The SCORAD is calculated as $A/5 + 7B/2 + C$. The maximum SCORAD score is 103. The objective SCORAD is calculated as $A/5 + 7B/2$. The maximum objective SCORAD score is 83. A higher score indicates worse condition. The SCORAD is only assessed at visits, once an AD diagnosis was made (study personnel or parental reported).

4.2.2.3 Patient Oriented Eczema Measure (POEM)

The POEM is a 7-item, validated questionnaire used in clinical practice and clinical trials to assess disease symptoms in children and adults with atopic eczema (Charman, Venn, and Williams 2004). The POEM is assessed from the first diagnosis of AD (AD_UKWP) and at each subsequent visit or telephone interview. The format is patient response to 7 items (dryness, itching, flaking, cracking, sleep loss, bleeding, and weeping) based on symptom frequency during the past week (i.e., 0 = ‘no days’, 1 = ‘1 to 2 days’, 2 = ‘3 to 4 days’, 3 = ‘5 to 6 days’, and 4 = ‘every day’). The total score is the sum of the 7 items which is ranged from 0 to 28; a high score is indicative of a poor quality of life. The following POEM banding scores have been established (Charman, Venn, and Williams 2004): 0 to 2 = Clear or almost clear; 3 to 7 = Mild eczema; 8 to 16 = Moderate eczema; 17 to 24 = Severe eczema; 25 to 28 = Very severe eczema. If two or more response options are selected for a question, then the response option with the highest score is recorded. If one question of the seven is left unanswered, then that question is scored as 0 and the scores are summed and expressed as usual out of a maximum of 28. If two or more questions are left unanswered, then the questionnaire is not scored and is set to missing. If two or more response options are ticked for one question, the response option with the highest score should be recorded. A higher score indicates worse condition. This questionnaire asks the caregiver to report their perception of the patient’s AD symptoms. It is therefore completed by the caregiver. It is

recommended that the same person complete the questionnaire on behalf of the patient throughout the study.

The POEM is assessed at visits and telephone interviews, once an AD diagnosis was made (study personnel or parental reported).

4.2.2.4 Infant Quality of Life Questionnaire (ITQOL)

ITQOL-SF47 is a simple, parent-completed, 47 items, validated, Infant and Toddler Quality of Life Questionnaire for use in clinical care and research. This shorter form (SF47) includes several scales (dimensions) that measure different aspects of quality of life, such as physical health, pain, sleep, development, and social functioning (Landgraf et al. 2013). The ITQOL questionnaire is assessed at 4, 12, 24, and 60 months of age. Recoding and recalibrating of all items will be done according to the instructions from ITQOL-SF47 scoring rules (confidential). Recoding ensures that all items are positively scored so that a higher score indicates better health. Recalibrating is performed for the two items with an “excellent to poor” response continuum to achieve a better linear fit with corresponding scales and to provide a better approximation of equal interval scaling. It is recommended that scale scores be calculated if half or more of the items within a scale have been answered or half plus one in the case of scales with an odd number of items. Next step is to compute raw scores by calculating the mean of items, followed by a transformation of the raw scores to standardized 0 to 100 scores.

4.2.2.5 Brief Infant Sleep Questionnaire (BISQ)

The Brief Infant Sleep Questionnaire assesses sleep patterns and behaviours in infants and young children (Sadeh 2004). Parents report on their child’s sleep behaviour over the past week. The BISQ is assessed at study entry, 4, 6, and 12 months of age.

The BISQ focuses on key areas of sleep, including:

- nocturnal sleep duration (between the hours of 7pm and 7am)
- daytime sleep duration (between the hours of 7am and 7pm)
- number of night wakings
- duration of wakefulness during the night hours (10pm to 6am)
- nocturnal sleep-onset time
- sleeping time (latency to falling asleep for the night)
- method of falling asleep (1 = While feeding; 2 = Being rocked; 3 = Being held; 4 = In bed alone; 5 = In bed near parents)
- location of sleep (1 = Infant crib in a separate room; 2 = Infant crib in parents’ room; 3 = In parents’ bed; 4 = Infant crib in room with sibling(s); 5 = Other)
- preferred body position (1 = On his/her belly; 2 = On his/her side; 3 = On his/her back)

Higher values in night wakings and longer sleep latency might suggest potential sleep issues, while higher total sleep duration and parental ratings of good quality indicate healthy sleep patterns.

4.2.2.6 Infant Gastrointestinal Symptom Questionnaire (IGSQ)

The Infant Gastrointestinal Symptom Questionnaire is a tool to assess feeding tolerance in infants (Riley et al. 2015). The IGSQ is assessed at study entry, 4, 6, and 12 months of age. This parent reported questionnaire on infant digestion and elimination behaviour over the previous week (7 days) includes 5 symptom clusters and results in a 13-items index. The five

symptom clusters are:

• **Stooling** (A: How many times did your baby pass a hard stool? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 to 3 times in the week; 4 = 4 to 6 times in the week; 5 = 7 or more times in the week; 0 = Don't know or no response; B: How many times did your baby have difficulty when passing a bowel movement? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 to 3 times in the week; 4 = 4 to 6 times in the week; 5 = 7 or more times in the week; 0 = Don't know or no response)

• **Spitting-up / Vomiting** (A: How many times did milk come out of your baby's mouth? 1 = 0 times in a usual day; 2 = 1 time in a usual day; 3 = 2 to 3 times in a usual day; 4 = 4 to 6 times in a usual day; 5 = 7 or more times in a usual day; 0 = Don't know or no response; B: How much milk usually came out each time? 1 = 5ml; 2 = 15ml; 3 = 30ml; 4 = About half the feeding; 5 = More than half the feeding; 0 = Don't know or no response; C: How often did your baby seem uncomfortable or fussy when milk came out of his or her mouth? 1 = Never; 2 = Almost never; 3 = Sometimes; 4 = Almost always; 5 = Always; 0 = Don't know or no response; D: How many times did your baby arch his or her back as if in pain when milk came out of his or her mouth? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 to 3 times in the week; 4 = 4 to 6 times in the week; 5 = 7 or more times in the week; 0 = Don't know or no response)

• **Crying** (A: How much total time did your baby usually cry in a day? 1 = Less than 10 minutes in a day; 2 = 10 to 30 minutes in a day; 3 = 30 minutes to 1 hour in a day; 4 = 1 to 2 hours in a day; 5 = 2 or more hours in a day; 0 = Don't know or no response; B: How many times were you unable to soothe your baby to stop his or her crying? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 times in the week; 4 = 3 times in the week; 5 = 4 or more times in the week; 0 = Don't know or no response; C: How many times did your baby cry during or right after a feeding because the milk bothered the baby? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 times in the week; 4 = 3 times in the week; 5 = 4 or more times in the week; 0 = Don't know or no response)

• **Flatulence** (A: How many times in usual day was your baby fussy? 1 = 0 times in a usual day; 2 = 1 time in a usual day; 3 = 2 to 3 times in a usual day; 4 = 4 to 6 times in a usual day; 5 = 7 or more times in a usual day; 0 = Don't know or no response; B: How often did gas seem to make your baby uncomfortable or fussy? 1 = Never; 2 = Almost never; 3 = Sometimes; 4 = Almost always; 5 = Always; 0 = Don't know or no response)

• **Fussiness** (A: On how many days was your baby fussy? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 times in the week; 4 = 3 times in the week; 5 = 4 or more times in the week; 0 = Don't know or no response; B: How many times were you unable to soothe your baby when she or he was fussy? 1 = 0 times in the week; 2 = 1 time in the week; 3 = 2 times in the week; 4 = 3 times in the week; 5 = 4 or more times in the week; 0 = Don't know or no response)

Scores for each item are summed to create a total gastrointestinal symptom score which will be used for analysis.

4.3 Other variables

4.3.1 Interruption and discontinuation of study formula consumption

Information about interruption of study formula consumption was collected at every time point (T0, T2, T3, V4, V6, T8, T10, and V12) during the intervention phase. The parents were

asked: Has the child interrupted study milk consumption for more than 3 consecutive days (or completely stopped) since the last contact? (yes/no). If this question was answered with yes, it was asked: The consumption of the study product was completely stopped? (yes/no).

If study formula was discontinued, the last date of consumption was recorded. Additionally, the parents were asked to indicate the reason(s) for discontinuing the study formula, selecting all that apply from the following options:

- Change of dietary preferences (other formula or milk)
- Medical problem
- Not enough product available
- Logistic reasons (e.g., travel, extraordinary situations)
- Obvious dislike of product by child
- An allergic reaction or intolerance
- Too much gas
- Constipation
- Too much spit up
- Diarrhoea
- Vomiting
- Belly pain
- Too much mucus
- Concern about lack of growth/ lack of weight gain
- Other concerns/ reasons

Medical problems were recorded as AE.

4.3.2 Solid food intake before 4 months of age

At 4 months of age, it was asked: Did you introduce regular solid foods before the age of 4 months? Regular means taking four or more teaspoons (i.e., 20 g) of complementary foods per day (yes/no).

4.3.3 Formula feeding before start of study formula

At baseline it was asked in a matrix format if formula was fed before inclusion with a row identifying the type of formula and columns with ticks for each week of life:

Which brand of formula did you feed in which week of age?

A variable of any formula consumption before inclusion (yes/no) and a second variable including the total weeks of formula consumption before inclusion (each ticked week signifies one week) will be constructed.

Furthermore, it was asked when the study formula was started to be fed to the child. If the child received formula at inclusion into the study, complete weeks (seven days) between study start and start date of formula feeding are considered as weeks of formula consumption.

Combining the information of formula feeding before inclusion and start of formula consumption will be used to define a variable of more than 4 weeks of formula consumption before feeding study formula.

4.3.4 Cow's milk allergy and intolerance

At each visit the question of an existing health issue was raised to each participant. If answered with "yes" following question is asked: Was your child diagnosed or tested (including (study) milk interruption) for cow's milk allergy or (study product) intolerance since the last planned study contact? If yes the following questions were used to clarify the diagnostic certainty of a milk allergy diagnosis:

- Was your child diagnosed by a physician to have cow's milk allergy or intolerance? (yes/no).
- Was the (study) milk intake interrupted and restarted to see if problems disappeared and re-appeared? (yes/no).
- Blood test for specific Immunoglobulin E (IgE) performed (yes/no)
- Skin prick test performed (yes/no)
- Standardized provocation test performed (yes/no).

4.3.5 Filaggrin null mutations

Following Filaggrin gene variants will be analysed: R501X, 2282del4, R2447X, and S3247X. The combined genotype is defined as the presence of at least one of the four considered variants and will be used for the subgroup analysis as presence of a Filaggrin risk variant (see 7.3.4).

4.4 Safety variables

4.4.1 Adverse events and serious adverse events

Adverse events (AEs) and serious adverse events (SAEs) were collected starting from the time of informed consent signature and at each contact until the end of the study. The criteria for determining whether a health problem should be reported as an AE were as follows:

- The child was treated with medication > 14 days
- The child was treated with oral antibiotics
- The child was treated with inhalation therapy
- The child was treated with steroids, salbutamol, antihistamines, montelukast
- The child was hospitalized
- The child was treated with a special diet for more than 7 days
- The child interrupted the intake of the study product for more than one day or completely discontinued consumption

Pre-existing diseases or conditions will not be considered AEs unless there is a worsening of the condition, or a change in the quality. Furthermore, AEs after 12 months of age were only considered up to 30 days after the last intervention formula consumption. There was no instruction/standard operating procedure (SOP) for the study personnel to report multiple AEs during a single event period as separate health problems, e.g., conjunctivitis accompanying an episode of bronchitis, if the AE was already concluded at the visit. For

ongoing AEs any changes in diagnosis, treatment or impact on feeding were documented by entering the end date for the health problem with previous diagnosis, therapy or diet and creating a new AE for the health problem with new diagnosis, therapy or diet and enter the new start date.

Reported adverse events were classified according to the Medical Dictionary for Regulatory Activities (MedDRA) Version 27.1. AEs were coded to a "Preferred Term" (PT), and were associated with "High Level Group Terms" (HLGTs) and the primary "System Organ Class" (SOC). If multiple AEs were given in one health problem, the medically most relevant or more general health problem was classified. Multiple occurrences of AEs of the same PT (or SOC) in the same subject will be counted only once for that PT (or SOC). The denominator for computation of percentage is the number of patients in each treatment group for the corresponding analysis period.

Following SOC exist in the MedDRA:

- Blood and lymphatic system disorders
- Cardiac disorders
- Congenital, familial and genetic disorders
- Ear and labyrinth disorders
- Eye disorders
- Gastrointestinal disorders
- General disorders and administration site conditions
- Hepatobiliary disorders
- Immune system disorders
- Infections and infestations
- Injury, poisoning and procedural complications
- Investigations
- Metabolism and nutrition disorders
- Musculoskeletal and connective tissue disorders
- Neoplasms benign, malignant and unspecified (incl. cysts and polyps)
- Nervous system disorders
- Pregnancy, puerperium and perinatal conditions
- Psychiatric disorders
- Renal and urinary disorders
- Reproductive system and breast disorders
- Respiratory, thoracic and mediastinal disorders
- Skin and subcutaneous tissue disorders
- Surgical and medical procedures
- Vascular disorders

Start and end date of the AE was recorded together with any relevant treatment according to above criteria. Inhalations therapy (separated by sodium chloride, steroid, adrenalin and spasmolytics) and study product interruption and discontinuation was explicitly asked via tick boxes, while any other treatment was entered as free text. All free text medications were manually coded into:

- Steroid i.a.
- Steroid i.v./p.o.
- Steroid skin
- Antibiotic i.v./p.o.
- Antihistamines
- Spasmolytics
- Montelukast
- Antibiotic skin, eye, nose
- Antifungal i.v./p.o.
- Antifungal skin
- Pro-/prebiotic
- Special diet
- Steroid, nasal
- Adrenaline, i.a.
- Steroid, eye
- Antihistamine, eye
- Antiviral treatment
- Other medication

4.4.2 Laboratory Safety Variables

Haemoglobin and eosinophils of the differential blood count are used. Abnormal findings are defined with

Anaemia: haemoglobin < 10.5 g/dl;

Eosinophilia: eosinophils > 1000 /L.

4.4.3 Anthropometry

Length, height and weight will be measured at time points according to Table 3. BMI will be calculated by weight in kg divided by length/height in cm squared. All anthropometric values will be expressed in z-scores according to the WHO growth standards (Onis 2006). Linear growth, weight gain between randomization and the visit at 12 months of age will be analysed using regression analysis adjusting for respective anthropometric measure at baseline. In line with the primary outcome analysis, we will use centre as a random effect. Weight-for-age z-score at 12 months of age is the primary safety measure.

5 Statistical Methods

5.1 General statistical methods

Categorical or ordinal variables will be summarized using frequencies and percentages for each category. For categorical variables, a Chi-square test will be utilized to assess whether there is a significant association or difference between each category. Continuous variables will be summarized by the number of participants reflected in the calculation (n), mean and standard deviation in case of normality, otherwise median and interquartile range (25. and 75

percent). To inspect for data normality, histograms will be used. The number of observations for each variable will be reported.

The frequency of dropouts will be compared between the goat milk intervention- and cow milk intervention-group using Chi-square-test or Fisher-exact-test if drop-out rates are low.

5.2 Confidence intervals and p-values

All statistical tests will be two-sided and conducted at a 5% significance level. Confidence intervals (CI) will be two-sided and set at 95%. For all estimated parameters CIs will be reported.

5.3 Demographic and Baseline Characteristics

Demographics and baseline characteristics will be summarized by intervention group (i.e., goat milk infant formula and cow milk infant formula) with descriptive statistics using the ITT. The list of demographic and baseline characteristics to be summarized will include the listed variables in Table 4.

The baseline characteristics will be presented by treatment group. Demographic and baseline characteristics will be summarized by intervention group for the ITT.

To assess imbalances in the randomization, baseline comparisons between the treatment groups will be performed for all variables listed in Table 4 using t-test for continuous variables and Chi-square-test for nominal and categorical variables.

5.4 Subject Disposition and Discontinuation

The following summaries will be provided for each intervention group and the study in total (unless otherwise specified):

- The total number of screened participants who signed Informed Consent Form (ICF)
- The total number of randomized participants (children, who received a randomization number)
- The total number of participants in each analysis set
- The total number of participants who completed the study up to 12, 24 and 60 months of age or discontinued the study with the reason of discontinuation.

Participants who discontinued for defined reasons will be summarized (*Table 6*). Medical problems as a reason for discontinuation will be listed for each study arm with number (and percentage) of occurrence according to SOC and HLGT.

Table 6: Drop-out reasons per intervention group

Drop-outs per group			
	Cow	Goat	p-value
	n (%)	n (%)	
Number of drop-outs	xx	xx	x
Drop out reasons			
Medical problem(s)/AE	xx	xx	x
Withdrawal	xx	xx	x

Loss-to-follow-up	xx	xx	x
Other	xx	xx	x

Flow of the participants through the study will be illustrated in a CONSORT diagram. The flow diagram will summarise the number of patients according to their status at each stage of the study, including:

- number of subjects screened
- eligible and randomised
- received the randomised allocation
- did not receive the randomised allocation*
- lost to follow-up*
- discontinued the intervention*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided.

5.5 Withdrawal / follow-up

Withdrawal from intervention: Subjects can discontinue consuming the study formula, either by their own decision or upon the decision of the Site Principal Investigator if continuation would be detrimental to the subject's well-being.

If infants do not consume any study formula before the infant is 4 months old, the infants are excluded from the study.

Withdrawal from follow-up: Subjects can withdraw consent to participate in the trial at any time without any given reason. Once consent is withdrawn, no further evaluations will be performed, and no additional data will be collected.

5.6 Protocol deviations

Protocol deviations were classified prior to unblinding of the intervention. Protocol deviations (PD) are departures from the study protocol procedures. Such deviations may be major or minor (Table 7). Protocol deviations are recorded as a tool of safety management of the study (immediate action is required) or for monitoring quality (affects the final quality of the data and long-or-short-term adjustments might be needed).

Major protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of key study data or that may significantly affect the subject's rights, safety or well-being. For example, major protocol deviations may include enrolling subjects in violation of key eligibility criteria or failing to collect data necessary to interpret primary outcomes, as this may compromise the scientific value of the trial.

Minor protocol deviations are all changes or alterations in the conduct of the trial which do not have a major impact on the subject's rights, safety or well-being, or the completeness, accuracy and reliability of the study data necessary to interpret primary outcomes.

Table 7: Summary table of major and minor protocol deviations

Protocol Deviation description	Class
Date signature before Date screening	Major
All parental signatures missing	Major
Second parental signature missing (if applicable)	Major
Consent procedure not followed	Major
Violation of inclusion / exclusion criteria, but enrolled	Major
Product dispensed different from product allocated	Major
Study product not consumed for more than 3 consecutive days	Major
Study product consumption discontinued	Major
No data on AD according UKWP criteria during the first year of life	Major
SAE reporting to sponsor/LMU [Timestamp of completion of SAE report > timestamp of first report entered into database + 24h (or) date report entered into database > date of awareness + 2days]	Major
Missing visit V12	Major
V12 out of visit windows: [12months of age +14days] < [Date V12] < [12months of age - 14days]	Major
Missing three or more visits	Major
Incorrect signed or dated	Minor
Study product not stored at participant's home according to instructions	Minor
Study product not shipped in time	Minor
Expired study product was consumed	Minor
Introduction of complimentary foods before age of 4 months	Minor
Data on AD according UKWP criteria missing at one or more time points (not all) during the first year of life	Minor
No data on AD according UKWP criteria after first year of life	Minor
AD severity (SCORAD) never assessed	Minor
AD severity (POEM) never assessed	Minor
Health assessment never done	Minor
Skin care and AD medication never assessed	Minor
Baseline information not collected	Minor
Allergy symptoms never assessed	Minor
Feeding practices never assessed	Minor
Compliance never assessed	Minor
ITQOL never assessed	Minor
BISQ never assessed	Minor

IGSQ never assessed	Minor
Blood sample never collected	Minor
Anthropometrics never measured	Minor
Blood sample preparation not done according to the lab manual	Minor
Missing phone call	Minor
T0 out of visit windows: [recruitment date + 14days +7days] < [Date T0] < [recruitment date + 14days -7days]	Minor
T2 out of visit windows: [2months of age +7days] < [Date T2] < [2months of age -7days]	Minor
T3 out of visit windows: [3months of age +7days] < [Date T3] < [3months of age -7days]	Minor
T8 out of visit windows: [8months of age +7days] < [Date T8] < [8months of age -7days]	Minor
T10 out of visit windows: [10months of age +7days] < [Date T10] < [10months of age -7days]	Minor
T18 out of visit windows: [18months of age +14days] < [Date T18] < [18months of age -14days]	Minor
T36 out of visit windows: [36months of age +14days] < [Date T36] < [36months of age -14days]	Minor
T48 out of visit windows: [48months of age +14days] < [Date T48] < [48months of age -14days]	Minor
Missing visit (except V12)	Minor
V4 out of visit windows: [4months of age +7days] < [Date V4] < [4months of age -7days]	Minor
V6 out of visit windows: [6months of age +7days] < [Date V6] < [6months of age -7days]	Minor
V24 out of visit windows: [24months of age +14days] < [Date V24] < [24months of age -14days]	Minor
V60 out of visit windows: [60months of age +14days] < [Date V60] < [60months of age -14days]	Minor

The number and percentage of patients having PD(s) will be summarized within category and subcategory of deviation by treatment group using the ITT population.

5.7 Subject's compliance

A compliant participant is defined as a child:

- 1) Who takes study formula continuously over the whole intervention period. Breaks of no longer than 3 consecutive days are accepted (as outlined in 4.3.1) during the intervention period or up to an AD_primary diagnosis;
- 2) Who is not introduced to solid foods before the age of 4 months as outlined in 4.3.2..

5.8 Observation time for incidence calculations

Time under observation is calculated to assess each individual's time at risk to develop AD. The observation period starts at the date of enrolment. For those who develop AD time ends with the AD diagnosis date, different for each AD outcome (in the following chronological priority: 1) AD_primary, 2) AD_UKWP, 3) AD_Dr), if AD date is before the respective outcome visit (12, 24, or 60 months of age). All other subject's time at risk ends with either

- a) the subject's date of AD assessment of the respective outcome visit (12, 24, or 60 months of age) or
- b) the last contact date before a) if the respective AD assessment is missed. A "contact" is either a visit or a phone call, depending on the AD outcome (e.g., AD_primary or AD_UKWP)

Missed visits will not be considered for interval truncating of the time under observation. Observation time is expressed in person-years.

6 Analyses of Outcome Variables

6.1 Primary and secondary AD endpoints

The primary endpoint is the cumulative incidence of AD up to the age of 12 months diagnosed by study personnel during a visit (AD_primary). The incidence of any other AD diagnoses is considered to be a secondary outcome. Incidence rates are calculated by:

$\frac{\text{Number of AD cases}}{\text{Sum of observation time}}$. The sum of observation time is defined by all subjects under observation as outlined in 5.8. For the primary outcome and all AD related secondary outcomes the incidence rates of respective endpoints are separately calculated for the cow and goat milk formula group which are compared with incidence rate ratios: $\frac{\text{Incidence rate}_{\text{cow}}}{\text{Incidence rate}_{\text{goat}}}$.

For descriptive purposes number of new AD cases and of subjects under observation at each visit as well as incidence rate with total number of AD cases and total observation time up to the age of 12 months will be summarized for the two intervention arms (Table 8).

Table 8: Descriptive table of new AD cases by visit

	Cow %(n/N) / IR (n/person- years)	Goat % (n/N) / IR (n/person- years)
AD_primary		
Visit 4	n/N	n/N
Visit 6	n/N	n/N
Visit 12	n/N	n/N
Crude incidence rate AD_primary	xx (xxx/xxx)	xx (xxx/xxx)
AD_UKWP		
Visit 4	xx	xx
Visit 6	xx	xx
Visit 12	xx	xx
Crude incidence rate AD_UKWP	xx	xx
AD_Dr		
Visit 4	xx	xx
Visit 6	xx	xx
Visit 12	xx	xx
Crude incidence rate AD_Dr	xx	xx
AD_any		

Visit 4	XX	XX
Visit 6	XX	XX
Visit 12	XX	XX
Crude incidence rate AD_any	XX	XX

Number of subjects completing the visit, by intervention group, along with the crude incidence rates (IR per 100 person-years) for the primary and secondary AD outcomes up to the age 12 of months (Visit 12).

6.2 Safety parameters

6.2.1 Adverse events

The number of participants reporting medical problems (AE) will be summarized overall throughout the study. A summary table will present the number (n) and percentage (% of all subjects in the ITT population) for each safety parameter, comparing the study milk groups using Chi-square test, as shown in *Table 9*, and according to the MedDRA Primary System Organ Class (SOC) and HLGt classification.

Table 9: List of reported AEs and SAEs

	Cow n (%)	Goat n (%)	p-value
AE occurred (individuals)			
No	XX	XX	X
Yes	XX	XX	
AE occurred			X
No	XX	XX	
Yes	XX	XX	
AE related* to study product (individuals**)			X
No	XX	XX	
Yes	XX	XX	
AE related* to study product			X
No	XX	XX	
Yes	XX	XX	
SAE occurred (individuals)			X
No	XX	XX	
Yes	XX	XX	
SAE occurred			X
No	XX	XX	
Yes	XX	XX	
Severity of the SAE			X
Mild	XX	XX	
Moderate	XX	XX	
Severe	XX	XX	
Relatedness of SAE to study product			X
Related	XX	XX	
Unlikely	XX	XX	
Unrelated	XX	XX	

p-values from Chi-square or Fisher exact test, as applicable * probable/definite related; ** Subjects who experienced one or more AEs or SAEs are counted only once with the most study product relatedness

6.2.2 Blood parameters and anthropometry

Blood and anthropometric safety parameters will be summarized as displayed in Table 10. Additionally, estimated differences in length, height, and weight gain between study products at 12 months of age will be calculated as outlined in 4.4.3.

Table 10: Descriptive table for safety measures

	Cow n (%) /Mean (SD)*	Goat n (%) / Mean (SD)*	p-value
Mean haemoglobin	xx	xx	x
Mean eosinophil count	xx	xx	x
Children with anaemia	xx	xx	x
Children with eosinophilia	xx	xx	x
Anthropometric measures	xx	xx	x
Weight-for-age z-scores			
Visit 4	xx	xx	x
Visit 6	xx	xx	x
Visit 12	xx	xx	x
Length-for-age z-scores			
Visit 4	xx	xx	x
Visit 6	xx	xx	x
Visit 12	xx	xx	x
BMI-for-age z-scores			
Visit 4	xx	xx	x
Visit 6	xx	xx	x
Visit 12	xx	xx	x

* as appropriate

6.2.3 Cow milk protein allergy (CMPA)

It is reasonable to expect that results may reveal differentiated diagnostic rates of CMPA between participants consuming cow milk formula and those consuming goat milk formula.

Table 11: Comparison of CMPA among the intervention groups and diagnostic certainty measures for the diagnosis

Cow milk allergy (CMPA)	Goat milk formula intervention group N (%)	Cow milk formula intervention group N (%)	p-value
Physician diagnosed CMPA			
Formula discontinuation and re-challenge performed			
Blood test for specific IgE			
Skin prick test performed			
Standardized provocation test performed			

6.2.4 Comparison of PP and ITT

A comparison of PP and ITT will be conducted for all baseline characteristics based on the baseline table in 4.1 to describe any deviation of PP and ITT in relation to the planned balanced population distribution of the intervention groups.

7 Analysis methods

The primary statistical analysis will be a comparison of the incidence rates of children with AD up to 12 months of age (AD_{primary}) between the goat milk formula group and cow milk formula group. The treatment difference will be summarized using a rate ratio with a 95% CI by employing a generalized linear mixed model with study product as a fixed and site as random effect following a Poisson distribution (formula shown in chapter 1.4) with exponentiated coefficients using a mean–variance adaptive Gauss–Hermite quadrature with 7 integration points and an independent covariance structure of the random-effects. Adjusted models are considered secondary analyses.

Table 12: Incidence rate ratio for AD up to 12 months of age

	Primary model		Adjusted model 1		Adjusted model 2	
	RR (95% CI)	p-value	RR (95% CI)	p-value	RR (95% CI)	p-value
Full analysis set (ITT)						

Per-Protocol (PP) data set						
---------------------------------------	--	--	--	--	--	--

Random-effects Poisson regression model. Primary model: fixed effect study product; random effect study site. Adjusted model 1: fixed effects for sex, country, any parental AD added to primary model. Adjusted model 2: Further adjustments with socio-economic status, any Filaggrin null mutation, antibiotic usage, or family size only in case effect estimates change considerably.

7.1 Missing data

Within the primary outcome analyses, implausible cases are corrected, if possible, otherwise set to missing. Missing values are not imputed.

7.2 General statistical principle for other secondary outcomes

Centre will be generally used in the regression analysis as a random effect in mixed models with study product as a fixed effect; depending on the distribution properties of the outcome variable, a Poisson, logit, or Gaussian distribution will be used. Adjustment will follow the same principle as in the primary analysis.

7.3 Subgroup analysis

Subgroup analyses will be performed as effect modification models as outlined in the protocol. We will look if the interaction effect of the intervention product (goat vs. cow milk) with the effect modifier is significantly improving the model fit compared to the model including the same fixed effect variables without an interaction. If the interaction effect by one of the effect modifiers is significant at the 5% significance level, estimates for interaction effect will be presented as secondary analysis.

The following potential modifiers will be considered: sex, age at randomization, number of weeks with formula consumption before use of study product, country, parental AD, and Filaggrin null mutation.

7.3.1 Early versus late randomization

The timing of the milk exposure and the feeding before the intervention start might impact the effect of the intervention. Thus, the effect of the intervention will be compared in subjects randomized before 4 weeks of age and in subjects with later randomization.

7.3.2 Weeks of formula feeding before start of study formula

Cow's milk intake during the early intervention phase might impact the effect of the intervention. As it is expected that only very few children will be exclusively breastfed before the start of the study formula, the effect of the intervention will be compared in children with formula feeding ≤ 4 weeks before start of intake of study formula (see 4.3.3) to all other children.

7.3.3 Study country

Spain and Poland differ in many aspects, including AD prevalence, family atopic history and nutritional aspects. Therefore, we will compare the effect of the intervention in subjects from Spanish to those from Polish study centres.

7.3.4 Parental atopic dermatitis and Filaggrin gene variants

Parental AD and Filaggrin mutation are considered a strong risk factor for AD. Therefore, we will compare the effect of the intervention for subjects from parent(s) with AD to those without and from subjects with a Filaggrin gene variant and without it.

8 Statistical interim analyses and stopping guidance

An interim analysis was not planned. Specific rules for stopping the trial have not been defined in the study protocol. However, the protocol defined that a safety analysis based on adequacy of growth had to be performed that required external unblinding. A non-inferiority boundary for weight-for-age z-score of minus 0.5 z-scores was chosen according to Onyango et al. (Onyango et al. 2007). Assuming a standard deviation = 1, a non-inferiority margin = -0.5, a power $1-\beta = 80\%$ and a type I error $\alpha = 0.025$ (one-sided), 128 children (63 per group) were needed for this analysis. In October 2022 a data and safety monitoring board report was produced by an external statistician and reviewed by an independent Data and Safety Monitoring Board. No safety issues were detected.

Statistical analyses will be performed, when all recruited infants have passed defined time points: All study infants have been seen at 12 months of age (visit 12) or their age exceeds the examination window of visit 12. All study infants have been seen at 24 months of age (visit 24) or their age exceeds the examination window of visit 24. All study infants have been seen at 60 months of age (visit 60) or their age exceeds the examination window of V60.

9 Unblinding

Unblinding will take place when the visit 12 has been performed in all children. Furthermore, all concerns regarding AD_UKWP (up to 12 months visit), all data concerning baseline characteristics, as shown in *Table 4*, as well as AEs and compliance have to be cleaned, and all related queries must have been resolved. Furthermore, the SAP has to be signed and approved prior to unblinding. The data manager from the LMU management team will carry out the unblinding. Consequently, the data manager will be unblinded, while the PIs, study participants, sponsor, and other members of the LMU management team will remain blinded.

10 Statistical software

All statistical analyses will be carried out with the software package R version R-4.3.2 or above, SPSS 29.0.0.0 (241) and Stata/MP 18.0.

11 References

- Alexander, D. D., and M. D. Cabana. 2010. "Partially hydrolyzed 100% whey protein infant formula and reduced risk of atopic dermatitis: a meta-analysis." *J Pediatr Gastroenterol Nutr* no. 50 (4):422-30. doi: 10.1097/MPG.0b013e3181cea52b.
- Charman, C. R., A. J. Venn, and H. C. Williams. 2004. "The patient-oriented eczema measure - Development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective." *Archives of Dermatology* no. 140 (12):1513-1519. doi: 10.1001/archderm.140.12.1513.
- Cuello-Garcia, C., A. Fiocchi, R. Pawankar, J. J. Yepes-Nunez, G. P. Morgano, Y. Zhang, A. Agarwal, S. Gandhi, L. Terracciano, H. J. Schunemann, and J. L. Brozek. 2017. "Prebiotics for the prevention of allergies: A systematic review and meta-analysis of randomized controlled trials." *Clinical and Experimental Allergy* no. 47 (11):1468-1477. doi: 10.1111/cea.13042.
- Draaisma, E., L. Garcia-Marcos, J. Mallol, D. Sole, V. Perez-Fernandez, P. L. P. Brand, and Eisl Study Grp. 2015. "A multinational study to compare prevalence of atopic dermatitis in the first year of life." *Pediatric Allergy and Immunology* no. 26 (4):359-366. doi: 10.1111/pai.12388.
- Ferry, Jill, Rafael Galera-Martínez, Cristina Campoy, M. Pipaon, Elzbieta Jarocka-Cyrta, Jaroslaw Walkowiak, Bartosz Romańczuk, Joaquin Escribano, Mariona Gispert, Paula Grattarola, Dariusz Gruszczyński, Iris Iglesia, Veit Grote, Hans Demmelmaier, Uschi Handel, Sophie Gallier, and Berthold Koletzko. 2023. "Effects of infant feeding with goat milk formula or cow milk formula on atopic dermatitis: protocol of the randomised controlled Goat Infant Formula Feeding and Eczema (GlraFFE) trial." *BMJ Open* no. 13:e070533. doi: 10.1136/bmjopen-2022-070533.
- Iskedjian, M., C. Dupont, J. Spieldenner, G. Kanny, F. Raynaud, B. Farah, and F. Haschke. 2010. "Economic evaluation of a 100% whey-based, partially hydrolysed formula in the prevention of atopic dermatitis among French children." *Curr Med Res Opin* no. 26 (11):2607-26. doi: 10.1185/03007995.2010.525472.
- Kamer, B., R. Pasowska, E. Dolka, A. Blomberg, and H. Rotsztejn. 2013. "Prevalence of atopic dermatitis in infants during the first six months of life: authors' observations." *Postepy Dermatologii i Alergologii* no. 30 (5):277-281. doi: 10.5114/pdia.2013.38355.
- Landgraf, J. M., I. Vogel, R. Oostenbrink, M. E. van Baar, and H. Raat. 2013. "Parent-reported health outcomes in infants/toddlers: measurement properties and clinical validity of the ITQOL-SF47." *Qual Life Res* no. 22 (3):635-46. doi: 10.1007/s11136-012-0177-8.
- Onyango, A. W., M. de Onis, M. Caroli, U. Shah, Y. Sguassero, N. Redondo, and B. Caroli. 2007. "Field-testing the WHO child growth standards in four countries." *J Nutr* no. 137 (1):149-52. doi: 10.1093/jn/137.1.149.
- Pedroza, C., and V. T. T. Truong. 2017. "Estimating relative risks in multicenter studies with a small number of centers - which methods to use? A simulation study." *Trials* no. 18 (1):512. doi: 10.1186/s13063-017-2248-1.
- Riley, A. W., J. Trabulsi, M. Yao, K. B. Bevans, and P. A. DeRusso. 2015. "Validation of a Parent Report Questionnaire: The Infant Gastrointestinal Symptom Questionnaire." *Clin Pediatr (Phila)* no. 54 (12):1167-74. doi: 10.1177/0009922815574075.
- Sadeh, A. 2004. "A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample." *Pediatrics* no. 113 (6):e570-7. doi: 10.1542/peds.113.6.e570.
- Scott, N. W., G. C. McPherson, C. R. Ramsay, and M. K. Campbell. 2002. "The method of minimization for allocation to clinical trials. a review." *Control Clin Trials* no. 23 (6):662-74. doi: 10.1016/s0197-2456(02)00242-8.
- Stalder, J. F., A. Taieb, D. J. Atherton, T. Bieber, E. Bonifazi, A. Broberg, A. Calza, R. Coleman, Y. Deprost, T. L. Diepgen, C. Gelmetti, A. Giannetti, J. Harper, B. Kunz, J. M. Lachapelle, T. Langeland, R. Lever, A. P. Oranje, C. Queilleroussel, J. Revuz, J. Ring, J. C. Roujeau, J. H.

- Saurat, M. Song, D. Tennstedt, D. Vanneste, D. Vieluf, and M. Poncet. 1993. "SEVERITY SCORING OF ATOPIC-DERMATITIS - THE SCORAD INDEX - CONSENSUS REPORT OF THE EUROPEAN TASK-FORCE ON ATOPIC-DERMATITIS." *Dermatology* no. 186 (1):23-&. doi: 10.1159/000247298.
- White, S. J., and L. S. Freedman. 1978. "Allocation of patients to treatment groups in a controlled clinical study." *Br J Cancer* no. 37 (5):849-57. doi: 10.1038/bjc.1978.124.
- Zhou, S. J., T. Sullivan, R. A. Gibson, B. Lonnerdal, C. G. Prosser, D. J. Lowry, and M. Makrides. 2014. "Nutritional adequacy of goat milk infant formulas for term infants: a double-blind randomised controlled trial." *Br J Nutr* no. 111 (9):1641-51. doi: 10.1017/S0007114513004212.