PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	'Barrier dysfunction in Atopic newBorns studY' (BABY): protocol of a Danish prospective birth cohort study
AUTHORS	Gerner, Trine; Halling-Sønderby, Anne-Sofie; Rasmussen Rinnov, Maria; Haarup Ravn, Nina; Hjorslev Knudgaard, Mette; Menné Bonefeld, Charlotte; Ewertsen, Caroline; Trautner, Simon; Jakaša, Ivone; Kezic, Sanja; Skov, Lone; Thyssen, JP

VERSION 1 – REVIEW

REVIEWER	Sara Brown University of Dundee, Scotland, UK I am part of the BIOMAP consortium which received funding from Sanofi, LEO Pharma, Boehringer Ingelheim International GmbH, Pfizer Limited and UCB Biopharma Sprl in addition to the EU.
REVIEW RETURNED	19-Sep-2019

GENERAL COMMENTS	This is an ambitious study aiming to collect very detailed observational data on a substantial cohort of infants. The cohort acronym is nice! The findings are likely to increase our knowledge about measurements of aspects of skin barrier function in preterm and term infants which is to be welcomed for the field. However the impact on understanding atopic dermatitis (AD) is less clear and it would be helpful for the specific hypotheses and plans for data analysis to be described more fully. My main questions are as follows: i) What involvement did the funders have in study design & on-going conduct of this study? Will the funders have any involvement in data analysis & outputs? - this should be clearly stated. ii) It would be helpful to include your definition of preterm & term in the abstract. iii) Line 112 - why are you focusing on early-onset AD? We know that often this is transient & self-limiting, so it will be important to distinguish transient/trivial AD from the start of more established AD. Is this possible within your study? iv) How will *predictors* of AD be distinguished from *early features* of AD? v) The objectives are not clearly stated. Does the primary objective refer to predictors early in life or AD early in life (or both?) - it would be helpful to clarify. The secondary objectives are also unclear: how many years of life? vi) How is 'severe congenital abnormailty' defined? vii) Is it acceptable to say that people who cannot communicate in Danish are excluded from this study? Would it be possible to use interpreters?

	viii) Why do term infants have more visits than preterm? How will the data at different timepoints be compared? ix) How will the UK WP diagnosis be made over the telephone? Because one of the criteria is 'visible flexural dermatitis'. x) Will the children have a standardised status in terms of recent washing / emollient application pre-TEWL and pre-Raman? This is important. xi) Do you have experience to show that 8 tapestrips from the cheek skin will be acceptable (not too traumatic) for baby and parents? What will you do if this is too uncomfortable and/or not acceptable? xii) Who will carry out the skin assessments? - with what training? xiii) Why are moles being recorded? How will you distinguish congenital from non-congenital melanocytic naevi? Why is this important? xiv) What is the consent process? There is no description of this, but clearly informed consent is very important, with specific consideration of the data storage. xv) Is this study simply observational, or do you have specific hypothese to test? If so, these should be clearly stated and the specific analyses pre-defined. I would like to make the following suggestions for improvements: a) The abstract states that 'skin barrier development ... has been scarcely studied.' There is quite an extensive literature dating back many years on skin barrier development so this sentance should be rephrased. b) The abstract focusses on TEWL as a measure of skin barrier function but in fact you are assessing many other aspects of skin barrier too; this is a strength and should be more clearly stated in the abstract. c) Please add more details regarding the 'custom build' Raman instrument. d) Please add more details regarding the methods & measurements to be made on the tape strip samples. e) I am not an expert on microbiome sampling but it would be helpful to state how the swab sample collection will be standardised. And how will it be analysed? f) Line 265 states that measurements are made 'right after birth' but the protocol describes the preterm babies having their first study visit during the first 31 days of life & term babies in first 3 days of life. Please clarify & correct this 'strength' statement if necessary. Minor comments: 1. Line 97 ... genetic risk variants. It would be more correct to say 'genetic risk effect' - many variants have been described but their contribution to the overall risk remains modest. 2. Lines 108-109: this is unclear. Do you mean that a smaller amount of thymus tissue may lead to a lower number of circulating T cells which in turn reduces the immune response to antigen in the skin? Please consider rephrasing this more clearly. 3. Line 111 typo - examine (not examines) 4. Line 159 - participate (not participates) 5. Line 162 includes a reference for POEM but this is not in the preceding text. 6. It will be difficult to fully sequence FLG in DNA from a buccal swab. The null mutation analysis is much more possible & also much more informative.

	7. Line 219 - typo: swabs (not swaps) 8. Line 237 - what do you mean by 'best' in this context?

REVIEWER	Yukihiro Ohya National Center for Child Health and Development, Japan
REVIEW RETURNED	25-Oct-2019

GENERAL COMMENTS	General Comments This study is a birth cohort study to identify infants at risk of developing AD early in life by examining the skin of newborns. To examine infants including neonates is very tough work, however, worth doing. Specific comments Page 6 from line 116 OBJECTIVES To identify predictors of AD in early childhood is described as primary objective. This expression is vague. Candidates for predictor variables should be stated. Usually, sample size estimation was calculated by using primary outcome. If this case is to be applied, the filaggrin breakdown product 2-pyrrolidone-5-carboxylic acid (PCA) seems to be a predictor as described in the second line of page 12. Is it right? At least another variable may exist as a predictor of AD, since the authors described “to identify predictors (plural form) of AD”. Generally speaking, predictors of AD are not only NMF such as PCA. How can you identify predictors of AD except NMF? Page 6 from line 125 METHODS AND ANALYSIS In the section of Study population and setting, details of recruiting methods and population should be described. Ex. Whether recruitment is done during pregnancy or after delivery? Has General population been targeted to recruit infants or hereditary high-risk population (one of the parents or siblings was affected with allergic disease) been targeted? Page 7 from line 135 Cohort design The authors described “Preterm children are scheduled for two study visits: during the first 31 days of life and approximately two months after their scheduled due date (Figure 1). Term children are scheduled for four study visits: during the first 3 days of life and approximately at 2, 6 and 12 months of age (Figure 2). “ Is the first visit of preterm children scheduled from when to the first 31 days of life? During the first month of life, skin barrier systems the authors plan to measure change dramatically. Therefore, to specify the date range of the first visit of preterm infant is important. In addition to that, acceptable date range of the second visit (2 months of age) of preterm infants and term infants should be described. As well, acceptable date range of 6 and 12 months of age of term infants visit should be described. Page12 from line 244 Sample size estimation The authors described “In our cohort, we hypothesized a 5% change of NMF in children developing AD compared to children without AD. With a 5% two-sided significance level and a power of

	80%, we calculated a sample size of 112 premature children and 223 in term children." This information is not sufficient to calculate the sample size of participants accurately. How did you calculate the sample size 112 for preterm children and 223 for term children? Page 2 and 3 Although the authors declared no conflicts of interest in page 2, they received financial support from private companies as described in page 3.

VERSION 1 – AUTHOR RESPONSE

COMMENTS FROM REVIEWER 1:

- **Comment 1:**
  What involvement did the funders have in study design & on-going conduct of this study? Will the funders have any involvement in data analysis & outputs? - this should be clearly stated.
  Response:
  We thank the reviewer for pointing out the need for clarification regarding any potential involvement of the funders.

  Changes made to our manuscript (line 52):

  The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- **Comment 2:**
  It would be helpful to include your definition of preterm & term in the abstract.
  Response:
  We agree with the reviewer that the definition of both preterm and term children needs to be defined in the abstract.

  Changes made to our manuscript (line 61):

  BABY Cohort is a prospective birth cohort study of 150 preterm children (gestational age (GA) below 37+0) and 300 term children (GA 37+0 to 41+6).

  (The BABY Cohort is a prospective birth cohort study of 150 preterm and 300 term children.)

- **Comment 3:**
  Line 112 - why are you focusing on early-onset AD? We know that often this is transient & self-limiting, so it will be important to distinguish transient/trivial AD from the start of more established AD. Is this possible within your study?
  Response:
  We thank the reviewer for the opportunity to elaborate on this important matter. In this study, we followed children during their first two years of life. It is known, that approximately 80 % of AD patients develop their disease within this time period therefore, we expect to identify and distinguish children with both transient and more established AD. Accordingly, we will seek

additional funding once this study is done to allow for further follow-up but with a slightly different aim.

- **Comment 4:**
  How will *predictors* of AD be distinguished from *early features* of AD?
  Response:
  We thank the reviewer for this very interesting, yet very philosophical question. Defining on tissue level if a change in skin barrier is a predictor or an early manifestation of AD is very difficult and will in many cases be impossible, for example determining if dry skin is a predictor or an early feature. In other cases, it will be straight forward assessing if environmental exposures act as predictors. Therefore, our distinguishing will be done on a case to case basis depending on the studied exposure variable.

- **Comment 5:**
  The objectives are not clearly stated. Does the primary objective refer to predictors early in life or AD early in life (or both?) - it would be helpful to clarify. The secondary objectives are also unclear: how many years of life?
  Response:
  We thank the reviewer for giving us the opportunity to clarify the objectives of this study.

  Changes made to our manuscript (line 127 and 132):

  To identify early predictors of AD during the first two years of life.

  (To identify predictors of AD in early childhood.)

  To closely describe the normal skin barrier development including immune activity and skin microbiome in preterm and term newborns during the first two years of life.

  (To closely describe the normal skin barrier development including immune activity and skin microbiome in preterm and term newborns during the first years of life.)

- **Comment 6:**
  How is 'severe congenital abnormality' defined?
  Response:
  We thank the reviewer for giving us the opportunity to elaborate on this. We did not include children born with conditions affecting their life expectancy. All children where included by trained medical doctors who were able to assess this. In case of any doubt a pediatrician was consulted at the department.

  Changes made to our manuscript (line 155):

  All study visits are conducted by trained medical doctors.

- **Comment 7:**
  Is it acceptable to say that people who cannot communicate in Danish are excluded from this study? Would it be possible to use interpreters?

Response:
The question from the reviewer addresses an interesting and difficult challenge for all clinical studies. The aim of this study was to evaluate skin barrier development in the general population in Denmark. As most of the population is Danish speaking, we believe that we would get the best representation by only including Danish speaking parents. Furthermore, in uncomplicated childbirths the women are managed as outpatients and are discharged a few hours after birth. Therefore, it would be difficult within this timeframe to use interpreters.

- **Comment 8:**
  Why do term infants have more visits than preterm? How will the data at different timepoints be compared?
  Response:
  The NICU at Rigshospitalet is a highly specialized department responsible for treatment of all extremely preterm children in eastern Denmark. A lot of these children continue to have many hospital visits after their discharge. Since many of the families lives far away from the hospital, it has been our assessment, that it would be unrealistic for the families to participate in more than one follow-up visit.
  Overall, assessment of development in preterm children is always based on the corrected age instead of the chronological age. Since it is established that preterm children have delayed skin maturation, we have chosen to assess them 2 months after their planned due date, to adjust for this difference between the preterm and term children.

  Changes made to our manuscript (line 153):
  For all study visits we register the time of the study visit, to be able to adjust for any effects that occur due to age differences.

- **Comment 9:**
  How will the UK WP diagnosis be made over the telephone? Because one of the criteria is 'visible flexural dermatitis'.

  Response:
  We thank the reviewer for letting us clarify this important matter. During the structured telephone interviews, we ask the parents if the child have visible flexural dermatitis in the elbows or knees. If the child has developed any sign of eczema, we invite the parents in for another study visit where we clinically assess the eczema and make an AD diagnosis using Hanifin and Rajkas diagnostic criteria.

  Changes made to our manuscript (line 179):

  AD assessment according to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis, with parental assessment of visible flexural dermatitis in the elbows or knees.

  (AD assessment according to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis.)

- **Comment 10:**
  Will the children have a standardised status in terms of recent washing / emollient application pre-TEWL and pre-Raman? This is important.

Response:
This question from the reviewer addresses a very important matter. During all study visits we interview parents about the child's emollient and shower habits including time of the last shower (line 165).

- **Comment 11:**
  Do you have experience to show that 8 tape strips from the cheek skin will be acceptable (not too traumatic) for baby and parents? What will you do if this is too uncomfortable and/or not acceptable?
  Response:
  We thank the reviewer for giving us the possibility to clarify this matter. Deciding how many tape strips to use varies a lot between studies, mainly depending on which outcomes to assess. Furthermore, tape stripping has only been done in a few studies with children. In our study, we chose 8 tape strips after consulting with researchers who have a lot of experience with tape stripping, also in newborns. Furthermore, McAleer et al. has in BJD 2018 published a study using 8 tape strips on newborns.
  Regarding the possibility of experiencing discomfort during the tape stripping procedure we consulted pediatricians for advice. They did not find any concerns about the procedure as the tapes are less sticky than other patches (i.e. ECG patches) that are used frequently on newborns in the NICU. Nonetheless, if a child experience any discomfort, we give the child sugar water solution for pain relief or stop tissue collection.

  Changes made to our manuscript (line 207):

During all study visits, SC is collected by tape stripping as previously described.[22, 23]

22. Kezic S, Kammeyer A, Calkoen F, et al. Natural moisturizing factor components in the stratum corneum as biomarkers of filaggrin genotype: evaluation of minimally invasive methods. Br J Dermatol. 2009;161(5):1098-104.

23. McAleer MA, Jakasa I, Raj N, et al. Early-life regional and temporal variation in filaggrin-derived natural moisturizing factor, filaggrin-processing enzyme activity, corneocyte phenotypes and plasmin activity: implications for atopic dermatitis. Br J Dermatol. 2018;179(2):431-41.

(During all study visits, SC is collected by tape stripping.)

- **Comment 12:**
  Who will carry out the skin assessments? - with what training?
  Response:
  We thank the reviewer for pointing out that this has not been thoroughly described in the manuscript. All study visits are done by medical doctors with a special interest in dermatology who have received thorough clinical training in skin assessment from a professor in dermatology.

  Changes made to our manuscript (line 157):

  All study visits are conducted by trained medical doctors.

- **Comment 13:**
  Why are moles being recorded? How will you distinguish congenital from non-congenital melanocytic naevi? Why is this important?
  Response:
  Thank you for giving us possibility to elaborate on this. In order to thoroughly describe the normal skin barrier development, we have decided to closely examine the skin on all children in study. This includes moles and capillary malformations. Examining the skin several times during the first year of life we will be able to distinguish which are congenital from those who are not. Also, several studies, have shown a lower number of nevi in AD patients, and this is even confirmed in meta-analyses, so we wanted to include this variable.

- **Comment 14:**
  What is the consent process? There is no description of this, but clearly informed consent is very important, with specific consideration of the data storage.
  Response:
  We apologize and thank the reviewer for letting us know that we have not described the consent process.

  Changes mad to our manuscript (line 315):

  Both parents or guardians will give written informed consent prior to entry to the study.

- **Comment 15:**
  Is this study simply observational, or do you have specific hypothese to test? If so, these should be clearly stated and the specific analyses pre-defined.
  Response:
  We thank the reviewer for giving us an opportunity to explaining this. Despite years of research, there is still a need for understanding the skin barrier development in newborns and the risk of AD. Interestingly preterm children develop less AD than term children, and we therefore hope, that assessing the skin barrier in both preterm and term with several modalities, we will learn more about the risk factors for AD.

  Changes made to our manuscript in the (line 136):

  The BABY Cohort is an ongoing prospective and observational birth cohort study recruiting 150 preterm and 300 term newborn infants.

  (The BABY Cohort is an ongoing prospective birth cohort study recruiting 150 preterm and 300 term newborn infants.)

- **Comment 16:**
  The abstract states that 'skin barrier development ... has been scarcely studied.' There is quite an extensive literature dating back many years on skin barrier development so this sentance should be rephrased.
  Response:
  We sincerely thank the reviewer for letting us know that our introduction in our abstract needs to be expressed more precisely.

Changes made to our manuscript (line 58):
Skin barrier development and dysfunction in premature and mature newborns is important for the risk of atopic dermatitis (AD).

(The skin barrier development in premature and mature newborns has been scarcely studied but may be important for the risk of atopic dermatitis (AD).)

- **Comment 17:**
  The abstract focusses on TEWL as a measure of skin barrier function but in fact you are assessing many other aspects of skin barrier too; this is a strength and should be more clearly stated in the abstract.
  Response:
  We thank the reviewer for pointing out the need for specification of our skin barrier assessments in our abstract.

  Changes made to our manuscript (line 62):

  Skin barrier is assessed through transepidermal water loss, tape stripping, Raman-spectroscopy and microbiome sampling.

  (Skin barrier function is assessed by transepidermal water loss. Biomolecules important for skin barrier function and immune response are investigated by Raman-spectroscopy and stratum corneum and microbiome sampling.)

- **Comment 18:**
  Please add more details regarding the 'custom build' Raman instrument.
  Response:
  We thank the reviewer for asking about more details on the Raman. ´Therefore, we have cited two extra manuscripts describing the method and technique in detail (line 199).

  - *Caspers PJ, Lucassen GW, Carter EA, et al. In vivo confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. J Invest Dermatol. 2001;116(3):434-42.*

  - *Caspers PJ, Lucassen GW, Puppels GJ. Combined In Vivo Confocal Raman Spectroscopy and Confocal Microscopy of Human Skin. Biophysical Journal. 2003;85(1):572-80.*

- **Comment 19:**
  Please add more details regarding the methods & measurements to be made on the tape strip samples.
  Response:
  We thank the reviewer for letting us know that we have not described the measurements on the tape strips enough. The last years, the possibilities of analyzing tape strips have evolved tremendously and we are expecting that new techniques will be developed before we have ended our study. However, most of our analyses have been planned and we have added information on this in the manuscript. Furthermore, we have cited two studies describing the NMF and morphology assessment techniques.

- *Kezic S, Kammeyer A, Calkoen F, et al. Natural moisturizing factor components in the stratum corneum as biomarkers of filaggrin genotype: evaluation of minimally invasive methods. Br J Dermatol. 2009;161(5):1098-104.*

- *Riethmuller C, McAleer MA, Koppes SA, et al. Filaggrin breakdown products determine corneocyte conformation in patients with atopic dermatitis. J Allergy Clin Immunol. 2015;136(6):1573-80 e2.*

Changes made to our manuscript (line 215):

SC samples will be analyzed for biomarkers of the immune response by multiplex immuno - assays, NMF using a liquid chromatography previously described by Kezic et al. and corneocyte surface morphology by atomic force microscopy

(SC will be examined for NMF, proteins, cytokines, lipids and morphology.)

- **Comment 20:**
  I am not an expert on microbiome sampling but it would be helpful to state how the swab sample collection will be standardised. And how will it be analysed?
  Response:
  We thank the reviewer for pointing out that we have not described the microbiome sampling enough in our manuscript. We have therefore rephrased the description. Regarding the microbiome analysis we have a cooperation with a laboratory with great knowledge about this, but we will not take part in this procedure ourselves.

  Changes made to our manuscript (line 253):

  Skin swabs are rubbed on the skin for 60 seconds and are immediately stored at -80° C until analysis.

  (All samples are immediately stored at -80° C until analysis.)

- **Comment 21:**
  Line 265 states that measurements are made 'right after birth' but the protocol describes the preterm babies having their first study visit during the first 31 days of life & term babies in first 3 days of life. Please clarify & correct this 'strength' statement if necessary.
  Response:
  We apologize for not being precise in description.

  Changes made to our manuscript (line 289):

  The major strength of this birth cohort study is the extensive and repeated skin barrier measurements beginning shortly after birth.

  (The major strength of this birth cohort study is the extensive and repeated skin barrier measurements beginning right after birth.)

- **Comment 22:**

Line 97 ... genetic risk variants. It would be more correct to say 'genetic risk effect' - many variants have been described but their contribution to the overall risk remains modest.
Response:
We thank the reviewer for giving us the opportunity to rephrase

Changes made to our manuscript (line 106):

…genome wide association studies have only identified a relatively small proportion of the genetic risk effect.

(…genome wide association studies have only identified a relatively small proportion of the genetic risk variants.)

- **Comment 23:**
Lines 108-109: this is unclear. Do you mean that a smaller amount of thymus tissue may lead to a lower number of circulating T cells which in turn reduces the immune response to antigen in the skin? Please consider rephrasing this more clearly.
Response:
We apologize that our description of the matter was perhaps somewhat unclear. Previous literature has proved lower number of circulating T-cells in children after total and partial thymectomy.
We have added a citation for this in our manuscript*:*

    - *Eysteinsdottir JH, Freysdottir J, Haraldsson A, et al. The influence of partial or total thymectomy during open heart surgery in infants on the immune function later in life. Clin Exp Immunol. 2004;136(2):349-55.*

Changes made to our manuscript (line 117):

Interestingly, the risk of AD is decreased in premature newborns and infants undergoing heart surgery, which often includes partial or total thymectomy, perhaps due to their reduced number of total lymphocytes and circulating of T-cells resulting in an inappropriate immune response to antigens encountered in the skin.

(Interestingly, the risk of AD is decreased in premature newborns and infants undergoing heart surgery, which often includes partial or total thymectomy, perhaps due to a lower number of circulating T cells and an inappropriate immune response to antigens encountered in the skin.)

- **Comment 24:**
Line 111 typo - examine (not examines) 4. Line 159 - participate (not participates) 5. Line 162 includes a reference for POEM but this is not in the preceding text.
Response:
We sincerely thank the reviewer who found mistakes in our manuscript. All these have been corrected.

- **Comment 25:**
It will be difficult to fully sequence FLG in DNA from a buccal swab. The null mutation analysis is much more possible & also much more informative.
Response:

We agree with the reviewer that this would very difficult. As stated in the 'Genetics' section in our manuscript we have only planned to screen for the most common mutations in the Northern European population. A procedure which we use regularly at the ordinary Dermatology Department at the hospital.

- **Comment 26:**
  Line 219 - typo: swabs (not swaps)
  Response:
  Thank you for finding this mistake. This has been changed.

- **Comment 27:**
  Line 237 - what do you mean by 'best' in this context?
  Response:
  We apologize that our description of the matter was perhaps somewhat unclear. The ultrasound images are assessed by a radiologist with great experience in choosing the best visualization for calculating the thymus index. The best images are those with a full visualization of the gland in the proper scan plane.

COMMENT FROM REVIWER 2:

- **Comment 1:**
  Page 6 from line 116
  OBJECTIVES
  To identify predictors of AD in early childhood is described as primary objective. This expression is vague. Candidates for predictor variables should be stated.
  Usually, sample size estimation was calculated by using primary outcome. If this case is to be applied, the filaggrin breakdown product 2-pyrrolidone-5-carboxylic acid (PCA) seems to be a predictor as described in the second line of page 12. Is it right? At least another variable may exist as a predictor of AD, since the authors described "to identify predictors (plural form) of AD".
  Generally speaking, predictors of AD are not only NMF such as PCA. How can you identify predictors of AD except NMF?
  Response:
  We thank the reviewer for giving us the opportunity to clarify the objectives of this study. Our cohort is strictly observational, and we are planning on assessing the skin through a variety of procedures, including TEWL, RAMAN, skin swabs, tape strips, parental interviews etc., to assess both the normal and atopic skin barrier development. With so many different procedures we hope to learn more about which factors that are predictors for AD and which are not.
  At first, we conducted our power analysis on a change in NMF for the purpose of ethical approval and to get an impression of the size needed to examine various barrier predictors with a conservative but realistic prediction. Since the intention of the study became increasingly observational and aimed to assess multiple predictors for AD we believe that the calculations are no longer relevant and do not offer a meaningful full picture on how to size the cohort.
  We respectfully suggest removing the paragraph on sample size calculation from the article (line 268-277).

- **Comment 2:**
  Page 6 from line 125

METHODS AND ANALYSIS

In the section of Study population and setting, details of recruiting methods and population should be described. Ex. Whether recruitment is done during pregnancy or after delivery? Has General population been targeted to recruit infants or hereditary high-risk population (one of the parents or siblings was affected with allergic disease) been targeted?

Response:

We thank the reviewer for giving us the opportunity to explain this matter. In our study we are recruiting parents after they have been given birth. As stated in the manuscript, we are recruiting eligible participants at the maternity ward, which is a department for women who already have given birth.

We recruit all newborns independently of their hereditary risk for AD. In our parental interviews we are asking information about parents and siblings' allergic diseases to account for this in future analyses.

Changes made to our manuscript (line 142):

Children are included independently of their hereditary risk for AD

- **Comment 3:**

Page 7 from line 135

Cohort design

The authors described "Preterm children are scheduled for two study visits: during the first 31 days of life and approximately two months after their scheduled due date (Figure 1). Term children are scheduled for four study visits: during the first 3 days of life and approximately at 2, 6 and 12 months of age (Figure 2). " Is the first visit of preterm children scheduled from when to the first 31 days of life? During the first month of life, skin barrier systems the authors plan to measure change dramatically. Therefore, to specify the date range of the first visit of preterm infant is important. In addition to that, acceptable date range of the second visit (2 months of age) of preterm infants and term infants should be described. As well, acceptable date range of 6 and 12 months of age of term infants visit should be described.

Response:

We thank the reviewer for letting us describe the inclusion process in detail.

In order to be able to recruit and examine newborns, we do not always have the possibility to examine them shortly after birth, especially regarding very immature born children, that are not stable and might receive extensive medical care. Therefore, we often must wait until the child is stable, which can take days or sometimes weeks. As soon as the child can cooperate, we perform the examinations.

Regarding the follow-up visits, we always try to schedule these when the child turns 2, 6 and 12 months. In rare cases it can be difficult for the parents to meet up for the visits at these time points. Since we register the current date of all study visits, we will be able to adjust for any effects that might occur due to the small age differences.

Changes made to our manuscript (line 151):

Overall, all children are recruited and examined as soon as possible after their delivery. Very immature born children often receive intensive medical care, and we wait until the child is stable until we perform the examinations. For all study visits the time of the study visit is registered, to be able to adjust for any effects that occur due to age differences.

- **Comment 4:**
  Page12 from line 244
  Sample size estimation
  The authors described “In our cohort, we hypothesized a 5% change of NMF in children developing AD compared to children without AD. With a 5% two-sided significance level and a power of 80%, we calculated a sample size of 112 premature children and 223 in term children.”
  This information is not sufficient to calculate the sample size of participants accurately. How did you calculate the sample size 112 for preterm children and 223 for term children?
  Response:
  We sincerely thank the reviewer for making us aware of the difficulties understanding the sample size calculation. As described previously, our study is strictly observational cohort and our sample calculation is only made on NMF change. Therefore, our result is impossible to use on the other predictors we asses on for AD. We respectfully suggest removing this paragraph from the article (line 268-277).

- **Comment 5:**
  Page 2 and 3
  Although the authors declared no conflicts of interest in page 2, they received financial support from private companies as described in page 3.
  Response:
  We thank the reviewer for giving us the opportunity to clarify this matter.

  Changes made to our manuscript (line 52):

  The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

VERSION 2 – REVIEW

REVIEWER	Sara Brown University of Dundee, Scotland, UK I am a collaborator on the UK A-STAR eczema case collection and the IMI-funded BIOMAP project, each of which have received part-funding from pharmaceutical companies.
REVIEW RETURNED	02-Dec-2019

GENERAL COMMENTS	Thank for responding to my questions. However, the statement about funders' involvement remains unclear. You have stated: 'The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript' - Does this refer to the protocol publication? I was actually asking about the publication of the birth cohort study findings in the future. Will the funders have any role in the data analysis and outputs? It would be helpful to include in the protocol the explanation and clarifications that you have made in response to comments 3, 4, 6, 7, 8, 13, 25 and 27.

	It is not correct to say that this study is using the UK WP criteria for diagnosis, without a trained person assessing the 'visual flexural dermatitis' in any flexure (not just elbows and knees). This wording should be corrected, for example as follows: 'AD assessment was carried using a modification to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis: in this modification visible flexural dermatitis was recorded based on parental-report of dermatitis in the elbows or knee creases.' The response to comment 10 states that you will enquire about the time since washing, but not that this will be standardised before key measurements of barrier function - this is an important limitation which should be clearly stated. Your response to comment 21 remains unclear: is the phrase 'shortly after birth' intended to include 3-31 days of life as appears to be described elsewhere in the protocol?

REVIEWER	Yukihiro Ohya National Center for Child Health and Development, Japan Grants from Ministry of Health, Labour and Welfare, National Centre for Child Health and Development,Japan Agency for Medical Research and Development, Joint research expenses from Yakult, lecture fees from Maruho, Mylan, Kyorin, Kyowa Hakko Kirin, Sanofi, Shiseido, Sysmex, Taiho Pharma, Thermo fisher Scientific, Torii Pharmacerutical, Towa Pharmaceutical outside the submitted work; .
REVIEW RETURNED	23-Dec-2019

GENERAL COMMENTS	General comment I would like to thank the authors for responding to my comments, however, some points still remain to be clarified. Specific comments Regarding the authors' responses to previous comment 1 and comment 4. The authors removed the paragraph on sample size calculation from the article, however, my previous comments 1 and 4 did not suggest the removal of the description of sample size calculation. NMF such as PCA seems to be one of the predictor variables treated in primary objective. My intention of previous comment 1 is to describe all candidate variables such as PCA measured by Raman spectroscopy and the other ones as predictors of AD. My intention of previous comment 4 is to describe more details of sample size calculation because the information of the following description is not enough for readers to reproduce same calculation of the sample size. "In our cohort, we hypothesized a 5% change of NMF in children developing AD compared to children without AD. With a 5% two-sided significance level and a power of 80%, we calculated a sample size of 112 premature children and 223 in term children." Regarding the authors' response to previous comment 5. According to International Committee of Medical Journal Editors (ICMJE) of which BMJ journals follow on the recommendation, the authors should disclose their COIs according to ICMJE Form for Disclosure of Potential Conflicts of Interest. The section 2 in that Form stated as follows:

	Did you or your institution at any time receive payment or services from a third party (government, commercial, private foundation, etc.) for any aspect of the submitted work (including but not limited to grants, data monitoring board, study design, manuscript preparation, statistical analysis, etc.)? Are there any relevant conflicts of interest? Yes No If yes, please fill out the appropriate information below. ICMJE Form for Disclosure of Potential Conflicts of Interest can be downloaded from the following site. (http://icmje.org/conflicts-of-interest/),

VERSION 2 – AUTHOR RESPONSE

COMMENTS FROM REVIEWER 1:

- **Comment 1:**

  The statement about funders' involvement remains unclear.

  You have stated: 'The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript'

  - Does this refer to the protocol publication? I was actually asking about the publication of the birth cohort study findings in the future. Will the funders have any role in the data analysis and outputs?

  Response:

  We thank the reviewer for pointing out the need for clarification of potential involvement of the funders.

  Changes made to our manuscript (line 61):

  …, as well as no role in future publications.

- **Comment 2:**

  It would be helpful to include in the protocol the explanation and clarifications that you have made in response to comments 3, 4, 6, 7, 8, 13, 25 and 27

  Response:

  We sincerely than the reviewer for letting us know, that we need to clarify the manuscripts further. Below is all or amendments to the manuscript

  Comment 3:

  Line 112 - why are you focusing on early-onset AD? We know that often this is transient & self-limiting, so it will be important to distinguish transient/trivial AD from the start of more established AD. Is this possible within your study?

  Response:

  We thank the reviewer for the opportunity to elaborate on this important matter. In this study, we followed children during their first two years of life. It is known, that approximately 80 % of AD patients develop their disease within this time period therefore, we expect to identify and distinguish children with both transient and more

established AD. Accordingly, we will seek additional funding once this study is done to allow for further follow-up but with a slightly different aim.
Changes made to our manuscript (line 300):

Since approximately 80% of AD patients develop their disease within the first two years of life, we expect to identify children with both transient and more established AD.

Comment 4:

How will *predictors* of AD be distinguished from *early features* of AD?

Response:
We thank the reviewer for this very interesting, yet very philosophical question. Defining on tissue level if a change in skin barrier is a predictor or an early manifestation of AD is very difficult and will in many cases be impossible, for example determining if dry skin is a predictor or an early feature. In other cases, it will be straight forward assessing if environmental exposures act as predictors. Therefore, our distinguishing will be done on a case to case basis depending on the studied exposure variable.
Changes made to our manuscript (line 313):

Due to our study design, we cannot discriminate clearly between early features and predictors.

Comment 6:

How is 'severe congenital abnormality' defined?

Response:
We thank the reviewer for giving us the opportunity to elaborate on this. We did not include children born with conditions affecting their life expectancy. All children where included by trained medical doctors who were able to assess this. In case of any doubt a pediatrician was consulted at the department.
Changes made to our manuscript (line 139):

Children eligible for enrolment are preterm newborns (GA below 37+0) excluding preterm newborns with severe congenital abnormality or conditions affecting their life expectancy…

(Children eligible for enrolment are preterm newborns (GA below 37+0) excluding preterm newborns with severe congenital abnormality …)

Comment 7:

Is it acceptable to say that people who cannot communicate in Danish are excluded from this study? Would it be possible to use interpreters?

Response:
The question from the reviewer addresses an interesting and difficult challenge for all clinical studies. The aim of this study was to evaluate skin barrier development in the general population in Denmark. As most of the population is Danish speaking, we believe that we would get the best representation by only including Danish speaking parents. Furthermore, in uncomplicated childbirths the women are managed as outpatients and are discharged a few hours after birth. Therefore, it would be difficult within this timeframe to use interpreters.
Changes made to our manuscript (line 142):

Children with parents unable to communicate in Danish are excluded, since it is not possible to use (for practical and financial reasons) interpreters right after birth given that we have to be very flexible and recruit at odd hours.
(Children with parents unable to communicate in Danish are excluded.)

Comment 8:

Why do term infants have more visits than preterm? How will the data at different timepoints be compared?

Response:
The NICU at Rigshospitalet is a highly specialized department responsible for treatment of all extremely preterm children in eastern Denmark. A lot of these children continue to have many hospital visits after their discharge. Since many of the families lives far away from the hospital, it has been our assessment, that it would be unrealistic for the families to participate in more than one follow-up visit.
Overall, assessment of development in preterm children is always based on the corrected age instead of the chronological age. Since it is established that preterm children have delayed skin maturation, we have chosen to assess them 2 months after their planned due date, to adjust for this difference between the preterm and term children.
Changes made to our manuscript (line 151):
Many premature born children continue to have many hospital visits after their discharge, and many of the families lives far away from the hospital, i.e. other parts of Denmark. Therefore, preterm children are only scheduled to participate in one follow-up visit. We can therefore only make certain comparisons across the two groups.

Comment 13:

Why are moles being recorded? How will you distinguish congenital from non-congenital melanocytic naevi? Why is this important?

Response:
Thank you for giving us possibility to elaborate on this. In order to thoroughly describe the normal skin barrier development, we have decided to closely examine the skin on all children in study. This includes moles and capillary malformations. Examining the skin several times during the first year of life we will be able to distinguish which are congenital from those who are not. Also, several studies, have shown a lower number of nevi in AD patients, and this is even confirmed in meta-analyses, so we wanted to include this variable.

Changes made to our manuscript (line 226):
A complete examination of the skin is performed at each study visit to describe the normal skin barrier development.
(A complete examination of the skin is performed at each study visit.)
Changes made to our manuscript (line 228):
Studies and meta-analysis have shown that the number of nevi is inverse with AD. However, we are not aware of prospective data collection.

Comment 25:

It will be difficult to fully sequence FLG in DNA from a buccal swab. The null mutation analysis is much more possible & also much more informative.

Response:

We agree with the reviewer that this would very difficult. As stated in the 'Genetics' section in our manuscript we have only planned to screen for the most common mutations in the Northern European population. A procedure which we use regularly at the ordinary Dermatology Department at the hospital.

Changes made to our manuscript (line 244):

Buccal swabs (Isohelix, Harrietsham, U.K.) are used to collect DNA to screen for the most common FLG mutations in Northern European populations (R501X, 2282del4 and R2447X) by TaqMan genotyping assay, a routine analysis in our Biochemical department, ….

(Buccal swabs (Isohelix, Harrietsham, U.K.) are used to collect DNA to screen for the most common FLG mutations in Northern European populations (R501X, 2282del4 and R2447X) …)

Comment 27:

Line 237 - what do you mean by 'best' in this context?

Response:

We apologize that our description of the matter was perhaps somewhat unclear. The ultrasound images are assessed by a radiologist with great experience in choosing the best visualization for calculating the thymus index. The best images are those with a full visualization of the gland in the proper scan plane.

'Changes made to our manuscript (line 267):

The best images with a full visualization of the gland are selected by a trained radiologist. (The best measurement in both planes is selected.)

- **Comment 3:**

  It is not correct to say that this study is using the UK WP criteria for diagnosis, without a trained person assessing the 'visual flexural dermatitis' in any flexure (not just elbows and knees). This wording should be corrected, for example as follows:

  'AD assessment was carried using a modification to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis: in this modification visible flexural dermatitis was recorded based on parental-report of dermatitis in the elbows or knee creases.'

  Response:

  We thank the reviewer for letting us clarify this important matter.

  Changes made to our manuscript (line: 183)

  AD assessment using a modification to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis, with parental assessment of visible flexural dermatitis in the elbows or knees. (AD assessment according to the U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis, with parental assessment of visible flexural dermatitis in the elbows or knees.)

- **Comment 4:**

  The response to comment 10 states that you will enquire about the time since washing, but not that this will be standardised before key measurements of barrier function - this is an important limitation which should be clearly stated.

  Response:

  We thank the reviewer for addressing this very important matter.

  Changes made to our manuscript (line: 309)

  Since bathing habits prior to study visits are not standardized, but only registered, this might impact our skin barrier assessments.

- **Comment 5:**

  Your response to comment 21 remains unclear: is the phrase 'shortly after birth' intended to include 3-31 days of life as appears to be described elsewhere in the protocol?

  Response:

  We apologize for not being precise in description.

  Changes made to our manuscript (line 294):

  The major strength of this birth cohort study is the extensive and repeated skin barrier measurements.

  (The major strength of this birth cohort study is the extensive and repeated skin barrier measurements beginning shortly after birth.)

COMMENT FROM REVIWER 2:

- **Comment 1:**

  Regarding the authors' responses to previous comment 1 and comment 4.

  The authors removed the paragraph on sample size calculation from the article, however, my previous comments 1 and 4 did not suggest the removal of the description of sample size calculation.

  NMF such as PCA seems to be one of the predictor variables treated in primary objective. My intention of previous comment 1 is to describe all candidate variables such as PCA measured by Raman spectroscopy and the other ones as predictors of AD.

  My intention of previous comment 4 is to describe more details of sample size calculation because the information of the following description is not enough for readers to reproduce same calculation of the sample size. "In our cohort, we hypothesized a 5% change of NMF in children developing AD compared to children without AD. With a 5% two-sided significance level and a power of 80%, we calculated a sample size of 112 premature children and 223 in term children."

  Response:

We thank the reviewer for making us aware of the need for clarification on the sample size calculation. We have conducted an updated power calculation and described our method below.

Changes made to out manuscript (line 275):

The sample size calculation is based on including preterm and mature children in a 1:2 ratio. The power calculation was based on an expected prevalence of AD in 20 % of the cohort population. Based on previous knowledge, where adult controls have and NMF of 0.095 +/- 0.029, we hypothesized a 12% change in NMF in children developing AD compared with children without AD. With a 5% two-sided significance level and a power of 80%. AD, as we calculated at sample size of 366 children. In order to account for possible drop-outs we decided on a study population of 450 participants in total, whereas 150 were preterm and 300 mature children.

- **Comment 2:**

  Regarding the authors' response to previous comment 5.

  According to International Committee of Medical Journal Editors (ICMJE) of which BMJ journals follow on the recommendation, the authors should disclose their COIs according to ICMJE Form for Disclosure of Potential Conflicts of Interest. The section 2 in that Form stated as follows:

  Did you or your institution at any time receive payment or services from a third party (government, commercial, private foundation, etc.) for any aspect of the submitted work (including but not limited to grants, data monitoring board, study design, manuscript preparation, statistical analysis, etc.)? Are there any relevant conflicts of interest? Yes No If yes, please fill out the appropriate information below.

  ICMJE Form for Disclosure of Potential Conflicts of Interest can be downloaded from the following site. (http://icmje.org/conflicts-of-interest/),

  Response:

  We sincerely thank the reviewer for letting us know the need attachment of disclosure forms.

VERSION 3 – REVIEW

REVIEWER	Yukihiro Ohya National Center for Child Health and Development, Japan
REVIEW RETURNED	01-Mar-2020

GENERAL COMMENTS	The authors responded almost properly to my comments 2, however, response to comment 1 has not yet been answered to my all requests. From line 124 to 125: Primary objective: To identify early predictors of AD during the first two years of life. This description is ambiguous. Please list candidate variables clearly as the other protocol papers accepted in the BMJ Open did. From line 227 to 282: Sample size estimation:

	As to Sample size estimation, the hypothesis authors described in their first draft was different from that in this third draft. In the first draft they stated that we hypothesized a 5% change of NMF in children developing AD compared to children without AD. But in this third draft they stated that we hypothesized a 12% change in NMF in children developing AD compared with children without AD. However, the description in the third draft was improved and still additional information remains to be described. What kind of statistical method was used to test the sample size estimation? The test seems to be based on a 2-Sided non-parametric Wilcoxon Mann Whitney test for continuous response data with significance level 0.05. Is it right? Did you hypothesize that NMF of infants developing AD will change from 0.0836mmol/g tissue (12 % reduction compared to the normal adult control) measured at their first visit to 0.095 mmol/g tissue in 2 years of age? Please describe the detail of your statistical calculation for the readers to reproduce the same result.

VERSION 3 – AUTHOR RESPONSE

COMMENTS FROM REVIWER 2:

- **Comment 1:**

  From line 124 to 125: Primary objective: To identify early predictors of AD during the first two years of life.

  This description is ambiguous. Please list candidate variables clearly as the other protocol papers accepted in the BMJ Open did.

  Response:

  We thank the reviewer for making us aware of the need further clarification on predictors for development of AD.

  Changes made to our manuscript (line 124):

  To identify early predictors of AD during the first two years of life. The study will assess patient and parental characteristics, family history of atopic comorbidities, exposures during pregnancy and in infancy and skin barrier function and development.

  (To identify early predictors of AD during the first two years of life.)

- **Comment 2:**

  From line 227 to 282: Sample size estimation:

  As to Sample size estimation, the hypothesis authors described in their first draft was different from that in this third draft.

  In the first draft they stated that we hypothesized a 5% change of NMF in children developing AD compared to children without AD. But in this third draft they stated that we hypothesized a 12% change in NMF in children developing AD compared with children without AD.

However, the description in the third draft was improved and still additional information remains to be described. What kind of statistical method was used to test the sample size estimation? The test seems to be based on a 2-Sided non-parametric Wilcoxon Mann Whitney test for continuous response data with significance level 0.05. Is it right?
Did you hypothesize that NMF of infants developing AD will change from 0.0836mmol/g tissue (12 % reduction compared to the normal adult control) measured at their first visit to 0.095 mmol/g tissue in 2 years of age?
Please describe the detail of your statistical calculation for the readers to reproduce the same result.

Response:
We sincerely thank the reviewer for asking us to improve the power calculation. We acknowledge that there have been unfortunate changes in our power calculation, since the calculation in our first draft was computed by an employee no longer affiliated with our department. We have no insight in the software or setting used at the time.
Since our cohort study is focusing on multiple important outcomes, a sufficient power calculation is a challenge, as we only can consider one selected outcome for our calculation. At the time we decided to move on with the study, we aimed for the highest possible number of children to allow for analysis of as many skin barrier measurement outcomes as possible, given that the field is evolving so quickly. At the time of study initiation, we focused on NMF as predictor of AD and therefore power calculations were done with this. We have clarified our sample size calculation and hope that this satisfy the reviewer.

<u>Changes made to our manuscript (line 277):</u>
The sample size calculation was based on including preterm and mature children in a 1:2 ratio. The power calculation was based on an expected prevalence of AD in 20% of the cohort population, assessing changes in NMF, which is one of multiple important endpoints in our study. Based on a previous study, where adult controls had an NMF of 0.095 +/-0.029,[30] we hypothesized a 12% change in NMF in newborns developing AD compared with children without developing AD. Using a two-sided parametric test with an alpha of 5% and a power of 80%, we calculated at sample size of 366 children. In order to account for possible drop-outs and the intention to study many other predictors for AD and skin barrier function in general, we decided on a study population of 450 participants in total, i.e. 150 preterm and 300 mature children.

(The sample size calculation is based on including preterm and mature children in a 1:2 ratio. The power calculation was based on an expected prevalence of AD in 20% of the cohort population. Based on previous knowledge, where adult controls have and NMF of 0.095 +/-0.029,[30] we hypothesized a 12% change in NMF in children developing AD compared with

children without AD. With a 5% two-sided significance level and a power of 80%. AD, we calculated at sample size of 366 children. In order to account for possible drop-outs we decided on a study population of 450 participants in total, whereas 150 were preterm and 300 mature children.)