Author’s response to reviews

Title: The Norwegian Preeclampsia Family Cohort Study: a new resource for investigating genetic aspects and heritability of preeclampsia and related phenotypes

Authors:

Linda Roten (linda.tommerdal@ntnu.no)
Liv Cecilie Thomsen (liv.vestrheim@uib.no)
Astrid Gundersen (astrid.solberg.gundersen@stolav.no)
Mona Fenstad (Mona.Hoyseter.Fenstad@stolav.no)
Maria Lisa Odland (marialisaodland@gmail.com)
Kristin Strand (kristin.m.strand@ntnu.no)
Per Solberg (Per.Solberg@helse-nordtrondelag.no)
Christian Tappert (christian.tappert2@stolav.no)
Elisabeth Araya (elisabeth.araya@stolav.no)
Gunhild Bærheim (gunhild@forusakutten.nhn.no)
Ingvill Lyslo (ingvill.lyslo@sus.no)
Kjersti Tollaksen (kjersti.tollaksen@sus.no)
Line Bjørge (line.bjorge@uib.no)
Rigmor Austgulen (rigmor.austgulen@ntnu.no)

Version: 1 Date: 15 Oct 2015

Author’s response to reviews:

First we would like to thank the reviewer for constructive comments, we also thank and the editor of BMC Pregnancy and Childbirth for the positive response regarding our manuscript.

The section «Characteristics of participants» in the manuscript has been revised by updating the information about the separate publication containing a comprehensive characterization of phenotypic subgroups in the total cohort. A paper by Thomsen et al. entitled «Refined phenotyping identifies links between preeclampsia and related diseases in a Norwegian
preeclampsia family cohort» was published online a head of print in Journal of Hypertension a week after we submitted our original manuscript entitled “The Norwegian Preeclampsia Family Cohort Study: a new resource for investigating genetic aspects and heritability of preeclampsia and related phenotypes”. (The published version of the paper by Thomsen et al. is attached to this resubmission). Thus, we want to refer to the published paper, which the corresponding author has coauthored, in our revised manuscript.

Below is a point-by-point response to the reviewer’s comments.

Reviewer #1:

Methods: identification of participants:

1. Were there any exclusion criteria? Multiple births do not appear to have been excluded, but what about lupus, diabetes, chronic hypertension, etc.? If not excluded, what were the proportions of women with these conditions? (Results) 2. Which hospitals were involved? Should be listed in methods and not solely appear in Figure.

Author’s response:

There were not any exclusion criteria in our study, since the inclusion criteria were quite strict and it is likely that development of preeclampsia is associated with several different underlying non-gestational disease phenotypes.

The proportion of self-reported non-gestational diseases related to development of preeclampsia such as diabetes, pulmonary disease, autoimmune (including systemic lupus erythematosus), atherothrombotic cardiovascular disease (aCVD), kidney disease, chronic hypertension, hypercholesterolemia, myocardial infarction/angina, stroke and thrombosis in the total cohort is presented in the paper by Thomsen et al.. The proportion of these phenotypes in women (nulliparous, parous, non-preeclamptic) and in men is also presented by Thomsen et al. [28].

The table below shows the proportion of these conditions in the index women in our cohort. If accepted by the reviewer and editor we would like to add a table showing these proportions in the paper as well (Table 6).

<table>
<thead>
<tr>
<th>Disease phenotype</th>
<th>Proportion (%)</th>
<th>Missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes mellitus type 1</td>
<td>1/214 (0.5)</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes mellitus type 2</td>
<td>9/213 (4)</td>
<td>7</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td>15/214 (7.0)</td>
<td>6</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>3/215 (1.4)</td>
<td>5</td>
</tr>
</tbody>
</table>
Pulmonary disease 29/215 (13.5) 5
Autoimmune/inflammatory disease 26/214 (12.1) 6
aCVD 66/215 (30.7) 5
Chronic hypertension 54/215 (25.1) 5
Hypercholesterolemia 20/215 (9.3) 5
Myocardial infarction/Angina 3/215 (1.4) 5
Stroke 0/214 (0) 6
Thrombosis 7/215 (3.3) 5

aCVD  Atherothrombotic cardiovascular disease

The definitions of the disease phenotypes listed are shown in table 1 in the paper published by Thomsen et al. [28]. Unfortunately, we are not able to give more detailed information on numbers of individuals with different types of these defined disease phenotypes. The conditions included in the term aCVD may be divided into treated risk factors for aCVD and established (meaning current or previously treated) aCVD. Established aCVD includes coronary heart disease, cerebrovascular disease and other peripheral arterial disease. Diabetes was phenotyped into diabetes mellitus type 1, type 2 and gestational diabetes. The registered inflammatory diseases included hypothyroidism, systemic lupus erythematosus, rheumatic disease, psoriasis, inflammatory bowel disease and the ‘other specified inflammatory diseases’. Kidney diseases included glomerulonephritis, nephrosis, consistent microhaematuria or macrohaematuria, need of dialysis, consistent albuminuria and kidney transplantation. Pulmonary diseases encompassed asthma and chronic obstructive pulmonary disease.

In addition, we would like to make it clear that information on non-gestational disease is self-reported. This information has not been validated by examining medical records.

2. Which hospitals were involved? Should be listed in methods and not solely appear in Figure.

Author’s response:

We acknowledge the reviewer’s comment and have included this information in the text in the methods chapter. The following five hospitals were involved in the study: St. Olavs University Hospital, Haukeland University Hospital, Stavanger University Hospital, Levanger Hospital, Namsos Hospital.

Methods: Linkage to Routine Data Sources
1. Is all follow-up in the cohort passive (i.e., by record linkage)? This should be made very clear. Currently, it seems that this is the case but the paper is vague in this regard.

Author’s response:

Yes, only passive follow-up in the cohort has been planned. The manuscript has been revised making this clear.

Methods: Statistical analysis:

1. I realize that it is not possible to calculate power without a study hypothesis but some discussion of power issues seems appropriate. What do the authors expect to have power to detect, with respect OR for genetic effects for a range of allele frequencies, for example? The authors have some leeway in how they can present this discussion, but some discussion of anticipated power is relevant.

Author’s response:

In general, the power to detect the effect of genes depends on the effect size, the allele frequency and the sample size. The ability to detect the genes (genetic susceptibility variants) increases when sample size and effect size increase. However, the choice of study design and research strategy is still crucial for the chances of a successful outcome in a genetic study. Figure 2 (adopted from Manolio et al. 2009 [25] and McCarthy et al. 2008 [26]) shows the relationship between potential effect size and allele frequency. Whereas Figure 3 (adopted from Gloyn et al. 2010 [27]) shows what genetic variants that are detected with three primary strategies (genome-wide linkage, targeted sequencing and genome-wide association). Genome-wide linkage analyses can only be performed in cohorts/collections of biologically related individuals and are preferred when searching for rare variants (allele frequency <0.3 %) expected to have large effect size (odds ratio (OR)>5). Whereas, genetic association analyses in case-control or cohort studies, are better suited to identify common variants (allele frequency >5 %) of modest effect size (OR<2).

The presented cohort is family-based and well suited for genome-wide linkage analysis. Other research strategies such as targeted resequencing and family-based association analyses are also possible. The present cohort consists of several nuclear families, and genetic studies of these are likely to represent a more homogeneous and limited set of causative genes and pathways. In addition, we have performed thorough phenotyping of the cohort (published in Thomsen et al. 2015 [28]). These features are likely to enhance statistical power for gene discovery in our cohort. Furthermore, significant heritability estimates in the cohort support that there is an increased susceptibility of preeclampsia [28]. High heritability implying a strong correlation between phenotype and genotype makes it easier to detect loci with an effect on the trait. Heritability denotes the proportion of phenotypic variance explained by additive genetic effects.
However, further research is needed as heritability does not provide information about the genetic architecture.

2. Why were the p-values chosen that were used for comparisons? Justifications are needed as they seem arbitrary (especially p<0.01). Why were the variables compared chosen? Were these of a priori interest based on previous knowledge, for example?

Author’s response:

The p-values given in the text were chosen to point out that the results were statistically significant at the given significance level. In addition, the exact p-values were given in table 4. In the revised manuscript the exact p-values are given both in the text and in table 4.

Yes, the variables compared were chosen based on previous knowledge. It is well known that women experiencing preeclamptic pregnancies are at an increased risk of labor and delivery complications. Since there is no curative treatment when the maternal health condition is threatened due to preeclampsia, birth has to be induced or in acute situations cesarean section is the outcome. This often leads to preterm delivery. It is shown that a very low or high maternal age is associated with increased risk of preeclampsia.

Results:

1. Suggest referring to invited women as eligible women. To do otherwise begs the question why all eligible women were not invited, however, they were.

Author’s response:

We want to thank the reviewer for this suggestion. We have made this revision in the resubmitted manuscript.

2. Refer to flow diagram in this section.

Author’s response:

We acknowledge the suggestion, and have made a reference to the flow diagram in the results chapter in the revised manuscript.

3. Low participation rates must be considered a limitation of the study -- can result in selection bias. Should be discussed in Discussion section.

Author’s response:

We agree with the reviewer that low participation rates should be considered a limitation of the study as the sample size affects the power to detect the effect of genes. In the presented cohort the total participation rate for all eligible index women was 51.6% (220/426). Participation was entirely voluntary and only the index women were personally invited to the study by the
hospitals. The recruitment of the family members depended fully on the index women’s motivation to participate in the study. In addition, family relations may be complex in some families. In our study we asked the index women to recruit the father(s) of their children since we also want to study the paternal contribution to development of preeclampsia. This might have reduced the participation rate among index women that no longer are in a relationship with the father of their children.

Regarding selection bias, one may speculate that index women and families being severely affected by preeclampsia were more positive to participation, and thus might be overrepresented in the cohort. However, regarding the search for genetic factors with large effect size this might in fact be an advantage. Because one assumes that there is a strong relationship between seriousness of disease in a family and the contribution of genetic factors in development of the disease.

4. Was there any missing data? If so, how was it handled? What is the plan for missing data in future analyses?

Author’s response:

Yes, some data is missing (see table 4). For example: placenta weight and birth weight was not available for all women. Index women with missing data were excluded from the statistical analyses when comparing preeclamptic with non-preeclamptic pregnancies.

A linkage with the Medical birth registry of Norway is planned. Hopefully we are able to complement some of the missing data by this linkage. However, this is not possible for women who gave birth before 1967 as the registry was not established until that year.

Another way of complementing missing data could be to collect data from medical records. However, this has not been planned.

Discussion:

1. (See above).

2. Comment on the generalizability of results from this study population.

Author’s response:

The participants in this cohort are recruited from a general population representing four of 20 Norwegian counties. Genetic findings in this study population may be specific to the Norwegian population or to populations of European ethnicity. Some genetic findings may also be specific to a smaller geographic region (a Norwegian county etc.) or even family specific. Regardless of this we anticipate that all genetic findings may help us pinpoint pathways or biological processes that are relevant and important for development of preeclampsia in general.