Dear Editors,

We are hereby submitting the manuscript “…”.

Large birth-cohort studies with voluntary participation play an important role in research on the aetiology of mental health and other illnesses. Because self-selection into studies and loss to follow up might lead to biased exposure outcome associations, a number of highly cited articles has examined selection bias in large cohort studies. Even though such studies investigated a selected set of exposures-outcome associations, their results are often cited as evidence against selection bias of un-related exposure-outcome associations in the same cohort.

For example, around 50% of the 50 most cited papers that cite Nilsen et al.’s (2009) article on self-selection bias in the Norwegian Mother and Child Cohort Study (MoBa), do so to argue that exposure outcome associations they report do not suffer from selection bias. Hence, the motivation for the submitted manuscript is to highlight that general statements about bias due to self-selection and loss to follow up are not possible for entire cohort studies. We believe that our research is of high relevance for epidemiological research using cohort studies, because the validity of such research can be improved by better controlling bias from self selection into studies and loss to follow up.

In particular, we propose to explicitly base examinations of selection bias on the structural approach of selection bias by (1) using directed acyclic graphs (DAG) to evaluate if selection bias is likely and (2) compare inverse probability of participation weighted and un-weighted association-estimates to estimate the size of selection bias.

The key innovations in our article are to (a) propose to exploit the increased availability of GWAS summary results to calculate genetic correlations as a way to evaluate the presence of common unobserved causes that is central to the manifestation of selection bias, (b) obtain estimates of bias and its highest density intervals by jointly estimating weighted and un-weighted association estimates in a Bayesian framework, and (c) show how a simple decision tree can be used to identify potential for selection bias without formulating a complete DAG.

We illustrate this approach by examining selection bias for associations between exposures and ADHD in MoBa. Our results indicate the presence of selection bias for most examined associations. More generally, genetic correlations between education--a key indicator of study participation--and mental health suggest that selection bias should be a concern for studies investigating mental health with non-representative samples.

The manuscript concludes by discussing selection bias and its relationship to representativeness, and by emphasising, that selection bias is not a property of entire studies, but rather a property of exposure-outcome associations.