nature research

Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits
Corresponding author name(s): Andries T. Marees & Eske M. Derks

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:
27th May 2020

Dear Dr Marees,

Thank you once again for your manuscript, entitled "Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits," and for your patience during the peer review process.

Your manuscript has now been evaluated by 3 reviewers, whose comments are included at the end of this letter. Although the reviewers find your work to be of interest, they also raise some important concerns. We are interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised manuscript before we make a decision on publication. Specifically, we would expect your revision to address all issues raised by the reviewers through additional analyses, clarification, and substantial rewriting of the manuscript.

Additionally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or

unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within three months. We understand that the COVID-19 pandemic is causing significant disruption for many of our authors and reviewers. If you cannot send your revised manuscript within this time, please let us know - we will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

• Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,

Charlotte Payne
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Genetic epidemiology of mental health

Reviewer #2: MR analysis of mental health traits

Reviewer #3: MR analysis of mental health traits; Genomic SEM; Genetic epidemiology of mental health

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
In the study, Marees et al. used a recently developed method, called Genomic SEM, to investigate the impact of genetic SES variance on the SNP-based heritability of 16 indicators of mental health and on the pattern of genetic correlations across these traits. Some conclusions are pretty striking, such as "attenuation of SNP-based heritability of mental health trait through SES". Overall, this study is a methodologically sound. However, that are multiple papers that have investigated the genetic correlation between psychiatric disorders and different traits.

Major issue
1. Introduction should be more concise. For example, the authors use a lot of space to describe the potential genetic correlation among psychiatric disorders, as well as the potential effect of low SES on psychiatric disorder susceptibility.

2. Why did the authors select these three SES indicators only? The study can be more powerful if investigation of the effect were performed for all those SES indicators on psychiatric disorder, if the summary statistics available?

3. The heritability estimates analyses showed that the SNP-based heritability of those three SES indicators are only between 0.03-0.09. They are more likely environmental factors than genetic traits. Therefore, they may contribute little to other traits's (such as psychiatric disorders) susceptibility?

4. The rationale why these three SES indicators were combined as the latent SES factor is not sufficiently described or documented.

Minor issue:
1. Full name of "SEM" should be given in the abstract section to be more clear.

2. Substantial negative genetic correlations (rg< - 0.4) of the latent SES
factor were found with ADHD, anxiety disorder, major depression, smoking initiation, cigarettes smoked per day, and smoking cessation, whereas substantial positive genetic associations (rg> 0.4) were found with frequency of alcohol consumption and age at smoking initiation (i.e. genetic variants underlying younger age of smoking were associated with genetic variants for lower SES).
Here, what's difference between smoking initiation and age at smoking initiation?

Reviewer #3:
Remarks to the Author:
In this manuscript, the authors present a novel application of the recently developed Genomic SEM method to assess the level of genetic overlap across 16 mental health traits (9 psychiatric disorders and 7 substance use phenotypes) and explore the influence of SES-associated genetic variation on the pattern of shared heritability. Whereas previous studies have shown very strong genetic correlations between mental health traits, the authors indicate that previous findings are likely to have been

biased by genetic confounding due to SES, and demonstrate significant changes in patterns of genetic correlations after partialling out SES-associated genetic variation. Overall, I think the manuscript is very clear and well written, and would offer an exciting and novel contribution to the literature. I have a few comments and questions:

1) The authors mention that PTSD was excluded from the analysis because it lacked power. How was adequate power demonstrated for the other traits, and could some power calculations be shown?
2) There is some repetition of the description of the method for partialling out SES in “Estimates of genetic variance and genetic correlations” and in “Genetic modelling”. It would make sense to describe the method in its entirety in the “Genetic modelling” section only to improve clarity.
3) The authors state that the LDSC is robust to population stratification, in which case why are they still sensitive to the impact of SES confounding (which has been suggested to be largely driven by population stratification?). The authors could perhaps highlight the limitations of LDSC approach, e.g. in relation to residual geographical clustering which may bias the LDSC estimates
4) In trying to understand the rationale for using a latent variable for SES, rather than the three independent SES traits, could the authors explain why they wanted to estimate the overlapping genetic variance rather than total genetic variance for the SES traits? Would the latter be possible using the approach proposed?
5) In the results, statements around altered genetic correlation with other traits after accounting for latent SES factor e.g. the sentence “genetic correlations with (a subset of) other traits also changed considerably for…” are quite vague. Could these changes be quantified more in terms of magnitude and direction? Which were the largest and smallest trait pair changes when accounting for SES?
6) Is sample overlap in GWAS likely to be a problem for Genomic SEM – could this be discussed?
7) Implications for MR analyses are mentioned briefly in the Discussion. Perhaps this approach should be briefly described for those unfamiliar with it. Is there potential to use the Genomic SEM approach to partial out SES confounding in MR analyses?
8) The authors mention the potentially major limitation that SES represents a collider rather than a confounder in some of the causal pathways described. Is there any way the impact of this could be investigated in the results presented, e.g. comparing changes in estimates with the magnitude/direction of the causal relationship with SES found from other studies?
9) How do the specific findings from this analysis compare with those in the literature e.g. the within vs family difference studies mentioned in the Introduction?

Reviewer #4:
Remarks to the Author:
Dear Editor and authors,
Many thanks for letting me review this interesting manuscript and well-written manuscript.

Key results:
The authors examine the role of the genetic influences underlying socioeconomic status in explaining the genetic correlations between substance use and psychiatric traits. They first show that genetics of SES (as operationalised by a common factor for educational attainment, income, an deprivation index) strongly correlate to genetics of substance use and psychiatric traits. What is more original and the main point of the paper is that accounting for the genetics of SES modifies heritabilities as well as patterns of genetic correlations within substance use and psychiatric traits and between them.

Validity:
The manuscript is clear, doesn't have major flaw, and innovative methods are globally well described and implemented. See below for more specific comments.

Originality and significance:

Findings raise a number of fundamental questions regarding the commonly made interpretations of genetic correlations between traits. SNP heritability estimates and genetic correlations have often been taken at face value, as reflecting either common underlying mechanisms or reciprocal causal relationships (i.e. meaningful pleiotropy). However, more recently, the extent of potential biases on such estimates has been highlighted, notably in within family studies as those cited by the authors. The role of the genetics of SES appears to be central in many of these biases, likely reflecting the importance of SES as a major confound at the phenotypic level. The objective of the authors to examine the role of the genetics of SES for substance use and psychiatric disorders is thus valuable. The authors employ a novel method, genomic SEM, to test whether the correlation between any pair of psychiatric substance use traits is changed after residualizing for SES. Findings show that this is clearly the case, with many genetic correlations being significantly altered by controlling for SES. The authors also show that controlling for SES somewhat change how those traits cluster together. As such, this finding should be of interest to many in the field of behavioural genetics and genetic epidemiology.

Data & methodology:

Overall, methods are adequate and well described. The description of the Monte-carlo approach to test for the difference between the residualized and the non-residualized correlations doesn't appear as commonly used and may warrant additional explanations, possibly in the supplementary material.

In addition, in this kind of studies, many small differences in analytical decisions can have a fairly significant impact on findings. Because this is all based on widely available summary statistics , I think the authors should share their script as part of the supplementary material. It is good practice, in line with open science, so that other researchers in the field can check statistical decisions and build on those findings.

Can you comment on what appears to be paradoxical, in that residualizing for SES seems to affect correlations within substance use cluster traits more than within psychiatric traits? However, clustering is unaffected for substance use, but more affected for psychiatric traits.

Also, it may be worth briefly mentioning the fact that clusters do not comprise traits with homogeneous within-cluster correlations as some 'subclusters' are clearly present (e.g. schizo + BP, or depression + anx). There is a danger of reifying such clusters, as their number and composition often depends on which traits when included in the analysis.

Preregistration: The authors do not appear to have preregistered their study. As mentioned above, sharing scripts would be in line with open science.

Appropriate use of statistics

The authors state that GWAS for post-traumatic stress disorder was not powerful enough, but give no information as to how this was reached. If the authors conducted formal power calculations, these should be summarised. If not, the decision to exclude PTSD should be further justified.

“applying the FDR outcome in five different sets” The phrasing is not clear, why five sets, please rephrase to make that clearer for the reader.

Conclusions:
I think interpretation and conclusions are the key issue for this paper. The authors mostly do a good job in explaining the limitations of their approach. A few developments might further improve the manuscript. For example, we can imagine that ADHD is a causal determinant of educational attainment (considerable evidence in the phenotypic literature in particular regarding the inattentive component). Therefore, adjusting the heritability in ADHD for SES may be the wrong way around… Similarly, educational attainement may in turn impact substance use and anxiety in later life. Such complex causal chains are only mentioned in the limitations as, in this case, controlling for SES may generate bias (e.g. if collider biais). The possibility of reverse causation and collider biais should be further developed in the main body of the discussion, together with other mechanisms that may explain why correlations change when controlling for SES.

As the authors say, the methods "do not allow us to separate between these alternative mechanistic explanations." The authors might need to spend more time justifying the interest of their approach We are left with a new set of correlations that are not necessarily more meaningful than the first one, in the absence of causal knowledge regarding the relationships between these variables (confounders, mediators, colliders). Can the authors further develop why this approach is interesting despite this key limitation.

In the conclusion, the authors say that “but the remaining genetic variance will be more trait-specific, and possibly more relevant from a clinical perspective” But is that the case if we don't know the underlying causal relationships as mentioned above? Any implications for functional analyses aiming to discover biological pathways?

The authors suggest that computing PRS with summary statistics residualized for SES may be beneficial. But in this case, why not simply generate a polygenic score for SES separately and implement a multivariable model with multiple polygenic scores, including for SES, as predictors. There are many papers using multipolygenic score approaches. Any reason why residualised scores would be better? Similarly, the paragraph on MR may want to refer to multivariable MR as recent papers have adjusted for educational attainment in an attempt to reduce confounding, similar to what the authors are proposing. e.g.
https://www.medrxiv.org/content/10.1101/2020.04.16.20067918v1.full.pdf
And author may also want to mention within-family MR. e.g.
https://pubmed.ncbi.nlm.nih.gov/31647093/

Suggested improvements:

Figure 4 is not easy to read, as we have to look for the adjusted and the unadjusted in the two parts

of the correlation matrix, which is cumbersome. Anyway that this could be presented as the figure of heritability, where before and after can be directly compared?

References:

ADHD includes impulsivity, which is often conceptualised as part of externalizing spectrum, which also includes substance use. There is a considerable literature on the subject of ADHD and substance use, which may need to be referred too e.g. PMID: 18020714

A few references are missing. For example the first two sentences of introduction have no reference, authors may want to refer to the role of mental health and substance use in the global burden of disease. In the discussion, the following sentence "investigated using genetic causal modelling" could do with some general references on genetics and causal inference. Please make sure all statements are justified by appropriate references.

Clarity and context:

Fine. Appreciated the "Western, Educated, Industrialized, Rich, and Democratic (WEIRD) :)"

Author Rebuttal to Initial comments

Dear editor,

We like to thank you for the opportunity to revise and resubmit our manuscript "*Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits*". We are pleased with the thorough reviews and excellent recommendations by the reviewers which we believe have significantly improved our manuscript. Please find our replies to the reviewers' comments below. Changes in the revised manuscript are printed in bold. Thank you again for considering our manuscript for publication in Nature Human Behavior.

Reviewer #1:

Remarks to the Author:

In the study, Marees et al. used a recently developed method, called Genomic SEM, to investigate the impact of genetic SES variance on the SNP-based heritability of 16 indicators of mental health and on the pattern of genetic correlations across these traits. Some conclusions are pretty striking, such as "attenuation of SNP-based heritability of mental health trait through SES". Overall, this study is a methodologically sound. However, that are multiple papers that have investigated the genetic correlation between psychiatric disorders and different traits.

Major issue

1. Introduction should be more concise. For example, the authors use a lot of space to describe the potential genetic correlation among psychiatric disorders, as well as the potential effect of low SES on psychiatric disorder susceptibility.

We have substantially shortened the introduction, particularly the parts about previously estimated genetic correlations. However, we believe that discussing the findings of these studies provides an important overview of what is already known regarding genetic correlations among mental health traits and creates a theoretical framework for our paper.

2. Why did the authors select these three SES indicators only? The study can be more powerful if investigation of the effect were performed for all those SES indicators on psychiatric disorder, if the summary statistics available?

SES is generally defined as one's access to financial, social, cultural, and human capital resources, which are measured by educational attainment, occupational status, and household or family income (see this report on the definition and measurement of SES by the US national center of educational statistics: https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf). As there is no powerful GWAS on occupational status (which also is a non-ordinal variable, so not suited for a GWAS in the traditional sense), we limited the SES measure to educational attainment and income, and added the Townsend index, which is used in the UK as a measure of regional SES and has powerful GWAS summary statistics available.

3. The heritability estimates analyses showed that the SNP-based heritability of those three SES indicators are only between 0.03-0.09. They are more likely environmental factors than genetic traits. Therefore, they may contribute little to other traits' (such as psychiatric disorders) susceptibility?

The SNP-based heritability estimates of the three SES indicators indeed range between 3% and 9%. These estimates are indeed in the lower range in comparison with the SNP-based heritabilities of the mental health traits (i.e., ranging between 5% and 25% before partialling out SES), but they are not markedly different. It should be noted that SNP-based heritabilities are not an indication of the magnitude of overall genetic effects (and therefore also not of the environmental effects), as SNP-based heritabilities are usually

much lower than the twin-based heritability of SES, which is estimated at 52%. In this study, our analyses are restricted to modelling the genetic covariances across traits as captured by genome-wide SNPs. The fact that we find attenuation of SNP-based heritability of mental health traits after controlling for genetic variance associated with SES, suggests that there is a contribution of SES, despite the relatively low SNP-based heritabilities.

4. The rationale why these three SES indicators were combined as the latent SES factor is not sufficiently described or documented.

We added the following sentences to the introduction describing a more detailed rationale behind combining the three SES indicators:
"SES represents an individual's or family's access to financial, social, cultural, and human capital resources, which are generally measured by educational attainment, occupational status, and household or family income[25]. In Great Britain, the authorities generally measure regional differences in SES with a composite of index variables called the Townsend Index (TI)[26]. We will use GWAS results of this latter indicator variable[21], and two other main components of SES, namely educational attainment (EA)[27] and household income (HI)[21]. First, we will explore the role of the three individual components separately. Next, we will use factor analysis to construct a latent variable that combines the three aspects of SES, because a factor that combines different indicators into a single composite may be a better indicator of SES and discard indicator variability that is not directly related to SES[28,29]. For this, we will apply a recently developed multivariate method called genomic structural equation modelling (genomic SEM[30]), …"

Minor issue:
1. Full name of "SEM" should be given in the abstract section to be more clear.

We have now provided the full name of SEM (Structural Equation Modelling) in the abstract.

2. Substantial negative genetic correlations (rg< - 0.4) of the latent SES factor were found with ADHD, anxiety disorder, major depression, smoking initiation, cigarettes smoked per day, and smoking cessation, whereas substantial positive genetic associations (rg> 0.4) were found with frequency of alcohol consumption and age at smoking initiation (i.e. genetic variants underlying younger age of smoking were associated with genetic variants for lower SES). Here, what's difference between smoking initiation and age at smoking initiation?

'Smoking Initiation' is a binary variable indicating whether someone has ever smoked (1) or not (0). 'Age at smoking initiation' is a variable capturing the age at which smokers started smoking (i.e. this variable is only measured in lifetime smokers). The variables are described in the methods section, page 5, and we now clarified the "Smoking Initiation" variable in the results section as well on page 8: "*Substantial negative genetic correlations (rg < - 0.4) of the latent SES factor were found with ADHD, anxiety disorder, major depression, smoking initiation (i.e., ever smoked cigarettes), cigarettes smoked per day, and smoking cessation*"

Accordingly, you would expect genetic correlations of these traits with other variables to be in opposite direction.

Reviewer #3:
Remarks to the Author:
In this manuscript, the authors present a novel application of the recently developed Genomic SEM method to assess the level of genetic overlap across 16 mental health traits (9 psychiatric disorders and 7 substance use phenotypes) and explore the influence of SES-associated genetic variation on the pattern of shared heritability. Whereas previous studies have shown very strong genetic correlations between mental health traits, the authors indicate that previous findings are likely to have been biased by genetic confounding due to SES, and demonstrate significant changes in patterns of genetic correlations after partialling out SES-associated genetic variation. Overall, I think the manuscript is very clear and well written and would offer an exciting and novel contribution to the literature. I have a few comments and questions:

1) The authors mention that PTSD was excluded from the analysis because it lacked power. How was adequate power demonstrated for the other traits, and could some power calculations be shown?

The inclusion criterium for this study was a SNP-based heritability Z-score (h^2_{SNP} Z-score) larger than 2, a threshold also used in a recent paper in Nature Genetics (see: https://pubmed.ncbi.nlm.nih.gov/31427789/). We calculated this score, using results obtained with LD score regression, by dividing the SNP-based heritability estimate by its standard error. The h^2_{SNP} Z-score of the PTSD GWAS was 1.0 and, therefore, we did not include this trait in our study.

We have now clarified this exclusion on page 5:
"… had a SNP-based heritability Z-score below our inclusion threshold of 2^{40}"

2) There is some repetition of the description of the method for partialling out SES in "Estimates of genetic variance and genetic correlations" and in "Genetic modelling". It would make sense to describe the method in its entirety in the "Genetic modelling" section only to improve clarity.

The reviewer is correct that there was some repetition in these 2 sections. As suggested by the reviewer they are now combined into one section and the repetitions are omitted (pages 5-6).

3) The authors state that the LDSC is robust to population stratification, in which case why are they still sensitive to the impact of SES confounding (which has been suggested to be largely driven by population stratification?). The authors could perhaps highlight the limitations of LDSC approach, e.g. in relation to residual geographical clustering which may bias the LDSC estimates.

We are not aware of any studies showing that the genetic signals for SES-related traits are substantially confounded by population stratification, and even if they were, LDSC regression should be able to separate confounding due to population stratification from the signals that we use. As you probably know, LDSC regression is a method that distinguishes inflation due to bias caused by cryptic relatedness or population stratification, i.e., ancestry differences, and inflation due to causal polygenic effects (see: https://www.nature.com/articles/ng.3211). It does so by leveraging the fact that SNPs that are in LD with more variants are more likely to be correlated with causal genetic variants when the trait is highly polygenic (which is the case for the traits we included in our study). This results in a relationship between the number of SNPs that are in LD with a SNP (i.e., its "LD Score") and GWAS signal due to polygenic effects. Because there is no such relationship between the LD score and ancestry differences, LD score regression can detect whether an inflation of the GWAS signal is due to population stratification by looking at the strength of the association between the LD Score and the GWAS test statistic. There is no reason to assume that the association between SES differences and ancestry within datasets of European descent are of a magnitude that would substantially confound the GWAS signal, and if there was such an association, the correction for population stratification within the GWAS should take care of that. And even if the correction for population stratification within the GWAS was not sufficient, LD Score regression makes sure that we analyse only the part of the GWAS signal that is due to polygenic effects.

It is possible however (likely even) that the polygenic effects are overestimated in SES-related GWASs due to environmental effects that are correlated with the polygenic effects for reasons explained in Abdellaoui et al. (2019) and Kong et al. (2018) (gene-environment correlations due to geographic clustering and "genetic nurturing" respectively). We explained these possible confounders in our introduction and discussion. We think it is likely that we remove some of this inflation by extracting the SES part of the GWAS signal, as that is the most likely part of the signal that suffers from inflation due to these gene-environment correlations. This is indeed an important motivation for our partialling out of SES genetic variance.

4) In trying to understand the rationale for using a latent variable for SES, rather than the three independent SES traits, could the authors explain why they wanted to estimate the overlapping genetic variance rather than total genetic variance for the SES traits? Would the latter be possible using the approach proposed?

Our *a priori* vision was based on the extant literature arguing that composite variables are better predictors than single indicators (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3907681/; https://psycnet.apa.org/record/1982-24392-001). Removing the total variance would be possible, but is in our view not a valid approach since indicator variables consist of more variation than that related to the factor (in the present case SES). Correcting for all variables would result in overcorrection. Nevertheless, we additionally provide models where we remove the variance of the single SES indicators. We have adjusted the introduction to provide a better rationale for using a factor approach (see also reviewer 1 comment 4 above.)

In the context of a genetic factor analysis as we performed here, partialling out EA variance appeared to show larger changes in h^2 and r_G than the SES factor, as we noted in the original text, page 9: "However, many more changes are significant after partialling out educational attainment than when removing the effects of household income, Townsend Index, or the latent SES factor." Nevertheless, this could constitute an overcorrection by a well powered GWAS. We therefore chose to provide all available information (i.e. correcting for the SES factor as well as all indicator variables separately).

5) In the results, statements around altered genetic correlation with other traits after accounting for latent SES factor e.g. the sentence "genetic correlations with (a subset of) other traits also changed considerably for…" are quite vague. Could these changes be quantified more in terms of magnitude and direction? Which were the largest and smallest trait pair changes when accounting for SES?

We have substantially re-written this paragraph to describe the changes more clearly. We did not describe the magnitude and direction for each individual change, however, as that would be too much text for data that we have visualized in Figure 4 (note that this was previously a Supplementary Figure, which we have now moved to the main text as a response to your comment and that of reviewer 4. This helps better visualize the magnitude and direction of the changes). In the text we now emphasize the three groups of most notable (largest) changes, namely (1) ADHD with substantially lower genetic correlations with several substance use and psychiatric outcomes, (2) anxiety and MDD with lower genetic correlations with substance use outcomes, and (3) weaker genetic correlations between the substance use outcomes (page 9):

"Notably, genetic correlations were substantially weakened between ADHD and several substance use and psychiatric disorders, consistent with the strong reduction in SNP-based heritability of ADHD in Figure 3. Other changes include weakened genetic correlations between anxiety and several substance use outcomes (age at smoking initiation, smoking initiation, frequency of alcohol consumption), and a weakened genetic correlations between major depression and several substance use outcomes (smoking cessation, cigarettes per day, age at smoking initiation, smoking initiation, frequency of alcohol consumption). Finally, the genetic correlations between the different substance use outcomes generally showed large decreases."

6) Is sample overlap in GWAS likely to be a problem for Genomic SEM – could this be discussed?

Sample overlap is no problem for Genomic SEM analysis. In the methods section, we have now added (page 5): "*Genomic SEM is robust to confounding due to population stratification, and against full or partial sample overlap and cryptic relatedness across GWAS samples.*"

7) Implications for MR analyses are mentioned briefly in the Discussion. Perhaps this approach should be briefly described for those unfamiliar with it. Is there potential to use the Genomic SEM approach to partial out SES confounding in MR analyses?

We have included a brief explanation of MR (page 15): "*Mendelian Randomization uses germline genetic variants as an instrument for the exposure to modifiable risk factors and can be used to investigate causality between traits*[67]".

Thank you. Your suggestion to use Genomic SEM to partial out SES is interesting, and we have added this possibility to the discussion (page 15). "*The results from our Genomic SEM analysis can be used*

to construct genetic instruments for mental health traits that control for the influence of SES allowing investigation of causal relations limiting the role of SES as a potential confounder"

8) The authors mention the potentially major limitation that SES represents a collider rather than a confounder in some of the causal pathways described. Is there any way the impact of this could be investigated in the results presented, e.g. comparing changes in estimates with the magnitude/direction of the causal relationship with SES found from other studies?

Our genetic correlation analyses do not support inference on the direction of causality and, therefore, conclusions about collider effects cannot be made. For a collider to exist, however, we would have to establish that both mental health traits affect SES. This possible avenue is discussed in the section where we discuss the application of MR. We also added a fuller exploration of the collider vs. confounder issue in response to reviewer 4. We now added the following text to the Discussion (page 13): "SES is usually considered a confounder in epidemiological research, warranting removal of its variability to investigate 'true' relations between variables. However, it is possible (and perhaps even likely) that SES sometimes acts as a collider, or shows bidirectional causal relations with mental health traits. For example, ADHD is known to affect educational performance and likely also HI[62-64]. In such cases, it is possible that the removal of SES variance biases the genetic correlation estimates. Nevertheless, the change in genetic correlations observed across our traits reveals that interpreting the genetic overlap at face value – as has been done in many cases[2,3] – may be too simple. The fact that so many of the SNP heritabilities and genetic correlations were affected suggests that SES is a central and important factor, either as cause or as consequence. Unfortunately, we cannot disentangle causal direction with these results; however, our results set the stage for such modelling, pointing to SES as an unneglectable factor in psychiatric genetics research."

9) How do the specific findings from this analysis compare with those in the literature e.g. the within vs family difference studies mentioned in the Introduction?

We added this comparison to the discussion, page 11. "*Our results showed attenuated SNP-based heritability estimates after controlling for SES, with the strongest effect observed for ADHD (43% attenuation). This observation is in line with the results of within-family GWAS that showed a reduction of genetic effects within families compared to between families*[16]*. For example, Selzam et al.*[16] *reported that the polygenic risk score for ADHD is a significantly stronger predictor between families (beta=-0.18) than within families (beta=-0.06); this difference disappeared after controlling for SES.*"

Reviewer #4:
Remarks to the Author:
Dear Editor and authors,
Many thanks for letting me review this interesting manuscript and well-written manuscript.

Key results:
The authors examine the role of the genetic influences underlying socioeconomic status in explaining the genetic correlations between substance use and psychiatric traits. They first show that genetics of SES (as operationalised by a common factor for educational attainment, income, a deprivation index) strongly correlate to genetics of substance use and psychiatric traits. What is more original and the main point of the paper is that accounting for the genetics of SES modifies heritabilities as well as patterns of genetic correlations within substance use and psychiatric traits and between them.

Validity:
The manuscript is clear, doesn't have major flaw, and innovative methods are globally well described and implemented. See below for more specific comments.

Originality and significance:

Findings raise a number of fundamental questions regarding the commonly made interpretations of genetic correlations between traits. SNP heritability estimates and genetic correlations have often been taken at face value, as reflecting either common underlying mechanisms or reciprocal causal relationships (i.e. meaningful pleiotropy). However, more recently, the extent of potential biases on such estimates has been highlighted, notably in within family studies as those cited by the authors. The role of the genetics of SES appears to be central in many of these biases, likely reflecting the importance of SES as a major confound at the phenotypic level. The objective of the authors to examine the role of the genetics of SES for substance use and psychiatric disorders is thus valuable.

The authors employ a novel method, genomic SEM, to test whether the correlation between any pair of psychiatric substance use traits is changed after residualizing for SES. Findings show that this is clearly the case, with many genetic correlations being significantly altered by controlling for SES. The authors also show that controlling for SES somewhat change how those traits cluster together. As such, this finding should be of interest to many in the field of behavioural genetics and genetic epidemiology.

We are pleased and grateful for the reviewer's extensive summary and the reviewer's positive words, in supports of the rationale behind our article and the importance of our findings.

Data & methodology:

Overall, methods are adequate and well described. The description of the Monte-Carlo approach to test for the difference between the residualized and the non-residualized correlations doesn't appear as commonly used and may warrant additional explanations, possibly in the supplementary material.

We have added a more detailed description, including a figure, in the supplementary information and we will provide the full code via GitHub (https://github.com/MareesAT/Genetic-correlates-of-socio-economic-status-influence-the-pattern-of-shared-heritability-across-ment).

In addition, in this kind of studies, many small differences in analytical decisions can have a fairly significant impact on findings. Because this is all based on widely available summary statistics, I think the authors should share their script as part of the supplementary material. It is good practice, in line with open science, so that other researchers in the field can check statistical decisions and build on those findings.

Thank you for this suggestion. We agree with the reviewer that it is good practice to share scripts. Therefore, all relevant scripts have now been uploaded to: https://github.com/MareesAT/Genetic-correlates-of-socio-economic-status-influence-the-pattern-of-shared-heritability-across-ment

Can you comment on what appears to be paradoxical, in that residualizing for SES seems to affect correlations within substance use cluster traits more than within psychiatric traits? However, clustering is unaffected for substance use, but more affected for psychiatric traits.

SES traits clearly affected SU traits more than most psychiatric traits. This is evident from the within cluster genetic correlations (edge weights) that were strongly affected. However, the between cluster r_G was also decreased. Clustering is based on relative positions, meaning that within-cluster absolute edge weights may decrease but as long as the effect is rather homogenous across all the traits in a cluster and the between-cluster edge weights also decrease systematically, the original cluster stays intact. Psychiatric traits changed differentially; some being affected more than others in their relation to other psychiatric traits. This is the source for the cluster reorganization. We now note this in the discussion section, page 12:

"*Paradoxically, the substance use cluster remained largely intact while the psychiatric clusters changed in composition. This can be attributed to the fact that the effect of partialling out SES had relatively consistent effects on interrelations between substance use traits, whereas it had strong effects on some genetic associations within the psychiatric clusters (either becoming stronger or weaker) but not on others (also see Figure 4).*"

Also, it may be worth briefly mentioning the fact that clusters do not comprise traits with homogeneous within-cluster correlations as some 'subclusters' are clearly present (e.g. schizo + BP, or depression + anx). There is a danger of reifying such clusters, as their number and composition often depends on which traits when included in the analysis.

We fully agree with this account, and we hope that we provided some ground for the assertion that clustering depends on the available data sources. We did not, however, investigate how clusters change with stepwise addition or removal of specific traits and disorders. Here, we aimed to adjust the genetic correlation matrix by removal of SES variance, a possible confounder. Clearly, some of the genetic clustering was driven by the effect that SES has on some of the mental health traits.

The stability of genetic factors (or clusters) is still open and unexplored but is very important given the investigations being performed by large consortia such as the Brainstorm Consortium and the cross-disorder workgroup of the psychiatric genomics consortium. The fact that these clusterings are affected by factors like SES gives room for thought.

Preregistration: The authors do not appear to have preregistered their study. As mentioned above, sharing scripts would be in line with open science.

All scripts have now been uploaded to:
https://github.com/MareesAT/Genetic-correlates-of-socio-economic-status-influence-the-pattern-of-shared-heritability-across-ment

Appropriate use of statistics

The authors state that GWAS for post-traumatic stress disorder was not powerful enough, but give no information as to how this was reached. If the authors conducted formal power calculations, these should be summarised. If not, the decision to exclude PTSD should be further justified.

We excluded PTSD because this GWAS did not SNP-based heritability Z-score higher than 2.

The inclusion criterium for this study was a SNP-based heritability Z-score (h^2_{SNP} Z-score) larger than 2, a threshold also used in a paper recently published in Nature Genetics (see: https://pubmed.ncbi.nlm.nih.gov/31427789/). We calculated this score, using results obtained with LD score regression, by dividing the SNP-based heritability estimate by its standard error. The h^2_{SNP} Z-score of the PTSD GWAS was 1.0, therefore we did not include this trait in our study.

We have now clarified this exclusion on page 5:
"*… had a SNP-based heritability Z-score below our inclusion threshold of 2*"

"applying the FDR outcome in five different sets" The phrasing is not clear, why five sets, please rephrase to make that clearer for the reader.

We have now clarified this as follows (page 6): *"In assessing whether a genetic correlation was significant, we used a Benjamini-Hochberg FDR correction to account for multiple testing: the 120 genetic correlations between all 16 mental health traits before partialling out genetic SES variance, and then for the 120 genetic correlations between all 16 mental health traits after partialling out the genetic SES variance, (i.e., applying the FDR outcome in five different sets of 120 p-values, the five sets being for the uncorrected genetic correlations and the genetic correlations corrected for EA, HI, TI, and the SES factor)."*

And Page 7: *"Benjamini-Hochberg FDR was used to correct for multiple testing for the changes in 120 genetic correlations between all 16 mental health traits after partialling out the SES-factor (i.e., applying the FDR outcome in four different sets of 120 p-values, the four sets being for the change in genetic correlations after correcting for EA, HI, TI, and the SES factor)."*

Conclusions:
I think interpretation and conclusions are the key issue for this paper. The authors mostly do a good job in explaining the limitations of their approach. A few developments might further improve the manuscript. For example, we can imagine that ADHD is a causal determinant of educational attainment (considerable evidence in the phenotypic literature in particular regarding the inattentive component). Therefore, adjusting the heritability in ADHD for SES may be the wrong way around… Similarly, educational attainement may in turn impact substance use and anxiety in later life. Such complex causal chains are only mentioned in the limitations as, in this case, controlling for SES may generate bias (e.g. if collider bias). The possibility of reverse causation and collider bias should be further developed in the main body of the discussion, together with other mechanisms that may explain why correlations change when controlling for SES.

Our research was prompted by the fact that genetic correlations across mental health traits showed highly recurrent patterns. These patterns were consistent with a SES factor influencing the genetic correlations between, for example, alcohol use variables and certain psychopathology. However, no study to date had systematically explored the influence of SES on genetic correlations. For this reason, we argue

that it is important to establish whether SES genetically correlates with mental health traits, and whether SES explains part of the genetic overlap across those mental health trait pairs.

We agree that collider bias is a very important aspect in (genetic) epidemiology. If SES acts as a confounder, SES should not be ignored, as neglecting this confounder will bias parameter estimates. If SES acts as a collider, correcting for SES will bias parameter estimates. Across the whole set of traits, we agree that SES is likely to show a complex pattern of interaction with mental health traits, with variable levels of magnitude of influence and variable directions of causation. Our results showed that many SNP-h^2 estimates and genetic correlations changed after including SES into the model (i.e., partialling out SES genetic variance). This fact by itself shows that SES is an important factor, either as cause or consequence. But we do note that the current approach cannot distinguish between the alternatives of confounder, mediator, and collider. We have extended the discussion as follows with a full paragraph:

"SES is usually considered a confounder in epidemiological research, warranting removal of its variability to investigate 'true' relations between variables. However, it is possible (and perhaps even likely) that SES sometimes acts as a collider, or shows bidirectional causal relations with mental health traits. For example, ADHD is known to affect educational performance and likely also HI[62-64]. In such cases, it is possible that the removal of SES variance biases the genetic correlation estimates. Nevertheless, the change in genetic correlations observed across our traits reveals that interpreting the genetic overlap at face value – as has been done in many cases[2,3] – may be too simple. The fact that so many of the SNP heritabilities and genetic correlations were affected suggests that SES is a central and important factor, either as cause or as consequence. Unfortunately, we cannot disentangle causal direction with these results; however, our results set the stage for such modelling, pointing to SES as an unneglectable factor in psychiatric genetics research."

As the authors say, the methods "do not allow us to separate between these alternative mechanistic explanations." The authors might need to spend more time justifying the interest of their approach We are left with a new set of correlations that are not necessarily more meaningful than the first one, in the absence of causal knowledge regarding the relationships between these variables (confounders, mediators, colliders). Can the authors further develop why this approach is interesting despite this key limitation.

We would not want to claim that the partial estimates are always the 'better set of genetic correlations' and point estimates of SNP-based h^2, although we believe that in some cases they are. To acknowledge the complexities regarding the interpretation of our results, we consistently use the phrase 'partialling out' and not e.g. 'correcting for SES'. Our results show that ignoring SES is definitely a potential source of error. The crucial information lies in the comparison between the two sets of estimates. First of all, the fact that estimates change means that some causation is present, which can then be investigated in causal models. The fact that it happens in so many cases means that SES an important factor, either as cause or as consequence. The next step could be MR investigations that establish the causal relation between SES and for example ADHD, MDD, and AN, which would elucidate whether collider bias plays a role in r_G estimates.

These observations are captured in the additional text provided in the previous point.

In the conclusion, the authors say that "but the remaining genetic variance will be more trait-specific, and possibly more relevant from a clinical perspective" But is that the case if we don't know the underlying causal relationships as mentioned above? Any implications for functional analyses aiming to discover biological pathways?

As above, we argue that establishing change in genetic correlation after partialling out SES genetic variation is informative in itself. Parsing the variance of a mental health trait will help identify SNPs that are related to non-SES biological mechanisms, and limit polygenic scores to the non-SES part. Although partialling out genetic SES variance will be crucial in many cases, it may not do so in all cases, hence our rephrasing of the sentence to highlight this uncertainty: "Partialling out SES may reduce the predictive power of PRS analyses as it removes part of the genetic variance, but the remaining genetic variance **may** more specifically reflect the phenotype of interest, which would increase their potential clinical utility."

The authors suggest that computing PRS with summary statistics residualized for SES may be beneficial. But in this case, why not simply generate a polygenic score for SES separately and implement a multivariable model with multiple polygenic scores, including for SES, as predictors.

There are many papers using multipolygenic score approaches. Any reason why residualised scores would be better? Similarly, the paragraph on MR may want to refer to multivariable MR as recent papers have adjusted for educational attainment in an attempt to reduce confounding, similar to what the authors are proposing. e.g.
https://www.medrxiv.org/content/10.1101/2020.04.16.20067918v1.full.pdf
And author may also want to mention within-family MR. e.g.
https://pubmed.ncbi.nlm.nih.gov/31647093/

We agree that multipolygenic scores can also be used for the same purpose, i.e. to create polygenic scores for a certain target trait after accounting for confounders (in this case SES). We do not claim that the method we used is better; our approach can be used to create a new set of SNP estimates, where the SNP effects are potentially closer to the biology of the trait of interest as the effects of confounders have been regressed out. The output of this method is versatile; these SNP estimates can be used for several purposes: to run secondary analyses to identify biological mechanism that underlie the trait of interest, to improve MR analyses, and also to create polygenic scores that are less influenced by confounding factors. Accordingly, creating more meaningful PRSs is merely one of the potential applications of this method.

Regarding within-family MR, we agree this is also a very good approach, achieving the same goal (although with the downside that you would need family-based data); we added the following sentence referencing the two suggested papers (page 15). "*An alternative approach would be to conduct within-family MR analyses*[69,70]."

Suggested improvements:

Figure 4 is not easy to read, as we have to look for the adjusted and the unadjusted in the two parts of the correlation matrix, which is cumbersome. Anyway, that this could be presented as the figure of heritability, where before and after can be directly compared?

Thank you for this comment. We had visualized the same information in Figure 4 in a different way in Supplementary Figure 3. We agree that looking up the changes in Figure 4 is more cumbersome than in Supplementary Figure 3. In response to your comment, we have now made Supplementary Figure 3 the new Figure 4. In addition, we will make use of the option of the Nature Human Behaviour journal to have 10 Extended Data Figures. We have thus renamed all of our Supplementary Figures as Extended Data Figures and will no longer have Supplementary Figures. This will make it easier for the reader to access the extra Figures, as Extended Data Figures are part of the pdf of the main text.

References:

ADHD includes impulsivity, which is often conceptualised as part of externalizing spectrum, which also includes substance use. There is a considerable literature on the subject of ADHD and substance use, which may need to be referred too e.g. PMID: 18020714

In the discussion, we included: …."*supporting the previously reported link between externalizing disorders and substance use*[51]" and reference the suggested manuscript.

A few references are missing. For example the first two sentences of introduction have no reference, authors may want to refer to the role of mental health and substance use in the global burden of disease. In the discussion, the following sentence "investigated using genetic causal modelling" could do with some general references on genetics and causal inference. Please make sure all statements are justified by appropriate references.

Additional references have now been included

Decision Letter, first revision:

22nd October 2020

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Marees,

Thank you once again for submitting your revised manuscript, entitled "Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits," and for your patience during the re-review process.

Your manuscript has now been evaluated by Reviewers 1 and 4 from the previous round of review. While Reviewer 3 was unable to re-review, we consulted with Reviewer 4 on your responses to Reviewer 3's comments to ensure their concerns were addressed. In the light of the feedback we received, I am delighted to say that we can in principle offer to publish it. First, however, we would like you to revise your paper to ensure that it complies with our Guide to Authors at http://www.nature.com/nathumbehav/info/gta.

One of the main reasons for delays in formal acceptance is failure to fully comply with editorial policies and formatting requirements. To assist you with finalizing your manuscript for publication, I attach a

checklist that lists all of our editorial policies and formatting requirements. I also attach a template document, which exemplifies our policies and formatting requirements.

Please attend to *every item* in the checklist and upload a copy of the completed checklist with your submission. I have highlighted in the checklist items that require your attention. I also mention here a few points that are frequently missed and can cause delays:

1) Ensure that all corresponding authors have linked their ORCID to their account on our online manuscript handling system. This is very frequently missed and invariably causes delays in formal acceptance.

2) Ensure that you provide all of the materials requested in the attached checklist and below with your final submission. Please note that the Licence to Publish needs to be hand-signed.

Nature Human Behaviour offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

We hope to hear from you within 10 days; please let us know if the revision process is likely to take longer.

To submit your revised manuscript, you will need to provide the following:
• Cover letter
• Point-by-point response to the reviewers (if applicable)
• Manuscript text (not including the figures) in .docx or .tex format
• Individual figure files (one figure per file)
• Extended Data & Supplementary Information, as instructed
• Reporting summary
• Editorial policy checklist
• License to publish
• Third-party rights table (if applicable)
• Suggestions for cover illustrations (if desired)

Consortia authorship:
For papers containing one or more consortia, all members of the consortium who contributed to the paper must be listed in the paper (i.e., print/online PDF). If necessary, individual authors can be listed in both the main author list and as a member of a consortium listed at the end of the paper. When

submitting your revised manuscript via the online submission system, the consortium name should be entered as an author, together with the contact details of a nominated consortium representative. See https://www.nature.com/authors/policies/authorship.html for our authorship policy and https://www.nature.com/documents/nr-consortia-formatting.pdf for further consortia formatting guidelines, which should be adhered to prior to acceptance.

Reviewer Recognition:
In recognition of the time and expertise our reviewers provide to Nature Human Behaviour's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Please use the following link for uploading these materials:
[REDACTED]

If you have any further questions, please feel free to contact me.

With best regards,

Charlotte Payne
Editor
Nature Human Behaviour

Reviewer #1:
Remarks to the Author:
All my concerns have been addressed. I have no more comments.

Reviewer #4:
Remarks to the Author:
The reviewers have appropriately answered all of my comments.

Final Decision Letter:

Dear Dr Marees,

We are pleased to inform you that your Article "Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits", has now been accepted for publication in Nature Human Behaviour.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

Nature Human Behaviour is a Transformative journal and offers an immediate open access option through payment of an article-processing charge (APC) for papers submitted after 1 January, 2021 . In the event that authors choose to publish under the subscription model, Nature Research allows authors to self-archive the accepted manuscript (the version post-peer review, but prior to copy-editing and typesetting) on their own personal website and/or in an institutional or funder repository where it can be made publicly accessible 6 months after first publication, in accordance with our self-archiving policy. Please review our self-archving policy for more information.

Several funders require deposition the accepted manuscript (AM) to PubMed Central or Europe PubMed Central. To enable compliance with these requirements, Nature Research therefore offers a free manuscript deposition service for original research papers supported by a number of PMC/EPMC participating funders. If you do not choose to publish immediate open access, we can deposit the accepted manuscript in PMC/Europe PMC on your behalf, if you authorise us to do so.

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript

submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,

Charlotte Payne
Editor
Nature Human Behaviour