PEER REVIEW HISTORY

BMJ Medicine publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The interplay between inflammatory cytokines and cardiometabolic disease: evidence from Mendelian randomization
AUTHORS	Karhunen, Ville; Gill, Dipender; Huang, Jian; Bouras, Emmanouil; Malik, Rainer; Ponsford, Mark; Ahola-Olli, Ari; Papadopoulou, Areti; Palaniswamy, Saranya; Sebert, Sylvain; Wielscher, Matthias; Auvinen, Juha; Veijola, Juha; Herzig, Karl-Heinz; Timonen, Markku; Keinänen-Kiukaanniemi, Sirkka; Dichgans, Martin; Salmi, Marko; Jalkanen, Sirpa; Lehtimäki, Terho; Salomaa, Veikko; Raitakari, Olli; Jones, Simon; Hovingh, G; Tsilidis, Konstantinos; Järvelin, Marjo-Riitta; Dehghan, Abbas

VERSION 1 - REVIEW

REVIEWER 1	Au Yeung, Shiu Lun Ryan; University of Hong Kong Faculty of Medicine. Competing Interest: I previous published with Jian Huang, a co-author of this paper.
REVIEW RETURNED	22-Feb-2022

GENERAL COMMENTS	This is a Mendelian randomisation (MR) study exploring the relation of cardio metabolic risk factors in cytokine level, and then cytokines in coronary artery disease risk. The authors also conducted colocalization and gene expression analyses. My main comments will be mainly focused on MR as I am most familiar with this method presented in the study. Major comments - How was the GWAS performed in this study similar/different from previous GWAS on cytokines (e.g. in terms of sample size, or range of cytokines), as referenced in the Introduction? I think the authors have briefly mentioned these in the Methods but would be good if the authors can discuss these in the Introduction. - Please clarify the meaning of "biologically plausible instruments" under the Methods - I wonder why MR-Egger was not considered? The intercept would also provide evidence for overall horizontal pleiotropy. - Given these cytokines could be correlated, I wonder if multivariable Mendelian randomisation would be helpful here when assessing the relation with CAD. However, these may be an issue with the level of dimensions presented here (i.e. many exposures, similar to issues we see for NMR Metabolomic profiling) - Have the authors considered the use of standard instrument selection approach for cytokine and compare and contrast the findings from the current analyses?

	Minor comments - I think it would be good to include the tests used for Mendelian randomisation in the main text.

REVIEWER 2	Chen, Lingyan; University of Cambridge, Public Health and Primary Care. Competing Interest: I became a full time employee of Novo Nordisk Ltd since June 2020.
REVIEW RETURNED	27-Feb-2022

GENERAL COMMENTS	Karhunen et al have investigated the interplay between circulating cytokines and cardiometabolic traits/diseases in a MR setting. They have identified cardiometabolic risk factors associated with inflammatory cytokines, cytokine signalling cascades and 3 putative targets for coronary artery disease. In general, this manuscript demonstrates a nice MR study. Major comments: 1. Cytokine instrument selection (1) While cis-pQTLs are less likely to be pleiotropy, using cis-pQTLs only as IVs for cytokines are likely to miss signals that might be good genetic proxies for some cytokines, such as IL6, as mentioned in Discussions. It would be nice to see MR estimates using IVs from both cis- and trans-pQTLs. (2) The authors derived pooled cis-eQTLs from GTEx using a fixed-effects meta-analysis across tissues. Is there any heterogeneity across tissues? What is the correlation between pooled eQTLs and blood eQTLs? eQTLGen might be a good resource for well-powered blood eQTLs. (3) Why use P-value < 1×10-4 as the threshold of cis-pQTLs / cis-eQTLs in IVs selection? 2. Selection of Cardiometabolic traits. It is not clear how the cardiometabolic phenotypes have been selected. The list of cardiometabolic traits as exposure is differed from that as outcomes? Minor comments: 1. Multiple testing. Does it make sense to use P<0.05/number of cytokines (N) as the significance threshold in the identification of significant associations/effect of circulating cytokines on other cytokines, when the multiple testing burden is N*47? Similar question applied to the MR of cytokine -> Cardiometabolic traits. 2. Some of the GWAS summary statistics for cardiometabolic phenotypes used in this study are relatively under-power, e.g. glycemic traits, why not use the latest publicly available GWASs (such as MAGIC) to maximise the power in MR analyses. 3. Reverse causation. In MR of cytokine -> Cardiometabolic traits, suggest doing a further sensitivity analysis, i.e. reverse causation. This can be achieved when using the same list of cardiometbolic traits in session 2 and session 4 (Figure 1). 4. In Sup Table 7, is the resource of cis-eQTLs Finnish and/or SCALLOP?

REVIEWER 3	Perera, Rafael; University of Oxford, Primary Care Health Sciences. Competing Interest: None
REVIEW RETURNED	03-May-2022

GENERAL COMMENTS	The present manuscript presents results arising from a series of Mendelian randomisation studies exploring the associations between inflammatory cytokines and CVD using data from three different Finnish cohorts. They use genetic data to create instrumental variables that will allow them to explore causal relationships. They go on to test for the impact of i) cardiometabolic traits on circulating cytokine levels, ii) circulating cytokines on other cytokine levels, and iii) circulating cytokine on cardiometabolic phenotypes. They report on relevant associations in each of these and conclude that this will help determine targets for therapeutics. I would recommend that a methodologist, familiar with MR, reviews this manuscript as there are multiple areas where I am unable to comment. As someone only partially familiar with MR studies, I found some of the reporting complex and not intuitive. The use of genetic data and MR as a design usually follows the identification of SNPs of the exposure that we believe is causal. Here it appears that the SNPs identified were for the cytokines which meant, I could not understand how this would support their first objective (impact of cardiometabolic traits on circulating cytokine levels). Secondly, it is unclear why this approach, identifying causal links, will allow the identification of targets for therapeutics. Even if there is a causal link, and a direct increase in levels, they have not linked these to relevant cardiometabolic outcomes. The risk factors might impact on cytokines and other relevant markers as well. Further clarification of this would be required.

REVIEWER 4	Bennett, Derrick. Competing Interest: None
REVIEW RETURNED	11-Jun-2022

GENERAL COMMENTS	1. Why doesn't the abstract contain any effect sizes or 95% CI? It is difficult to gauge the importance of the results without them. 2. The authors should provide more details on how the colocalization and posterior probabilities are arrived? 3. What is the rationale for a cut-off >0.7 for PPshare? A reference is needed to support this. 4. Did the authors perform power calculations to assess the effect sizes that could be detected reliably by their MR analyses? 5. Most associations were with genetically predicted BMI and cytokines. Did the authors consider any other adiposity traits such as WHR and WC (central adiposity measures)? These would also have been publicly available in GIANT. 6. The results section does not report any effect sizes and 95% CI for the effects of cardiometabolic traits on cytokines, cytokines on other cytokines or effects of cytokines on cardiometabolic traits. Some selected quantitative as well as qualitative statements would be helpful.

	7. The authors mention that they conducted sensitivity analyses (weighted-median and MR-PRESSO) for MR analyses of circulating cytokines with cardiometabolic traits (page 11: (lines 17-20). Did the authors also conduct sensitivity analyses for MR analyses of the effects of cardiometabolic traits on cytokines? 8. On page 12 (lines 25-26) the authors not that several of these inflammatory cytokines are directly targeted by biological drugs used in routine clinical practice or late-stage clinical trials. It would be useful to give some concrete examples of which cytokines and identify some ongoing trials. 9. On page 5 – supplementary material the authors state that "Prior to imputation, samples and probes with high missingness were excluded (MIND > 0.05 and GENO < 0.05)." Some guidance on MIND and GENO would be helpful for the reader of a general medical journal. 10. On page 6 of the supplementary material lines 17-23. What was the rationale for performing the RINT twice? A supporting reference for this approach should be supplied? 11. Were the genetic association estimates for SBP unadjusted or adjusted for BMI? 12. The authors used DIAGRAM for genetic variants related type 2 diabetes mellitus. Did they consider using data from the MAGIC consortium for continuous traits such as fasting glucose, HBA1c, fasting insulin as exposures? This may also have alleviated some power issues. 13. More details of how the clumping was performed in the TwoSampleMR package would be helpful for the cardiometabolic traits. 14. Why wasn't MR-Egger used as part of the sensitivity analyses? The authors should provide a rationale for this. 15. In supplementary figure 1 what does the colour coding (red, blue or black) of the cytokines represent? 16. In supplementary figure 2 are the numbers in brackets 95% CI? 17. Supplementary figure 4 the authors should perform a Demming regression line and add the slope and its associated 95% CI to this plot 18. In supplementary figure 6 it not very clear what the effect size metrics are (better labelling on the figure needed). Is it an OR per 1SD for CAD and a beta per 1SD for continuous traits? Need to make it clear that these are 95% CI, if not CI explain what they are. 19. Supplementary figure 7 the authors should perform a Demming regression line and add the slope and its associated 95% CI to this plot 20. Supplementary Table 2 does not mention GWAS data for fasting glucose or fasting insulin (which are mentioned on page 9 line 20 of the supplementary material).

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
This is a Mendelian randomisation (MR) study exploring the relation of cardio metabolic risk factors in cytokine level, and then cytokines in coronary artery disease risk. The authors also conducted colocalization and gene expression analyses. My main comments will be mainly focused on MR as I am most familiar with this method presented in the study.

We thank the Reviewer for these insightful and helpful comments.

Major comments
- How was the GWAS performed in this study similar/different from previous GWAS on cytokines (e.g. in terms of sample size, or range of cytokines), as referenced in the Introduction? I think the authors have briefly mentioned these in the Methods but would be good if the authors can discuss these in the Introduction.

We have expanded the list of 16 cytokines analysed in Sliz et al. [http://dx.doi.org/10.1136/jmedgenet-2018-105965] by including 31 cytokines from Ahola-Olli et al. [https://doi.org/10.1016/j.ajhg.2016.11.007] which were measured only in FINRISK or YFS studies (ten cytokines were measured in all three studies). Contrary to the previous work, we present the GWAS results unadjusted for body mass index to minimise the risk of collider bias affecting the results [https://doi.org/10.1016/j.ajhg.2015.12.019].

We have now added the following text in the Introduction:

“The list of cytokines consists of all cytokines measured in at least one of three Finnish cohorts (Methods), with the GWAS unadjusted for body mass index to minimise the risk of collider bias [https://doi.org/10.1016/j.ajhg.2015.12.019]”.

- Please clarify the meaning of "biologically plausible instruments" under the Methods

We now write:

“We then integrated publicly available GWAS summary statistics as well as eQTL data to generate ~~biologically plausible~~ instrumental variables for the cytokines, which have biological relevance to the cytokine under consideration through their presence at the relevant gene locus, and an association with the corresponding circulating protein, or additionally with the gene expression levels.”

- I wonder why MR-Egger was not considered? The intercept would also provide evidence for overall horizontal pleiotropy.

We thank the reviewer for this comment and now provide MR-Egger results. MR-Egger may be underpowered to detect evidence for pleiotropy when the confidence interval for the Egger intercept is large, as is typically the case when few instruments are available, therefore we restrict the MR-Egger to those exposures with ≥10 variants as instruments. We also note that care must be taken in

interpreting MR-Egger intercept due to potential violations of its assumptions (https://doi.org/10.1007/s10654-017-0255-x). We now write in the Supplementary Methods:

"MR-Egger regresses the variant-outcome associations on the variant-exposure associations, weighted by the precision of the variant-outcome association estimates. The regression slope represents the MR estimate, and presence of directional pleiotropy can be assessed by testing whether the intercept differs from zero [10.1093/ije/dyv080]. MR-Egger assumes that any pleiotropic effects of the genetic variants are uncorrelated with the variant-exposure associations, and the method may be biased due to outliers [10.1007/s10654-017-0255-x]. Therefore, we report MR-Egger results only for exposures with ≥10 variants available as instruments."

- Given these cytokines could be correlated, I wonder if multivariable Mendelian randomisation would be helpful here when assessing the relation with CAD. However, these may be an issue with the level of dimensions presented here (i.e. many exposures, similar to issues we see for NMR Metabolomic profiling)

We thank the reviewer for this important comment. However, as we are sure the Reviewer will also appreciate, multivariable Mendelian randomization would require more variants than exposures, which unfortunately we do not have for these exposures. Therefore, we have not conducted multivariable MR analyses here.

- Have the authors considered the use of standard instrument selection approach for cytokine and compare and contrast the findings from the current analyses?

Selecting variants from throughout the genome would almost certainly introduce horizontal pleiotropy, where the variants used as instrumental variables affect the outcome via pathways not related to the exposure. In particular, based on our results, cytokines studied here have multiple causal associations between each other. If the instruments are selected across the genome, one would therefore include variants that are not only specific to the cytokine at hand, but also variants that influence the upstream determinants of this cytokine. This results in a loss of specificity of the instruments proxying this cytokine. Hence, the results from such MR analyses are likely to be unreliable. We have therefore preferred to maintain the stringent criteria in our main MR analysis to maximise the likelihood of using valid instruments (https://doi.org/10.12688/wellcomeopenres.16544.2).

However, we do now additionally provide MR results using the conventional genome-wide instrument selection approach in secondary analyses.

We now write in the Methods:

"In secondary analyses, we additionally performed the same MR analysis by selecting instruments as independent (r^2<0.001) variants across the genome, associating with the cytokine levels at $P < 5\times10^{-8}$"

Supplementary Methods:

"As we only considered variants from within a biologically relevant genomic locus, we did not use the conventional threshold of $P < 5\times10^{-8}$ (which is usually applied when instruments are selected across the full genome) in our main analysis."

"As a supplementary analysis, we also report the results considering cytokines as exposures and cardiometabolic traits as outcomes, when selecting the instruments at $P < 5\times10^{-8}$ across the full genome from the GWAS summary statistics of the three Finnish cohorts, clumping at $r^2 < 0.001$."

Results:

"The MR results using genome-wide selection of the instruments for the cytokines (Supplementary Table 4) provided evidence for association (P<0.0033) for 16 cytokine-outcome pairs, most notably genetically proxied circulating levels of sE-selectin being associated with eight cardiometabolic traits (|β| from 0.01 to 0.07; Supplementary Table 11 and Supplementary Figure 9)."

Discussion:

"The MR results using genome-wide selection for instruments provided distinct results to the main results. In particular, genetically proxied circulating sE-selectin levels were associated with eight cardiometabolic traits, mostly driven by strong association in ABO locus [http://dx.doi.org/10.1136/jmedgenet-2018-105965]. Selecting instrument across the full genome allows for detection of trans-QTL proxies for cytokines. However, this strategy is also likely to include variants that are not specific only to the cytokine at hand, therefore potentially mis-specifying the exposure and introducing pleiotropic effects violating the MR assumptions."

Minor comments
- I think it would be good to include the tests used for Mendelian randomisation in the main text.

We now write in the main text Methods section:

"Inverse-variance weighted method was applied as the main MR analysis. We also conducted analyses using weighted median, MR-Egger and MR-PRESSO methods that are potentially more robust to violations of MR assumptions due to horizontal pleiotropy (Supplementary Methods)."

"The ratio method (if one instrument available) or inverse-variance weighted method (if two or more instruments available) was applied as the main MR analysis, complemented with weighted median, MR-Egger and MR-PRESSO methods to investigate whether the results were driven by pleiotropic effects (Supplementary Methods)."

Reviewer: 2

Comments to the Author
Karhunen et al have investigated the interplay between circulating cytokines and cardiometabolic traits/diseases in a MR setting. They have identified cardiometabolic risk factors associated with inflammatory cytokines, cytokine signalling cascades and 3 putative targets for coronary artery disease. In general, this manuscript demonstrates a nice MR study.

We thank the expert Reviewer for their diligent consideration and constructive feedback.

Major comments:

1. Cytokine instrument selection

(1) While cis-pQTLs are less likely to be pleiotropy, using cis-pQTLs only as IVs for cytokines are likely to miss signals that might be good genetic proxies for some cytokines, such as IL6, as mentioned in Discussions. It would be nice to see MR estimates using IVs from both cis- and trans-pQTLs.

As the Reviewer suggests, we now also provide results using the conventional genome-wide instrument selection criteria, using p<5e-8 as the threshold.

We highlight that in the main analysis, we use the criterion of selecting instruments from the vicinity coding gene of the corresponding cytokine. While we agree that this strategy is likely to miss some signals that might serve as proxies for some cytokines, the instruments selected are more likely to be specific to the considered cytokine (https://doi.org/10.12688/wellcomeopenres.16544.2), and not due to pervasive pleiotropic effects among cytokines (such as those highlighted in our cytokine-to-cytokine analyses).

We now write in the Methods:

"In secondary analyses, we additionally performed the same MR analysis by selecting instruments as independent (r^2<0.001) variants across the genome, associating with the cytokine levels at $P < 5\times10^{-8}$"

Supplementary Methods:

"As we only considered variants from within a biologically relevant genomic locus, we did not use the conventional threshold of $P < 5\times10^{-8}$ (which is usually applied when instruments are selected across the full genome) in our main analysis."

"As a supplementary analysis, we also report the results considering cytokines as exposures and cardiometabolic traits as outcomes, when selecting the instruments at $P < 5\times10^{-8}$ across the full genome from the GWAS summary statistics of the three Finnish cohorts, clumping at $r^2 < 0.001$."

Results:

"The MR results using genome-wide selection of the instruments for the cytokines (Supplementary Table 4) provided evidence for association (P<0.0033) for 16 cytokine-outcome pairs, most notably genetically proxied circulating levels of sE-selectin being associated with eight cardiometabolic traits (|β| from 0.01 to 0.07; Supplementary Table 11 and Supplementary Figure 9)."

Discussion:

"The MR results using genome-wide selection for instruments provided distinct results to the main results. In particular, genetically proxied circulating sE-selectin levels were associated with eight cardiometabolic traits, mostly driven by strong association in ABO locus [http://dx.doi.org/10.1136/jmedgenet-2018-105965]. Selecting instrument across the full genome allows for detection of trans-QTL proxies for cytokines. However, this strategy is also likely to include variants that are not specific only to the cytokine at hand, therefore potentially mis-specifying the exposure and introducing pleiotropic effects violating the MR assumptions."

(2) The authors derived pooled cis-eQTLs from GTEx using a fixed-effects meta-analysis across tissues. Is there any heterogeneity across tissues? What is the correlation between pooled eQTLs and blood eQTLs? eQTLGen might be a good resource for well-powered blood eQTLs.

We now provide the comparison of the cis-eQTL cross-tissue gene expression in GTEx and blood gene expression in eQTLGen. We write in the Methods: "The cross-tissue gene expression associations were further compared with blood expression statistics (Supplementary Methods)."

Supplementary Methods:

"To validate the relevance of the cross-tissue cis-eQTL instruments, we examined the correlation of the expression Z-scores of the cis-eQTL variants with the Z-scores in blood from eQTLGen consortium [https://doi.org/10.1038/s41588-021-00913-z]"

Results:

"For the cis-eQTL instruments, the Pearson correlation between cross-tissue expression Z-scores and blood expression Z-scores was 0.38 (95% CI 0.01 to 0.66) (Supplementary Table 7)."

(3) Why use P-value < 1×10-4 as the threshold of cis-pQTLs / cis-eQTLs in IVs selection?

As the variants used as instruments were selected based on the genomic location, the relevance assumption of the instruments is likely to be met (https://doi.org/10.12688/wellcomeopenres.15555.2). Considering instruments from only a subsection of a genome would therefore not require a stringent genome-wide multiple testing correction as is the case for the traditional threshold of P<5e-8. We further also evaluated the instrument strength using the F-statistic (Supplementary Methods, section "Cytokine Instrument Selection").

Our inclusion of instruments with weaker associations with the exposure may also contribute to increased risk of horizontal pleiotropy and violation of the exclusion restriction assumption, where the variants used as instruments affect the outcome via pathways not related to the exposure (https://doi.org/10.1038/s41576-018-0020-3). We have aimed to protect against this "by design", in that the instruments are only selected from a relevant genomic locus, where any horizontal pleiotropic effects are unlikely (https://doi.org/10.12688/wellcomeopenres.16544.2).

We now write in the Supplementary methods:

"As we only considered variants from within a biologically relevant genomic locus, we did not use the conventional threshold of $P < 5\times10^{-8}$ (which is usually applied when instruments are selected across the full genome) in our main analysis."

2. Selection of Cardiometabolic traits.

It is not clear how the cardiometabolic phenotypes have been selected. The list of cardiometabolic traits as exposure is differed from that as outcomes?

For exposures, we considered the classical cardiovascular risk factors (https://doi.org/10.1016/S0140-6736(19)32008-2), to explore whether these risk factors may be exerting their effects through inflammatory cytokines. In contrast, for outcomes, we considered the most prominent cardiometabolic disease outcomes, as well as physiological risk factors. However, following the Reviewer's

suggestion, we consider the same full set of traits as both exposures and outcomes, and have revised the Methods, Results and Discussion sections accordingly.

Minor comments:

1. Multiple testing. Does it make sense to use P<0.05/number of cytokines (N) as the significance threshold in the identification of significant associations/effect of circulating cytokines on other cytokines, when the multiple testing burden is N*47? Similar question applied to the MR of cytokine -> Cardiometabolic traits.

We appreciate the comments by the Reviewer. As the strongest MR results (P<0.05/number of outcomes) were taken further for colocalization analysis to be screened for potential false positive findings, we were deliberately lenient with the initial multiple testing correction and considered an aggressive approach to multiple testing inappropriate. Of note, using our threshold for multiple correction corresponds very closely to false-discovery-rate corrected threshold of P<0.05.

We now write in the Supplementary Methods:

"We accounted for multiple testing by applying a Bonferroni correction for the number of outcomes, resulting to P<0.05/47 = 0.0011 when cytokines were considered as outcomes, and P<0.05/15 = 0.0033 when considering the cardiometabolic traits as outcomes. The lenient approach to multiple testing correction was taken as the MR results were further validated in colocalization analysis, detailed below."

2. Some of the GWAS summary statistics for cardiometabolic phenotypes used in this study are relatively under-power, e.g. glycaemic traits, why not use the latest publicly available GWASs (such as MAGIC) to maximise the power in MR analyses.

We thank the Reviewer for highlighting this. The original analyses were performed before the publication of the most recent GWASs on lipids and glycemic traits. We have now updated the results using the most recent GWAS summary statistics for these traits.

3. Reverse causation. In MR of cytokine -> Cardiometabolic traits, suggest doing a further sensitivity analysis, i.e. reverse causation. This can be achieved when using the same list of cardiometabolic traits in session 2 and session 4 (Figure 1).

These additional analyses have now been performed, and the manuscript is revised accordingly throughout.

Results:

"We investigated causal effects of cardiometabolic traits on circulating cytokine levels using MR, and found positive associations (P < 0.0011 after a Bonferroni-correction for 47 outcomes) for genetically proxied body-mass index (BMI), waist circumference, systolic blood pressure, high and low-density lipoprotein cholesterol (HDL-C, LDL-C) levels, total cholesterol levels, triglycerides and smoking liability being associated with circulating levels of at least one cytokine (Figure 2, Supplementary Figure 2, Supplementary Tables 4-5). The absolute values of these effect sizes (|β|, per 1-standard-deviation increase in the genetically proxied exposure) varied from 0.08 to 0.48."

"Genetically proxied LDL-C levels were associated with circulating levels of IL6 (β = 0.12, 95% confidence interval [CI] 0.05 to 0.19)"

Discussion:

"Our analyses also revealed that cigarette smoking contributes to elevating CRP and MCP1 levels, higher SBP elevates CRP levels, and LDL-C elevates circulating IL6 levels. Taken together, these risk factors seem to increase cardiovascular disease pathogenesis at least partly through inflammatory mediators"

4. In Sup Table 7, is the resource of cis-eQTLs Finnish and/or SCALLOP?

In the Supplementary Table 7, the column "Source" indicates the source for each variant. The source for both cis-pQTL and cis-eQTL instruments were based on the availability of the GWAS summary statistics, the correspondence of the association estimates between the Finnish studies, INTERVAL and SCALLOP, and the sample size. This is detailed in the Supplementary Methods:
"We sought GWAS summary statistics from INTERVAL and SCALLOP consortia for the same cytokines that were analysed in our current work. We first tested whether the genetic associations were comparable between our work, INTERVAL, and SCALLOP consortia considering the different measurement methods. Specifically, for each cytokine, we examined the correlation between the beta estimates between our work and INTERVAL study and between our work and SCALLOP consortium based on the SNPs associated with the cytokine at P-value<10^{-5} and r^2<0.1. For 11 cytokines which beta estimates were correlated (regression coefficient P-value < 0.1 of regressing the beta estimates of the current study on the beta estimates of a previous study), we calibrated the beta estimates of the INTERVAL (3 cytokines) and SCALLOP (8 cytokines) consortium to match with the beta estimates from the current studies and performed the meta-analyses. For the rest of the cytokines, we used summary statistics from the GWAS with the largest sample size (the current studies (26 cytokines), the INTERVAL study (2 cytokines), and SCALLOP consortium (8 cytokines); Supplementary Table 3)."

Reviewer: 3

The present manuscript presents results arising from a series of Mendelian randomisation studies exploring the associations between inflammatory cytokines and CVD using data from three different Finnish cohorts. They use genetic data to create instrumental variables that will allow them to explore causal relationships. They go on to test for the impact of i) cardiometabolic traits on circulating cytokine levels, ii) circulating cytokines on other cytokine levels, and iii) circulating cytokine on cardiometabolic phenotypes. They report on relevant associations in each of these and conclude that this will help determine targets for therapeutics. I would recommend that a methodologist, familiar with MR, reviews this manuscript as there are multiple areas where I am unable to comment.

We thank the Reviewer for their consideration and feedback.

As someone only partially familiar with MR studies, I found some of the reporting complex and not intuitive. The use of genetic data and MR as a design usually follows the identification of SNPs of the exposure that we believe is causal. Here it appears that the SNPs identified were for the cytokines which meant, I could not understand how this would support their first objective (impact of cardiometabolic traits on circulating cytokine levels). Secondly, it is unclear why this approach, identifying causal links, will allow the identification of targets for therapeutics. Even if there is a causal link, and a direct increase in levels, they have not linked these to relevant cardiometabolic outcomes. The risk factors might impact on cytokines and other relevant markers as well. Further clarification of this would be required.

We thank the Reviewer for this consideration. As correctly stated, our paper tests for the impact of

i) cardiometabolic traits on circulating cytokine levels,

ii) circulating cytokines on other cytokine levels, and

iii) circulating cytokine on cardiometabolic phenotypes.

Point i) aims to explore how cardiometabolic traits affect levels of circulating cytokines. Point ii) aims to identify cytokine cascades and pathways. Point iii) that directly investigates putative therapeutic targets – namely circulating cytokines. We now detail this explicitly in the Introduction as follows:

"…we undertook MR to investigate the effect of cardiometabolic risk factors on levels of these circulating cytokines. Then, pooling these GWAS data with publicly available GWAS summary statistics in 21,735 individuals and further incorporating gene expression quantitative trait loci (eQTL) data from 15,201 samples across 49 tissues in 838 individuals, we identified biologically plausible genetic variants to proxy the effect of varying circulating cytokine levels. We performed MR and colocalization to investigate cytokine cascades and pathways. Finally, we used MR and colocalization to study the effects of circulating cytokines on cardiometabolic outcomes, to unravel potential therapeutic targets."

The instrument selection criteria for all these aspects are detailed in the Methods as follows:

"Instruments for each cardiometabolic trait were selected as single-nucleotide polymorphisms (SNPs) that associated with that trait at $P<5\times10^{-8}$ ($P < 1\times10^{-6}$ for fasting insulin) and were independent with correlation $r^2<0.001$."

"To define the instrumental variables for circulating cytokine levels… genetic instruments were generated using two different criteria. In the first approach, we selected variants within ±500 kb of their corresponding gene locus that associated with corresponding circulating cytokine levels at $P < 1\times10^{-4}$, which we term cis-protein QTL (cis-pQTL) instruments. In the second approach, we chose variants within ±500 kb of the corresponding gene locus that associated with both gene expression aggregated across tissues at $P < 1\times10^{-4}$, and circulating cytokine levels at $P < 0.05$, which we term cis-expression QTL (cis-eQTL) instruments."

Further details are given in the Supplementary Methods.

We also now clarify how "this approach… will allow the identification of targets for therapeutics" in the Introduction:

"Specifically, for molecular exposures that can be targeted by pharmacological interventions and which can be proxied by genetic variants, MR can be applied to study potential drug effects [https://doi.org/10.12688/wellcomeopenres.16544.2]."

Reviewer: 4

We thank the Reviewer for their time, consideration and constructive feedback.

Comments to the Author

1. Why doesn't the abstract contain any effect sizes or 95% CI? It is difficult to gauge the importance of the results without them.

BMJ Medicine is primarily directed towards clinical and medical audiences. As we detail in the paper (also detailed below), genetic effect estimates should not be directly translated to inform on the effect of clinical interventions, but rather should be used to identify potential causal effects. We therefore do not provide these estimates in the Abstract to avoid potential misrepresentation of our results, and instead focus on appropriate interpretation. If the Reviewer or Editors feel strongly, we could include P values to highlight the strength of the associations.

In the Discussion, we write:

"MR analyses should not be directly extrapolated to infer the effect of a clinical intervention, as the instruments employed represent the cumulative effect of lifelong genetic predisposition, while a clinical intervention typically represents a discrete event at a particular time point [https://doi.org/10.1093/ije/dyz236]."

2. The authors should provide more details on how the colocalization and posterior probabilities are arrived?

We appreciate the reviewer's comment and we have now substantially amended the description of colocalization and the posterior probabilities in the Supplementary Methods:

The colocalization analysis method 'coloc' proposed by Giambartolomei et al. (doi:10.1371/journal.pgen.1004383) as applied here investigates whether the data are consistent with a shared variant influencing both exposure and outcome trait. Under the assumption of a maximum of one causal variant per each trait within a genomic locus, the alternative hypotheses are:

H_0: No causal variants in the locus.
H_1: A causal variant on the exposure only.
H_2: A causal variant on the outcome only.
H_3: Distinct causal variants on exposure and outcome.
H_4: A shared causal variant on exposure and outcome.

Given data $D = \left(\hat{\beta}_j \;\; \mathrm{SE}(\hat{\beta}_j)\right), j = 1, \ldots, m$ for m variants from GWAS summary statistics, and prior probabilities p_1 for a variant being causal for the exposure, p_2 for a variant being causal for the outcome, and p_{12} for a variant being causal for both exposure and outcome, we can compute the posterior probabilities (PP) for all hypotheses. Specifically, the PP for a shared causal variant ($\mathrm{PP}_{\mathrm{shared}}$) is:

$$PP_{\mathrm{shared}} = P(H_4|D) = \frac{P(H_4|D)/P(H_0|D)}{1 + \sum_{k=1}^{4}\left(P(H_k|D)/P(H_0|D)\right)},$$

And the PP for distinct causal variants ($\mathrm{PP}_{\mathrm{distinct}}$) is:

$$PP_{\mathrm{distinct}} = P(H_3|D) = \frac{P(H_3|D)/P(H_0|D)}{1 + \sum_{k=1}^{4}\left(P(H_k|D)/P(H_0|D)\right)}.$$

In the above equations,

$$\frac{P(H_1|D)}{P(H_0|D)} = p_1 \times \sum_j ABF_j^{(1)} \qquad \frac{P(H_2|D)}{P(H_0|D)} = p_2 \times \sum_j ABF_j^{(2)}$$

$$\frac{P(H_3|D)}{P(H_0|D)} = p_1 p_2 \times \sum_{j \neq k} ABF_j^{(1)} ABF_k^{(2)} \quad \frac{P(H_4|D)}{P(H_0|D)} = p_{12} \times \sum_j ABF_j^{(1)} ABF_j^{(2)},$$

and *ABF* is the Approximate Bayes Factor (https://doi.org/10.1002/gepi.20359):

$$ABF = \sqrt{\frac{\mathrm{SE}(\hat{\beta})^2 + W}{\mathrm{SE}(\hat{\beta})^2}} \exp\{-\frac{1}{2}\left(\frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}\right)^2 \frac{W}{\mathrm{SE}(\hat{\beta}) + W}\}, \ \hat{\beta} \sim N(0, W),$$

where *W* is the prior variance for causal effect sizes, set here as $0.15 \times \mathrm{sd}(X)$ for continuous traits and 0.2 for binary traits (the default values used in coloc).

3. What is the rationale for a cut-off >0.7 for PPshare? A reference is needed to support this.

In light of the Reviewer's comment and a recent publication on combining Mendelian randomization (MR) and colocalization results (https://doi.org/10.1016/j.ajhg.2022.04.001), we have revised the colocalization criterion. We now write in the Supplementary Methods:

"Colocalization may suffer from reduced power if the association with the outcome is weak, which is more commonly the case when colocalization is done as a sensitivity analysis following MR. Additionally, when colocalization is conducted to examine the possibility of confounding by LD, already in the presence of MR evidence, it is of relevance to examine the probability of a shared causal variant conditioned on there being a causal variant for both traits. Therefore, we defined colocalization as being present when (i) PPshared + PPdistinct > 0.5 (to ensure sufficient power), and (ii) PPshared /(PPshared + PPdistinct) > 0.5 (to support that the probability of a shared causal variant is greater than the probability of distinct causal variants) [https://doi.org/10.1016/j.ajhg.2022.04.001]."

4. Did the authors perform power calculations to assess the effect sizes that could be detected reliably by their MR analyses?

We selected the list of cardiometabolic traits *a priori* based on their clinical relevance and the availability of GWAS summary statistics, and we did not perform power calculations prior to MR analysis. We prefer not to calculate the power *post-hoc*, as this is a direct transformation of the obtained p-value and provides no additional information of the detected associations. The corresponding precision of the analyses (and ability to detect effect sizes) can be inferred from the reported confidence intervals.

5. Most associations were with genetically predicted BMI and cytokines. Did the authors consider any other adiposity traits such as WHR and WC (central adiposity measures)? These would also have been publicly available in GIANT.

We have now amended our analyses to also consider waist-hip-ratio and waist circumference as additional cardiometabolic traits, for which the genetic associations were derived from the GIANT summary statistics. We now write in the Discussion:

“Our results suggest that elevated BMI increases multiple mediators of inflammation that affect various processes, including thrombosis (via plasminogen activation inhibitor-1), metabolism (HGF) and endothelial dysfunction (MCP1, TRAIL, sICAM1 and sE-selectin). These results for BMI were corroborated by similar associations for waist circumference and waist-hip-ratio.”

6. The results section does not report any effect sizes and 95% CI for the effects of cardiometabolic traits on cytokines, cytokines on other cytokines or effects of cytokines on cardiometabolic traits. Some selected quantitative as well as qualitative statements would be helpful.

We thank the reviewer for this suggestion and have now implemented selected quantitative and qualitative descriptions in the Results section. We now report the range of the absolute effect sizes throughout, and highlight some selected associations:

“We investigated causal effects of cardiometabolic traits on circulating cytokine levels using MR, and found positive associations ($P < 0.0011$ after a Bonferroni-correction for 47 outcomes) for genetically proxied body-mass index (BMI), waist circumference, systolic blood pressure, high and low-density lipoprotein cholesterol (HDL-C, LDL-C) levels, total cholesterol levels, triglycerides and smoking liability being associated with circulating levels of at least one cytokine (Figure 2, Supplementary Figure 2, Supplementary Tables 4-5). The absolute values of these effect sizes (|β|, per 1-standard-deviation increase in the genetically proxied exposure) varied from 0.08 to 0.48.”

“The odds ratio (OR) for CAD risk per 1-standard-deviation increase in genetically proxied MCSF = 1.13 (95% CI 1.06 to 1.20); for the quantitative traits, |β| of the top hits ranged from 0.02 to 0.38.”

“OR per 1-standard-deviation increase in genetically proxied sICAM on T2D risk: 0.79 [95% CI 0.67 to 0.92]; for quantitative trait associations, |β| ranged from 0.03 to 0.25).”

7. The authors mention that they conducted sensitivity analyses (weighted-median and MR-PRESSO) for MR analyses of circulating cytokines with cardiometabolic traits (page 11: (lines 17-20). Did the authors also conduct sensitivity analyses for MR analyses of the effects of cardiometabolic traits on cytokines?

We thank the reviewer for highlighting this and now specify the sensitivity analyses for MR analyses of the effects of cardiometabolic traits on cytokines in the Methods section:

“Inverse-variance weighted method was applied as the main MR analysis. We also conducted analyses using weighted median, MR-Egger and MR-PRESSO methods that are more robust to violations of MR assumptions due to horizontal pleiotropy (Supplementary Methods).”

8. On page 12 (lines 25-26) the authors note that several of these inflammatory cytokines are directly targeted by biological drugs used in routine clinical practice or late-stage clinical

trials. It would be useful to give some concrete examples of which cytokines and identify some ongoing trials.

We now write in the Discussion section:

“Indeed, VEGF signalling is already targeted clinically in the treatment of certain cancers and ophthalmic conditions [https://doi.org/10.1056/nejmoa1716948; https://doi.org/10.1056/nejmoa1414264].”

And

“Of relevance, anakinra, which is a recombinant and modified human IL1ra protein, is already used in the treatment of rheumatoid arthritis [https://doi.org/10.26355/eurrev_202112_27630].”

9. On page 5 – supplementary material the authors state that “Prior to imputation, samples and probes with high missingness were excluded (MIND > 0.05 and GENO < 0.05).” Some guidance on MIND and GENO would be helpful for the reader of a general medical journal.

We agree with the Reviewer and now instead write: “Prior to imputation, samples and probes with high missingness were excluded (missingness-per-individual > 0.05 and missingness-per-variant > 0.05).”

10. On page 6 of the supplementary material lines 17-23. What was the rationale for performing the RINT twice? A supporting reference for this approach should be supplied?

The inverse-normal rank transformation was performed twice for the results to be consistent with the previous studies by Ahola-Olli et al. [https://doi.org/10.1016/j.ajhg.2016.11.007] and Sliz et al. [http://dx.doi.org/10.1136/jmedgenet-2018-105965]. We write in the Supplementary Methods:

“The data pre-processing and transformations were done in a similar manner to previous GWAS analyses [https://doi.org/10.1016/j.ajhg.2016.11.007; http://dx.doi.org/10.1136/jmedgenet-2018-105965]”.

11. Were the genetic association estimates for SBP unadjusted or adjusted for BMI?

The genetic association estimates for SBP were unadjusted for BMI. We now explicitly state this in the Supplementary Methods, section “Genetic associations and instrument selection for cardiometabolic traits”:

“We used GWAS summary statistics that were unadjusted for other heritable phenotypes to avoid potential collider bias in MR (https://doi.org/10.1016/j.ajhg.2015.12.019)”

12. The authors used DIAGRAM for genetic variants related type 2 diabetes mellitus. Did they consider using data from the MAGIC consortium for continuous traits such as fasting glucose, HBA1c, fasting insulin as exposures? This may also have alleviated some power issues.

We have now amended the list of the considered cardiometabolic traits and updated the data sources for the most recent publicly available GWAS summary statistics. For glycemic traits, these were derived from the MAGIC Consortium, as the Reviewer highlights.

13. More details of how the clumping was performed in the TwoSampleMR package would be helpful for the cardiometabolic traits.

We now write in the Supplementary Methods:

"In clumping, a variant with the lowest p-value is detected within a specific genomic window, and any variants with r^2 above the pre-specified threshold with the lead variant are excluded. This procedure is iteratively repeated for any other remaining variants, also excluding the lead variants from the previous iterations. The final list of instruments consists of the lead variants for each iteration. We used the window size of 10,000 kilobase pairs, which is the default value in ld_clump() function."

14. Why wasn't MR-Egger used as part of the sensitivity analyses? The authors should provide a rationale for this.

We now provide MR-Egger results. MR-Egger may be underpowered to detect evidence for pleiotropy when the confidence interval for the Egger intercept is large, as is typically the case when few instruments are available, therefore we restrict the MR-Egger to those exposures with ≥10 variants as instruments. We also note that care must be taken in interpreting MR-Egger intercept due to potential violations of its assumptions (https://doi.org/10.1007/s10654-017-0255-x). We now write in the Supplementary Methods:

"MR-Egger regresses the variant-outcome associations on the variant-exposure associations, weighted by the precision of the variant-outcome association estimates. The regression slope represents the MR estimate, and presence of directional pleiotropy can be assessed by testing whether the intercept differs from zero [10.1093/ije/dyv080]. MR-Egger assumes that any pleiotropic effects of the genetic variants are uncorrelated with the variant-exposure associations, and the method may be biased due to outliers [10.1007/s10654-017-0255-x]. Therefore, we report MR-Egger results only for exposures with ≥10 variants available as instruments."

15. In supplementary figure 1 what does the colour coding (red, blue or black) of the cytokines represent?

The colour coding before was used for stylistic effect only. To prevent any ambiguity, we now use the same colour (blue) for all cytokines.

16. In supplementary figure 2 are the numbers in brackets 95% CI?

We now clarify this in the figure text: "The results …given as effect size estimate [95% confidence interval]"

17. Supplementary figure 4 the authors should perform a Demming regression line and add the slope and its associated 95% CI to this plot

We have now revised the figure to include the Deming regression line and its 95% confidence band.

18. In supplementary figure 6 it not very clear what the effect size metrics are (better labelling on the figure needed). Is it an OR per 1SD for CAD and a beta per 1SD for continuous traits? Need to make it clear that these are 95% CI, if not CI explain what they are.

We appreciate the reviewer's comment and have now explained the figure in more detail as follows (edits highlighted):

Supplementary Figure 6. Mendelian Randomization (MR) results (point estimates and their 95% confidence intervals) for the effect of genetically predicted circulating cytokine levels on cardiometabolic phenotypes considered as outcomes. The x-axis is in standard deviation scale (for quantitative outcomes) or on a log-odds-ratio scale (for binary outcomes). The effect size estimates are per 1-standard-deviation increase in the exposure. The results are plotted only for effects with strong evidence for causality (see Supplementary Methods). pQTL: protein quantitative trait loci; eQTL: expression quantitative trait loci. The abbreviations for cytokines are given in Supplementary Table 1.

19. Supplementary figure 7 the authors should perform a Demming regression line and add the slope and its associated 95% CI to this plot

This has now been implemented in the figure.

20. Supplementary Table 2 does not mention GWAS data for fasting glucose or fasting insulin (which are mentioned on page 9 line 20 of the supplementary material).

We thank the Reviewer for this highlighting this oversight and have now amended Supplementary Table 2 with all cardiometabolic traits that were used in the analyses.

VERSION 2 – REVIEW

REVIEWER 3	Perera, Rafael; University of Oxford, Primary Care Health Sciences. Competing Interest: None
REVIEW RETURNED	30-Sep-2022

GENERAL COMMENTS	Many thanks to the authors for their detailed response to my comments and also those made by the other reviewers and editors. I believe their updated version of their manuscript has addressed adequately all comments made and therefore is suitable for publication in BMJ Medicine.

REVIEWER 4	Bennett, Derrick. Competing Interest: None

REVIEW RETURNED	23-Sep-2022

GENERAL COMMENTS	The authors have addressed all of my comments and suggestions satisfactorily. I have no further comments to make.