Date: 21 Jan 2022

Journal: World Journal of Gastroenterology

Manuscript #: 73235

Title: Risk factors for major gastrointestinal bleeding in the general population in Finland.

Authors: Pareen Vora, Ronald Herrera, Arto Pietila, Ulrich Mansmann, Gunnar Brobert, Markku Peltonen and Veikko Salomaa

To the Editorial office,
World Journal of Gastroenterology

Thank you for considering our manuscript (no. 73235) and giving us the opportunity to submit a revision. Please find enclosed our response to each of the reviewers' and editors' comments with corresponding edits made in our revised manuscript.

We hope that we have satisfactorily addressed the comments and that you will find our revised manuscript suitable for publication in World Journal of Gastroenterology.

Kind regards,
Pareen Vora (corresponding author)

Responses to Editors comments

Science Editor

Comment 1: This is an interesting study which confirmed previously known risk-factors for GIB and identified potential risk factors not previously substantiated such as unemployment, BMI, GGT, SBP and coffee consumption as risk factors for major gastrointestinal bleeding in the general population in Finland.

Authors' response: We thank the editors for their interest in the findings of this study.

Comment 2: However, some data in the article are not consistent with common sense, such as the amount of coffee, please check it carefully.

Authors' response: We understand your comment on above average consumption of coffee however, this is specific to Finland where the coffee consumption is high. Finland ranks among the world's top coffee consuming nations per capita ~10-12kg per person per year with an average of 3-5 cups per day.[1-3] Therefore, it is not unexpected to find small proportion of individuals who drink 6-10 cups or >10 cups per day in Finland. In our study population, only 1.9% reported to consume >10 cups per day and 27.1% reported to consume 6-10 cups per day. Additionally, after each FINRISK survey was completed, the distributions of each variable and their maximum and minimum values were routinely checked, and obvious outliers were either checked or deleted. Previous studies [4-6] published using FINIRSK data also report coffee consumption as high as ≥10 cups per day.

Comment 3: In addition, the supplementary chart does not have corresponding instructions.

Authors' response: We have now added further explanation in the supplementary figures' legends. –

"**Supplementary figure 1. Study design overview.** Baseline data collected at enrollment includes demographics, socioeconomic factors, lifestyle factors, blood pressure measurements, and laboratory measurements. Occurrence of gastrointestinal bleedings (GIB) was observed within 10 years after enrollment.

Follow-up was censored at GIB event, death or end of follow-up and medical history was observed any time before censoring."

"**Supplementary figure 2. Cross validation plot for the penalty term generated from the least absolute shrinkage and selection operator Cox regression**. The top x-axis shows the number of variables in the model from 1-46. The bottom x-axis shows log (λ) values for the dashed lines corresponding to the minimum λ value (λ_{min} - left dashed line) and λ within one standard deviation (λ_{1se} - right dashed line). λ_{min} is the value for which the model with the respective number of variables has the lowest partial likelihood deviance i.e., smaller the deviance – better the fit. This deviance has a certain variability as shown using grey error bars to every red point. λ_{min} and λ_{1se} gives us a range of variables (12-39) which balances between model accuracy and model parsimony."

Company editor-in-chief

Comment 1: I have reviewed the Peer-Review Report, the full text of the manuscript, and the relevant ethics documents, all of which have met the basic publishing requirements of the World Journal of Gastroenterology, and the manuscript is conditionally accepted.

Authors' response: We thank the editors for their interest in the findings of our study and for a positive evaluation.

Comment 2: I have sent the manuscript to the author(s) for its revision according to the Peer-Review Report, Editorial Office's comments and the Criteria for Manuscript Revision by Authors.

Authors' response: We have revised the manuscript based on the comments from the editors and peer-reviewers and followed the instructions to upload the requested files to process the submission.

Comment 3: Please provide decomposable Figures (in which all components are movable and editable), organize them into a single PowerPoint file.

Authors' response: We have now organized and uploaded the supplementary figures in a single PowerPoint file as decomposable figures.

Comment 4: Please authors are required to provide standard three-line tables, that is, only the top line, bottom line, and column line are displayed, while other table lines are hidden. The contents of each cell in the table should conform to the editing specifications, and the lines of each row or column of the table should be aligned. Do not use carriage returns or spaces to replace lines or vertical lines and do not segment cell content.

Authors' response: We have now formatted the tables to conform the editing specifications.

Responses to Reviewer's comments

Reviewer #1

This is an interesting and well-written paper.

Authors' response: We thank the reviewer for this positive overall assessment.

Several minor suggestions for revision are listed below:

Comment 1: The paper requires a minor polishing of the English language. Use at least Grammarly.

Authors' response: Thank you for the feedback. We got the manuscript reviewed by a native English speaking medical writer from EpiMed Communications Ltd. 121 South Avenue, Abingdon, Oxon, OX14 1QS, UK – to improve the English. We hope that it has improved now and fulfilled your expectations.

Comment 2: Please add the approval number of the ethics committee (I know this is a secondary database).

Authors' response: We have now added the ethics committee approval number under 'Ethics statement' (page 5) – "THL/66/0.05.00/2015".

The secondary use of the survey data was approved by Finnish Institute of Health and Welfare as our study plan was consistent with the existing ethical permissions and the FINRISK permissions are only granted if the study in question does not violate the existing ethical approvals.

Comment 3. The consumption of coffee seems unrealistic to me: how can someone consume 6-10 or >10 cups of coffee per day? Which volume of coffee was taken into account?

Authors' response: We understand your comment on above average consumption of coffee, however, this is specific to Finland where the coffee consumption is high. Finland ranks among the world's top coffee consuming nations per capita ~10-12kg per person per year with an average of 3-5 cups per day.[1-3] Therefore, it is not unexpected to find small proportion of individuals who drink 6-10 cups or >10 cups

per day in Finland. In our study population, only 1.9% reported to consume >10 cups per day and 27.1% reported to consume 6-10 cups per day.
The question in the FINRISK survey regarding the 'no. of cups of coffee per day' specified the volume of coffee cups as '1 cup = c. 1 deciliter'. After each FINRISK survey, the distributions of each variable and their maximum and minimum values were routinely checked, and obvious outliers were either checked or deleted. Additionally, previous studies [4-6] published using FINIRSK data also report coffee consumption as high as ≥10 cups per day.

Comment 4: The anticancerous effects of coffee consumption can be further detailed in the Discussions. It contains a large amount of flavonoids which have antineoplastic effects. Look for meta-analyses of this subject.
Authors' response: We thank the reviewer for suggesting additional explanation regarding antineoplastic effects of coffee. We found several meta-analyses that show that coffee consumption is not associated with Gastric cancer [7-12] however there are several meta-analyses that indicate a protective effect on colorectal cancer[13-18] and have added the following statement in the Discussion section page 11 -
"Nonetheless, several meta-analyses have reported a protective effect of coffee consumption on colon cancer which is a major cause of GIB."

Reviewer #2:
Comment 1: Is the sample representative of the general population?
Authors' response: We thank the reviewer to highlight this point which is one of the strengths of the study. Yes, the FINRISK study was a cross-sectional survey, which invited a random and representative sample of population from several geographic regions of Finland and enrolled participants who responded to the invitation. The FINRISK surveys were initiated in 1972 and carried out every 5 years with a cohort size of 6000–8800 per survey. The survey stratified the participants to contain at least 250 subjects of each sex and 10-year age group from each geographical area. The participation rate in 1972 survey was approximately 80% with a gradually declining trend to 57% in men and 67% in women in 2012. [19] However these participation rates are still higher compared to other countries. Therefore, we have added the

following statement in the 'Study design, data source, and study population' section page 6 – "Survey participants were randomly chosen using the population register of Finland to obtain a representative sample of individuals across several geographic regions of Finland; those who responded to the invitation were subsequently enrolled in the study as participants in the first quarter of each survey year. From each geographical area, the surveys enrolled at least 250 subjects of each sex and 10-year age group. From the 1972 survey to the 2007 survey, the participation rate gradually decreased from approximately 80% to 65%"

Comment 2: Is the data expected to be robust / homogenous with regard to missing numbers under each variable?

Authors' response: We thank the reviewer for this question. There was missing data for some variables however it was minimal (ranging from 0.3 – 2.8%) with respect to the large sample size of the study. Therefore, we do not expect missing values to considerably affect results of this study. Further, participants with missing values were excluded from the final modelling exercise and no imputations were performed. We have already described this in 'Statistical analyses' section page 8-9 - "There were few participants with missing baseline data (<3%); they were excluded from the analysis and no imputations were performed."

Comment 3: Was a subgroup analysis between the single vs multiple bleeders, dead vs live, no of transfusions considered?

Authors' response: We thank the reviewer for raising this question. The focus of our analyses was to evaluate the risk factors for the first GIB, and hence we stopped the follow-up at first "incidence of GIB that led to hospitalization/death, death from any cause, or maximum of ten years, whichever occurred first" as described in 'Study design, data source, and study population' section page 6. The case definition of GI bleed included "GIB that led to hospitalization or GIB-specific death" as described in 'Study outcome' section page 7. Unfortunately, we do not have large enough number of multiple GI bleeding cases or deaths due to GI bleeding to do a stratified sub-group analysis. However, in our previous publication using the same data source, we

reported the numbers and the incidence rate, recurrent rates, mortality and case-fatality for GIB stratified by age, gender, and type of GIB.[20]
Unfortunately, we do not have information on the no. of transfusions as this was not collected as part of the FINRISK study and the interventions used in the inpatient setting are not collected in the hospital discharge register used for this study.

Comment 4: Was a co-relation with the final outcome (better/rebleed/death) considered?
Authors' response: As the focus of the analyses was the first GIB, we did not further investigate outcomes beyond it. However, in our previous publication using the same data source, we reported the proportion of particpiants experiencing recurrent GIB and death due to GIB.[20]

Comment 5: Was the effect of time over change in lifestyle / alteration in variables taken into consideration?
Authors' response: Unfortunately, repeated measurements of these variables were not collected over time hence we could not account for changes in lifestyle variables. We acknowledge this as a limitation and have added a statement in the 'Strengths and Limitation' section page 13 –
"Lastly, the data on demographic, lifestyle, and laboratory parameters were only collected at baseline and no repeated measurements were conducted during the study period. Therefore, we could not account for lifestyle modifications on the risk of GI bleeding in our analyses"

Comment 6: In case of Education a different way of categorisation (ug /pg /etc) could have been better
Authors' response: We thought about different classifications for education, however, the education standards and level have changed/improved over the years of the study period. Therefore, to be comparable/equal across different time periods, we standardized education across 20-year period. This has also been used similarly in previously published studies using the same data source.[21]

Comment 7: The reliability of self-quantification smoking / alcohol is of concern

Authors' response: We agree with the reviewer that smoking and alcohol consumption can be underestimated in self-reported questionnaires. However, self-reported smoking has been validated by using data collected in the 1992 FINRISK survey and reported to be high (≥95%).[22] Additionally, publications show that alcohol consumption levels in FINRISK rightly predicts liver diseases as expected.[23] Further, these questionnaires were mailed to the participants who agreed to take part in the study and answered these questions privately, ensuring that the participants are not hesitant to respond to these questionnaires. Unfortunately, in a large population-based surveys it is not practical to get biochemical evidence for smoking and alcohol from all the participants. Hence, we acknowledge this as a limitation and have added a statement in the 'Strengths and Limitations' section page 13 –
"Information on lifestyle factors such as smoking, and alcohol consumption are self-reported and might be under-reported"

Comment 8: Was the bearing of modifications in lifestyle on prevention / altering risk taken into consideration

Authors' response: Similar to the comment 5 as repeated measurements of these variables were not collected over time hence, we could not account for modifications in lifestyle on the risk of GI bleeding. We acknowledge this as a limitation with the statement in the 'Strengths and Limitation' section page 13 –
"Lastly, the data on demographic, lifestyle, and laboratory parameters were only collected at baseline and no repeated measurements were conducted during the study period. Therefore, we could not account for lifestyle modifications on the risk of GI bleeding in our analyses"

Comment 9: Several statistical data sets / graphs are included without elaborate discussion of the contents (supplemental fig 2, table 3)

Authors' response: We thank the reviewer for suggesting to add more explanation to supplementary fig 2, and table 3. Therefore, we have now added additional explanation in the results section page 9-10 where the supplementary table 3 is referenced –

"The LASSO method identified the most important predictors from larger set of variables. Variables with negative coefficients exhibit decreased risk, positive coefficients exhibit increased risk, and coefficient with value zero are the least important predictor variables in the model to predict gastrointestinal bleeding and can be removed from the final model. The aim of LASSO method is model prediction by selecting the most important predictor variables and therefore statistical significance of regression coefficients is not computed here (see supplementary table 3). Using these results from LASSO, we excluded variables such as DBP, HDL, VTE, and Inflammatory connective tissue diseases from the final cox model. Categorical variables for which one of the strata had a zero coefficient were kept in the final model."

We have also added brief explanation in the Supplementary table 3 legend in supplementary material page 4 –
"Negative coefficients show decreased risk, positive coefficients show increased risk, and zero coefficients are the least important variables in the model to predict gastrointestinal bleeding"

We have also added additional explanation as figure legend for Supplementary figure 2 in supplementary material page 8 –
"The top x-axis shows the number of variables in the model from 1-46. The bottom x-axis shows log (λ) values for the dashed lines corresponding to the minimum λ value (λ_{min} - left dashed line) and λ within one standard deviation (λ_{1se} - right dashed line). λ_{min} is the value for which the model with the respective number of variables has the lowest partial likelihood deviance i.e., smaller the deviance – better the fit. This deviance has a certain variability as shown using grey error bars to every red point. λ_{min} and λ_{1se} gives us a range of variables (12-39) which balances between model accuracy and model parsimony."

References

1. Clausnitzer J. Consumption of roasted coffee per capita in Finland 2010 to 2020, https://www.statista.com/statistics/708603/coffee-consumption-per-capita-in-finland/ (2021, accessed 21 Dec 2021).
2. Armstrong M. The Countries Most Addicted to Coffee, https://www.statista.com/chart/8602/top-coffee-drinking-nations/ (2020, accessed 21 Dec 2021).
3. CBI. Exporting coffee to Finland, https://www.cbi.eu/market-information/coffee/finland (2019, accessed 17 Jan 2022).
4. Tuomilehto J, Hu G, Bidel S, et al. Coffee Consumption and Risk of Type 2 Diabetes Mellitus Among Middle-aged Finnish Men and Women. *JAMA* 2004; 291: 1213-1219. DOI: 10.1001/jama.291.10.1213.
5. Wang Y, Tuomilehto J, Jousilahti P, et al. Coffee consumption and the risk of heart failure in Finnish men and women. *Heart* 2011; 97: 44-48. 2010/10/29. DOI: 10.1136/hrt.2010.206045.
6. Tanskanen A, Tuomilehto J, Viinamäki H, et al. Heavy coffee drinking and the risk of suicide. *Eur J Epidemiol* 2000; 16: 789-791. 2001/04/12. DOI: 10.1023/a:1007614714579.
7. Li L, Gan Y, Wu C, et al. Coffee consumption and the risk of gastric cancer: a meta-analysis of prospective cohort studies. *BMC Cancer* 2015; 15: 733. DOI: 10.1186/s12885-015-1758-z.
8. Liu H, Hua Y, Zheng X, et al. Effect of Coffee Consumption on the Risk of Gastric Cancer: A Systematic Review and Meta-Analysis of Prospective Cohort Studies. *PLOS ONE* 2015; 10: e0128501. DOI: 10.1371/journal.pone.0128501.
9. Deng W, Yang H, Wang J, et al. Coffee consumption and the risk of incident gastric cancer—A meta-analysis of prospective cohort studies. *Nutrition and Cancer* 2016; 68: 40-47. DOI: 10.1080/01635581.2016.1115093.
10. Botelho F, Lunet N and Barros H. Coffee and gastric cancer: systematic review and meta-analysis. *Cad Saude Publica* 2006; 22: 889-900. 2006/05/09. DOI: 10.1590/s0102-311x2006000500002.

11. Shen Z, liu H and Cao H. Coffee consumption and risk of gastric cancer: An updated meta-analysis. *Clinics and Research in Hepatology and Gastroenterology* 2015; 39: 245-253. DOI: https://doi.org/10.1016/j.clinre.2014.09.005.
12. Xie Y, Huang S, He T, et al. Coffee consumption and risk of gastric cancer: an updated meta-analysis. *Asia Pac J Clin Nutr* 2016; 25: 578-588. 2016/07/22. DOI: 10.6133/apjcn.092015.07.
13. Wang A, Wang S, Zhu C, et al. Coffee and cancer risk: A meta-analysis of prospective observational studies. *Scientific Reports* 2016; 6: 33711. DOI: 10.1038/srep33711.
14. Sartini M, Bragazzi NL, Spagnolo AM, et al. Coffee Consumption and Risk of Colorectal Cancer: A Systematic Review and Meta-Analysis of Prospective Studies. *Nutrients* 2019; 11 2019/03/27. DOI: 10.3390/nu11030694.
15. Wang Y, Chen J, Zhao R, et al. Dose–response meta-analysis of coffee consumption and risk of colorectal adenoma. *European Journal of Clinical Nutrition* 2020; 74: 297-306. DOI: 10.1038/s41430-019-0467-0.
16. Li G, Ma D, Zhang Y, et al. Coffee consumption and risk of colorectal cancer: a meta-analysis of observational studies. *Public Health Nutrition* 2013; 16: 346-357. 2012/06/14. DOI: 10.1017/S1368980012002601.
17. Gan Y, Wu J, Zhang S, et al. Association of coffee consumption with risk of colorectal cancer: A meta-analysis of prospective cohort studies. *Oncotarget* 2014; 8. DOI: 10.18632/oncotarget.8627.
18. Micek A, Gniadek A, Kawalec P, et al. Coffee consumption and colorectal cancer risk: a dose-response meta-analysis on prospective cohort studies. *International Journal of Food Sciences and Nutrition* 2019; 70: 986-1006. DOI: 10.1080/09637486.2019.1591352.
19. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. *International Journal of Epidemiology* 2017; 47: 696-696i. DOI: 10.1093/ije/dyx239.
20. Vora P, Pietila A, Peltonen M, et al. Thirty-Year Incidence and Mortality Trends in Upper and Lower Gastrointestinal Bleeding in Finland. *JAMA Network Open* 2020; 3: e2020172-e2020172. DOI: 10.1001/jamanetworkopen.2020.20172.

21. Lindbohm JV, Kaprio J, Jousilahti P, et al. Risk Factors of Sudden Death From Subarachnoid Hemorrhage. *Stroke* 2017; 48: 2399-2404. 2017/07/26. DOI: 10.1161/strokeaha.117.018118.
22. Vartiainen E, Seppälä T, Lillsunde P, et al. Validation of self reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Community Health* 2002; 56: 167-170. 2002/02/21. DOI: 10.1136/jech.56.3.167.
23. Sahlman P, Nissinen M, Puukka P, et al. Genetic and lifestyle risk factors for advanced liver disease among men and women. *J Gastroenterol Hepatol* 2020; 35: 291-298. 2019/07/02. DOI: 10.1111/jgh.14770.