Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-21-747

Reviewer A

With colon cancer screening becoming increasingly important and needed (with many facing delayed screening in COVID era and lowered age of screening recommendations) This group of authors have composed an interesting and important manuscript regarding a potential less invasive test that may be used for this purpose. While there are some areas within the manuscript that will require some revision, this is an important manuscript that should be publishable with attention paid to the noted areas. The strengths of the manuscript lie in the relevance of the topic to an important issue in medicine, the prospective enrollment of patients, preselected population size and potential use of a non-invasive and cost effect measure compared to the current standard. Weaknesses include lack of discussion surrounding blinding and missing tables and raw data. Please see my specific comments below:

General:

1. There is note of table 1 in the manuscript although I am unable to see this. Is this my mistake or is this missing? the information that it likely contains would be very helpful in understanding the manuscript better.

Reply: We re-attach Table 1 with its legend in the Tables section.

2. The number of patients with identified cancers seems somewhat high (45/221, 20.4%) in relation to expected number to find on colonoscopy in a population of standard risk individuals. Any potential explanation for this that you found? Could help to add something about this in the discussion, given potential that this population may not fit as much as hope with others.

Reply: The target population was highly selected because it came from a rapid colon cancer circuit made up of patients with a high suspicion of having colon cancer, and where preferential colonoscopies are performed.

3. Plots showing the raw data that is used for each laboratory value studied and used in the models would be helpful to understand the true meaning of the results.

Reply: We included them in the article. They are figures 1 and 2 in the figures section.

4. I potentially missed this, but I do not see mention of whether there was blinding of colonoscopy outcomes and lab values to the researchers during the study.

Reply: Yes, there was. There was blinding, since the results of the colonoscopy were only known to the data manager of the study and from the laboratory they were processed without knowing the results of the colonoscopy.

5. There was no further mention of survey items outside of the methods. Even if brief, addition of this information also would be helpful in clarifying the population risk and such.

Reply: It is described in Table 1. When statistical analysis was performed between these elements, they were not significant between the different groups.

6. Family history of colon cancer and higher meat consumption being more common in the non-cancer group in unexpected. Some comment on that in discussion may be helpful.

Reply: In this case, the family history is greater, since the target population is a population at risk that undergoes a colonoscopy. Regarding meat consumption, the results of the sample are paradoxical. The explanation may be that the study is not designed to study risk factors and the sample is very small compared to the large epidemiological studies that have analyzed these factors.

7. Paragraph 2 in the results section is overall confusing to me and could use some rewording and some more details:

a. Does the first sentence mean that each individual lab discriminated between each group?

Reply: It is modified in the article so that it is better understood (beginning of the second paragraph of results)

b. What is meant by "high reactivity" of patients with IBD? Is this referring to elevated levels of inflammation?

Reply: In this case, it referred to active inflammatory disease or in the acute phase.

Are these few sentences trying to say that the elevated levels of the noted markers (ferritin, CA125) in IBD makes them less helpful in discriminating between CRC and no CRC in patients with IBD?

Reply: Ferritin is a pleiotropic protein related not only to iron metabolism, but also plays an important role as a regulator of immunity, and as a mediator of inflammation and microcirculatory dysfunction. For this reason, serum ferritin levels increase in various inflammatory processes, which means that it is considered an acute phase reactant, although it is not a marker of the severity of the inflammatory component of the disease in all cases. There are several non-tumor clinical entities that can present with elevated ferritin levels, depending on the degree of inflammation of the underlying pathology. In our context, serum ferritin increases, as it is an acute phase reactant, in chronic inflammatory processes or those associated with acute outbreaks, as usually occurs in the group of non-tumor pathologies, which may present with a certain inflammatory component, such as ulcerative colitis and Crohn's disease in the active phase. Unlike CRC, of a lesser inflammatory nature, and generally associated with incipient or clinically manifest iron deficiency situations, which would generally manifest with decreased levels of this biochemical marker, whose dual antagonistic behavior would give it a notable discriminant value in the logistic regression model proposed.

8. Can it be explained at all how the cut off points were determined in the algorithms?

Reply: Initially, the Youden index supplied by the statistical package was used on the assumption that the two errors derived from the practical application of the model, false positives (FP) and false negatives (FN), were equally relevant. This objective indicator establishes an optimal initial equilibrium solution between both errors and their corresponding statisticians (sensitivity and specificity), so that the cut-off point is chosen as the one that minimizes the sum of both errors (PF+FN). With this initial cut-off point, this mathematical algorithm based on the multivariate logistic regression model allows, with the results of only a simple blood test, to calculate the diagnostic probability of CRC in asymptomatic patients but with some risk factor, and in patients with some clinical suspicion of CRC that justifies performing a colonoscopy. In other words, this initial cutoff predicts with high sensitivity and specificity the risk of malignant tumor before performing the colonoscopy or any other invasive technique. Our immediate project is the external and multicenter validation of the study, an objective validation of the formulas proposed in this training study, and that will allow us to

ascertain their potential impact in the clinical context of early diagnosis of CRC.
This initially proposed cut-off point can be modified depending on the clinical utility of the proposed algorithm. This initial study raises new questions that must be answered with new research projects. Would the introduction of the algorithm in the care process mean a more profitable (cost-efficient) diagnosis than the current one? With regard to the target population, should screening with the proposed algorithm be performed in the asymptomatic population or only in patients with some clinical suspicion, or in both? Could this discriminant function be used as a replacement for the current fecal occult blood (FOB) test? Should it be used in conjunction with FOBT in current colon cancer population screening programs to reduce the number of current colonoscopies? From a logistic point of view, should the biochemical tests of the algorithm be performed before, during or after the FOBT? Only in the positive SOH or in all? Would its systematic application produce a significant reduction in the number of protocolized colonoscopies based on the degree of risk derived from the evaluated logistic regression model? That is, patients with very low risk according to the biochemical algorithm could be excluded from performing the protocolized colonoscopy in this clinical context of the selected population. And last but not least, from the point of view of hospital management. , the possibility of using the probability of CRC generated by the formula as a priority, additional or complementary criterion for managing the waiting list for conventional colonoscopies would be interesting. The authors of the study were satisfied that the proposed algorithm could be used to prioritize patients with a higher risk of CRC in the care process according to the tested formula versus patients with a low probability of CRC. In all these situations, the cut-off point should be modified depending on the questions we intend to answer in future research studies, so initially, we propose this value based on the Youden index.

9. Would add a little more explanation as to the choice of the markers used in the study. Expanding on what is stated in the sentence on line 151-153 with some more specifics on how these performed in those with high tumor burden would give some more support to the use in this study.

Reply: Experimental markers have little scientific evidence and when the increase has been detected in different tumors, it has generally been correlated with a high tumor burden. In the field of early diagnosis, there is very little experience, which is why we found it attractive to include them in our study.

10. If possible, would insert the IRB approval number and date in the sentence on lines 167-169.

Reply: We added it in line 169 of the Methods section.

Reviewer B

The present manuscript describes a biomarker study aimed at blood-based early detection of colorectal cancer (CRC), which is an important scientific and clinical subject, as colorectal cancer is a frequent cancer type in both men and women all over the world, and screening has proven to reduce morbidity and mortality, as well as incidence of CRC. A blood-based alternative to the current feces tests could increase screening program effectiveness.

The study aimed to

i) investigate predictive biomarkers for CRC screening, either individually or combined in an algorithm.
ii) whether they could discriminate between precancerous lesions and benign lesions as well as inflammatory bowel disease.
iii) the association between the biomarker and UICC stage, clinical, endoscopic, and histological variables.

The study was described as a descriptive, prospective, cross-sectional study. Though the data was collected prospectively, the biomarkers were analysed after colonoscopy results were available, and thus in a retrospective manner.
Reply: The study was blinded for the investigators, when the samples were processed the results of the colonoscopy were unknown.
Colonoscopies were repeatedly referred to as 'bloody'. It would be more scientific if authors abstained from the use of this incorrect expression, as colonoscopies are rarely bloody; I have yet to cause a patient intestinal bleeding by performing a colonoscopy. 'Invasive' may be a more accurate and scientific term.
Reply: The word "bloddy" is changed to "invasive" (line 118)

Materials and methods:

In the materials and methods section, it would be nice if authors included a figure (flow-chart) describing how many patients were eligible for inclusion; how many of the eligible patients were not included; due to which reasons were they excluded; how many were lost of the included patients and why etc. This would help peers to evaluate if there could be any bias in the selection of the study cohort.
Reply: Eligible patients were 227. Losses were 6. Of these, 4 for not accepting the analytical extraction and 2 cases because the extracted sample was insufficient to perform a complete analysis.
Authors refer to CRC stage in the text, is this UICC stage? And if it is, what is stage 0 (line 230)?
Reply: Yes, it is the UICC TNM Project the 8th Edition. Stage 0 includes carcinoma in situ.
The authors write in the discussion section that there were some data losses in the control group, but given the characteristics of the multivariate analysis, they did not affect the final outcome of the study. This data loss needs to be documented in the result section, for peers to evaluate if authors assessment is correct.
Reply: The losses were not analyzed, since they did not meet the requirements for processing, either because they refused the extraction or because there was not enough sample. In this case, no analysis was processed, nor included in the study. In any case, they were patients without CRC. We corrected it in the text, discussion section on lines 314 to 316.

The included patients were not screening individuals, but individuals referred to colonoscopy due to clinical symptoms; thus this patient group is not comparable to CRC screening individuals. The latter are healthy individuals, whereas the former have gastro-intestinal symptoms, and the two populations are therefore not necessarily comparable. This is a point

that should be addressed in the discussion, as the authors aim to apply the proposed test to individuals in CRC screening programs.
Reply: We clarified it in the discussion section, lines 316 to 321.

The methods section describes that a blood sample was collected, that the "clinicopathological characteristics of the biopsy were collected" and that serum levels of tumor markers were determined (line 189-193). This description is rather shallow, and the lack of detail leaves peers unable to determine whether biomarker analyses of the tumor markers live up to good scientific practice.
Reply: All international standards of good practice and the recommendations of the reagent manufacturers were followed. All reagents have FDA approval for their marketing and use, as well as the mandatory UNE standards for their marketing and use in Europe.
How much blood was sampled?
Reply: Two 8 mL gelose biochemistry tubes were collected.
Were all biomarkers proteins?
Reply: Yes, they were.
Were all biomarkers analysed in one sample, or in different samples (raising costs), with which methods etc. The lack of laboratory description in the materials and methods section should be addressed.
Reply: The biomarkers were analyzed in the same sample.
The authors write that analyses was done according to manufactures instructions, but if some of the markers are not commercially available (as they are only emerging), what are those instructions? Further, there is no reference for peers to look up whether these laboratory methods are gold standard.
Reply: Classical tumor markers were by immunochemiluminescence from Roche Diagnostics SL, ferritin by immunoturbimetry from Roche Diagnostics SL, calprotectin by Enzymoimmunoassay from Palex Medical, SL. Hemogram was performed by cytometry in Syxmex 9760, general biochemistry was performed in Cobas 8000 from Roche Diagnostic SL. All non-classical markers were determined by ELISA using commercial kits purchased from Cayman Chemical, 1180 East Ellsworth Road, Ann Arbor, Michigan 48108 USA (8-OHdG), Human Lipocalin-2/NGAL Quantikine ELISA Kit were from R&D Systems Biotechne, 614 McKinley Place NE, Minneapolis, MN 55413, USA. EGFR and Cyr-61 were from abcam, Discovery Drive, Cambridge, Biomedical Campus, Cambridge, CB2 0AX, UK, and CA 72.4 were from MyBioSource, Inc. P.O. Box 153308, San Diego, CA 92195-3308, USA.

Further, it is not clear what biopsies are collected for. To confirm diagnosis of IBD or benign colonic lesions, or are they used for biomarker testing? Line 191 reads: "the rest of the biopsy sample", which indicates that something has been done to the biopsies, but this has not been described.
Reply: Biopsies were taken in order to confirm the detected findings (as in the case of IID or neoplasms) and to remove premalignant lesions as treatment in the case of polyps. The definitive diagnosis and the classification of each group was taking into account the result of the biopsy.

In the statistical section, a lot of statistical models are described. The authors state that extra patients were recruited to "compensate for expected losses". This extra inclusion should be described in the participants section, and a statement on the size of the expected loss should be included in the results section.

Reply: When the sample size was calculated, it was assumed that there would be a 10% loss. Following the recommendations of the TRIPOD Guide, almost all studies of prediction models have some missing result (missing values) of some predictor variables. This missing data, in our case, was less than 5%, so they were omitted from any analysis, but without performing the so-called full case analysis. We avoided including only participants with complete data, which is not only inefficient (as it can greatly reduce sample size), but can also lead to biased results when patients without missing data are not representative of the study sample full original.

Results:

In the Results section, it would be preferable, if the alle the selected clinical-epidemiology results were presented in a table, with the statistical test for differences between groups listed as well (especially for demographic variables, to minimize the risk of bias between groups). I see authors have referred to Table 1, but no table is accompanying the manuscript?

Reply: We attach Table 1 of Bivariate Analysis in the Tables section.

Likewise, the quantitative biomarkers levels for each group and whether they are different between groups would benefit from being displayed either as box plots or listed in a table (line 234-245).

Reply: Figures 1 and 2 are included in the article in the figures section.

When quantitative values are not disclosed, it is not possible to compare the values to other publications or standard clinical intervals, to assure that results are reasonable. Were the quantitative values normally distributed or not, are they comparable to other publications describing the same biomarkers? If these data are not presented in full, it is impossible to truly evaluate the result as a peer.

Reply: To check the goodness of fit to the normal of the numerical variables, the Shapiro-Wilk test was used. All the quantitative parameters evaluated showed a non-Gaussian pattern, with a significant degree of asymmetry, with a statistical behavior similar to that shown by these markers in other publications, which forced the use of non-parametric hypothesis contrast tests, such as Mann's U test. -Whitney (to compare two groups, CRC versus NO CRC) and the Kruskal Wallis test, for the comparison of the 4 initial groups of the study.

Line 248 describes that authors "proceeded to combine the most significant covariates form the bivariate analysis ($p < 0.25$) to use them jointly in multivariate diagnostic algorithms". However, the manuscript presents no univariate or bivariate analyses for me to see, and it does not list what variables are selected for the final model. As no result tables sum up all the analysed variables; the statistics used; and the p-values rendered, I have no possibility of finding the variables with a p-value above 0.25. included in the "multivariate diagnostic algorithms".

Reply: We attach Bivariate Analysis Table; Table 2 in the Tables section.

Authors do write that the final algorithm is subjected to intellectual property and therefore not disclosable; however, its is not possible for me as a peer to evaluate the scientific work if the results and the statistical models are not revealed properly and therefore I am unable to comment on whether the final multivariate algorithm seems a promising predictive biomarker for CRC screening, other than to solely rely on authors display of a nice AUC for their algorithm; this oppose the very nature of peer reviewing.
Reply: You are right in this indication. However, it is compatible, we think, to maintain the privacy imposed in commercial agreements with the dissemination and promotion of our work. We are totally open to any researcher who wants to learn about our work or carry out an external validation, to have our support. In addition, in this new version of the article we incorporate more research results, such as the bivariate analysis tables or the figures of the results of the selected markers, so that the final results of our work are easily deducible and interpretable.

In the aim, authors wish to investigate whether their biomarkers can discriminate between benign lesions, IBD and CRC. It is unclear whether the multivariate algorithm discriminate between CRC and controls or CRC and all other.
Reply: The algorithm fulfills the objective of discriminating CCR from the rest of the groups. In this case, the compared control group was that of "no cancer" and that they belonged to any of the other 3 groups.

Another study aim is to relate the biomarkers to other clinic-pathological features of the included patients. I do not find those results presented, neither as a Table or in the result section. It would be nice if it was.
Reply: There was no data with statistical correlation.

Discussion:

In the discussion section, authors suggest that their blood-based test could supplement CRC screening and prevent unnecessary colonoscopies. In such setting, it would be compared to iFOBT, the current marker for CRC screening. It would therefore be prudent if authors in the discussion reflected upon the specificity of iFOBT versus their new multimarker: The lower the specificity, the higher the number of unnecessary colonoscopies. Though a specificity of 79,7% for male and 83,1% for female is high, it would generate much more colonoscopies that iFOBT, which holds a specificity of 93%.
Reply: It is true, but the problem with iFOBT is that it decreases its sensitivity and therefore can generate false negatives, which is an important clinical problem. We think that when it comes to cancer detection, sensitivity should be prioritized over specificity. In any case, our algorithm is compatible with iFOBT since they would be complementary.

Another important point for discussion is the choice of CRC markers. The authors suggest some experimental markers (which might be included in the final algorithm?), which are only scarcely investigated in CRC: neutrophil-gelatinase-associated-lipocalin (NGAL), epidermal growth factor receptor (EGFR), Cyr61, and 8-hydrodeoxyguanosine (8OHdG).
However, NGAL is also well-established as a marker of kidney disease; EGFR is suggested a

marker of lung cancer; a meta-analysis from 2019 found that 8-OHdG in tumor tissues may be a predictor of prognosis in most solid tumors; and Cyr61 has been proposed as a biomarker for lung cancer.

That the proposed markers are also suggested as biomarkers in other diseases and cancer types present a problem, if they were to be used as specific markers for CRC screening; it is possible that individuals in a CRC screening cohort could have lung cancer or kidney disease, and then be false positive. The lack of marker specificity for CRC is thus an issue, which the authors need to address in the discussion section as well.

Reply: For this reason, the algorithm is in the population at risk of CRC, not in the general population. Currently there are few tumor markers that are not elevated in other clinical circumstances. We performed a study of comorbidities and ruled out the existence of other neoplasms. However, the lack of specificity of some markers is always a limitation, but in our case it was not an obstacle to discriminate between both groups.

Finally, the biomarker algorithm needs to be externally and independently validated with fixed cut-offs in a new and relevant cohort, to confirm the initial performance shown here. This is a necessary step, if the multimarker algorithm is to rise above the many previously published blood-based biomarkers for CRC screening. From the discussion, it seems that such a study is designed. The results from this future study would supplement and strengthen the results of this discovery study.

Reply: Yes, our new study is designed to carry out a multicenter validation and at the same time there are centers that are going to initiate their own validation of said algorithm.