ARTICLE DETAILS

| TITLE (PROVISIONAL) | Association between Proteinuria and Incident Colorectal Cancer: Analysis of a Nationwide Population-Based Database |
|---------------------|------------------------------------------------------------------------------------------------------------|
| AUTHORS             | Matsuoka, Satoshi; Kaneko, Hidehiro; Okada, Akira; Fukui, Akira; Yano, Yuichiro; Itoh, Hidetaka; Morita, Kojiro; Fujiu, Katsuhito; Michihata, Nobuaki; Jo, Taisuke; Takeda, Norifumi; Morita, Hiroyuki; Yamaguchi, Satoko; Nakamura, Sunao; Nishiyama, Akira; Yokoo, Takashi; Node, Koichi; Yamauchi, Toshimasa; Nangaku, Masaomi; Yasunaga, Hideo; Komuro, Issei |

VERSIGN 1 – REVIEW

| REVIEWER          | Sumida, Keiichi |
|-------------------|-----------------|
| University of Tennessee Health Science Center, Division of Nephrology, Department of Medicine |

| REVIEW RETURNED   | 30-Nov-2021 |

| GENERAL COMMENTS  | In a retrospective study of more than 2.8 million individuals who underwent health check-ups in Japan, Matsuoka et al. examined the association between dipstick proteinuria and incident CRC and found that trace and positive proteinuria were associated with a greater risk of incident CRC independent of potential confounders. Given the high burden of CRC and the increasing recognition of prognostic impact of proteinuria on several adverse clinical outcomes, the findings of this study may be of clinical importance with its strength of large sample size. The manuscript is generally well-written, and the authors acknowledge several limitations. However, there are some concerns and limitations that need to be addressed. |

Major:
1. One of the major limitations of this study is a potential misclassification of participants with proteinuria due to the use of random spot urine obtained at a single time point. Although the importance of the quantitative assessment for proteinuria has been acknowledged under the limitation, this possibility should also be mentioned.
2. Along this line, it would be interesting to see if a similar association can be observed for predicted ACR from urine dipstick protein using a recent dipstick-ACR conversion equation (PMID: 32658569).
3. Although the authors excluded participants undergoing hemodialysis or peritoneal dialysis in a sensitivity analysis, these participants should be excluded from the main analytical cohort as their urine protein seems to have little clinical significance for incident CRC.
4. The exclusion of several groups of participants in their sensitivity analyses may lose potentially important signal of the risk among...
them. I would instead suggest performing a subgroup analysis of participants with or without certain conditions (e.g., comorbidity, eGFR levels, mediation use).

5. I wonder if the authors assessed the proportional hazard assumption?

Minor:
1. Please spell out the JMDC at its first appearance in the manuscript.
2. The number (or proportion) of missingness should be provided for each variable with missing data.

---

**GENERAL COMMENTS**

The study examines the association of proteinuria, a marker of kidney function, with CRC risk. The study is of retrospective design and based on a large medical claims database in Japan. After multivariable adjustment, trace or high proteinuria was associated with an elevated risk of CRC, and this association remained in several sensitivity analyses.

The study appears to have been performed diligently and I find the manuscript well written and presented. The results are discussed pragmatically. I find some deficiencies, but these can be easily addressed.

Here are some specific comments:

**Abstract**

I find the term “negative proteinuria” to sound odd. Would it be better expressed as “no proteinuria/without proteinuria”?

**Introduction**

Line 90: for the known literature on CKD and CRC (refs 4-6), was CKD defined by the presence of proteinuria or using another clinical definition? It is not clear whether CKD and proteinuria can be considered interchangeable. Also, could the authors add a sentence of explanation on why these studies are conflicting and controversial?

More detail on exactly what originality the study brings would be appreciated here.

**Methods**

Line 127 and line 150: “Physical inactivity” is mentioned in the exclusions but not defined until later in the methods section. Perhaps this could be changed to “physical activity” for the exclusions only for clarity.

The methodology lacks some detail. What clinical assays were used for the measurement of proteins and other bioassays? Please could the authors mention the platform or instrumentation used?

Line 165: “Numbers” should be changed to “proportions” or similar.
Line 171: Were the authors able to include non-steroidal anti-inflammatory drugs (NSAIDs) in this adjustment? Their use is a known risk factor for CRC.

Should models have additionally been adjusted for assessment region and socio-economic status? I presume this data would be easily available.

Since ICD-10 codes C18, 19 or 20 were used, was information on tumour sub-site available? There is known heterogeneity between risk factors for colon and rectal tumours.

Given the large sample size and number of CRC cases recorded, it would be perhaps add novelty to perform another sensitivity analysis for early-onset CRC, i.e. before 55 years old.

Results

Line 212: The follow-up is quite short compared to the range of dates over which the data were collected. Were most medical checks taken towards the end of the period?

Line 219-221: It should be mentioned that the negative proteinuria group was the reference group.

Discussion

Is there any evidence from mendelian randomisation studies on the link between proteinuria and CRC to support the association found in the study?

The data suggests an association of proteinuria with CRC regardless of kidney function. Since this suggests metabolic perturbation is the driver of increased CRC risk, why did the authors not perform a sex-stratified analysis, rather than a sex-adjusted model only?

VERSION 1 – AUTHOR RESPONSE

Response to the Reviewer 1

Major:

1. One of the major limitations of this study is a potential misclassification of participants with proteinuria due to the use of random spot urine obtained at a single time point. Although the importance of the quantitative assessment for proteinuria has been acknowledged under the limitation, this possibility should also be mentioned.

Response:

Thank you very much for the important comment, and we agree with the reviewer. We described this point in the revised manuscript.
“Although we conducted a semiquantitative assessment using the urine dipstick test in this study, the latest guidelines recommend quantitative assessment for proteinuria using calculation of the random urine protein-to-creatinine ratio, urine albumin-to-creatinine ratio, and 24-h urine collection testing. Because we used the random spot urine assessed at a single occasion in this study, we need to acknowledge the possibility of misclassification.”

2. Along this line, it would be interesting to see if a similar association can be observed for predicted ACR from urine dipstick protein using a recent dipstick-ACR conversion equation (PMID: 32658569).

Response:
Thank you for your constructive feedback. Using this conversion equation, we conducted Cox regression analyses to identify the association between predicted ACR and the incidence of CRC. As shown below, our main results did not change.

Changes:

Page 7, line 10-12 (METHODS):
“Urine dipstick protein was rated on a five-point scale: negative, trace, +, ++, ++++, and we categorized the study participants as no proteinuria, trace proteinuria (10–20 mg/dL), or positive proteinuria (≥ 30 mg/dL, urine dipstick protein ≥ +).”

Page 10, line 2-8 (METHODS):
“Fifth, we divided participants into three groups according to predicted urine albumin-to-creatinine ratio (pACR) (< 30 mg/g, 30-299 mg/g, ≥ 300 mg/g) and examined the association of pACR category with incident CRC using multivariable analysis. We calculated pACR using the following conversion equation for urine dipstick protein adjusted for sex, hypertension, and diabetes mellitus: pACR = exp (2.0373 + 0.7270 [if trace] + 1.6775 [if +] + 3.2622 [if ++] + 4.5435 [if ++++] + 0.0822 [if female] + 0.27249 [if diabetes mellitus] + 0.33627 [if hypertension]).

Page 13, line 13-15 (RESULTS):
“Fifth, we investigated the relationship between pACR and incident CRC using multivariable analysis. The risk of CRC increased with pACR (Supplemental table 4).
Supplemental Table 4. The Frequency of Events, Corresponding Incidence Rates, and Hazard Ratios for Colorectal Cancer Events After Classification by Predicted Urine Albumin-to-creatinine Ratio

| predicted urine albumin-to-creatinine | < 30 mg/g (n=2,665,780) | 30-299 mg/g (n=72,576) | ≥ 300 mg/g (n=6,940) |
|--------------------------------------|-------------------------|------------------------|----------------------|
| No. of Colorectal Cancer Events      | 10,206                  | 353                    | 56                   |
| Incidence Rate (per 10,000 person-years) | 11.7 (11.5-12.0) | 15.3 (13.8-17.0) | 25.6 (19.7-33.2) |
| Model 1 (Unadjusted)                 | 1 [Reference]           | 1.36 (1.22-1.51)      | 2.39 (1.84-3.10)    |
| Model 2                              | 1 [Reference]           | 1.31 (1.18-1.45)      | 1.71 (1.31-2.22)    |
| Model 3                              | 1 [Reference]           | 1.25 (1.12-1.39)      | 1.58 (1.21-2.06)    |

We analyzed 2,745,296 participants. The incidence rate was per 10,000 person-years. Predicted urine albumin-to-creatinine was calculated using an equation that adjusted for sex, hypertension, and diabetes mellitus. Unadjusted and adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group are shown. Model 1 is unadjusted. Model 2 includes adjustment for age. Model 3 includes adjustment for age, obesity, dyslipidemia, cigarette smoking, alcohol consumption, and physical inactivity.
3. Although the authors excluded participants undergoing hemodialysis or peritoneal dialysis in a sensitivity analysis, these participants should be excluded from the main analytical cohort as their urine protein seems to have little clinical significance for incident CRC.

Response:
We appreciate this educational comment. As pointed out, we revised the exclusion criteria and excluded participants with a prior history of CRC, colorectal disease, renal disease, and renal replacement therapy (n=114,888) from the main cohort. We performed all analysis using this cohort, and found that our primary conclusion did not change.

Changes:

Page 2, line 8-14 (ABSTRACT):

“PARTICIPANTS: We selected records of participants (n=3,543,705) who underwent health check-ups, including physical examinations, blood tests, and urine dipstick tests. We excluded participants who were aged < 20 years (n=25,577), had a history of CRC, colorectal disease, renal disease, and renal replacement therapy (n=114,888), or had missing data on medications (n=170,145), cigarette smoking (n=14,835), alcohol consumption (n=366,414), or physical activity (n=106,550). Finally, we analyzed 2,745,296 participants.”

Page 2, line 16-27 (ABSTRACT):

“RESULTS: Participants were categorized as having no proteinuria (n=2,435,872), trace proteinuria (n=231,153), or positive proteinuria (n=78,271). Over a mean follow-up period of 1,189±914 days, 10,615 CRC diagnoses were recorded. The incidence of CRC (95% confidence interval [CI]) was lowest in participants without proteinuria (11.7; 95% CI, 11.5-11.9 per 10,000 person-years), followed by trace proteinuria (12.5; 95% CI, 11.7-13.3 per 10,000 person-years), and positive proteinuria (16.1; 95% CI, 14.6-17.7 per 10,000 person-years). After multivariable adjustment, compared with no proteinuria, hazard ratios for incident CRC were 1.20 (95% CI, 1.12-1.29) and 1.23 (95% CI, 1.11-1.36) for trace and positive proteinuria, respectively. The association between proteinuria and incident CRC existed in participants after multiple imputations for missing data, with a follow-up period of ≥ 365 days, regardless of age, sex, obesity, hypertension, diabetes mellitus, and estimated glomerular filtration rate.”

Page 6, line 11-17 (METHODS):

“For the current analyses, we selected records of participants (n=3,543,705) who underwent health check-ups, including physical examinations, blood tests, and urine dipstick test. We excluded participants who were aged <20 years (n=25,577), had a history of CRC, colorectal disease, renal disease, and renal replacement therapy (n=114,888), or had missing data on medications for hypertension (n=169,239), diabetes mellitus (n=698), or dyslipidemia (n=208), cigarette smoking (n=14,835), alcohol consumption (n=366,414), or physical activity (n=106,550) (Figure 1).”
For the current analyses, we selected records of individuals (n=3,543,705) who underwent health check-up including physical examination, blood test, and urine dipstick test enrolled in the JMDC Claims Database between January 2005 and April 2020. We excluded individuals < 20 years of age (n=25,577), those with a history of CRC (ICD-10 codes C18, C19, C20), colorectal disease, renal disease, and renal replacement therapy (n=114,888), and missing data on medications for hypertension, diabetes mellitus, or dyslipidemia (n=170,145), cigarette smoking (n=14,835), alcohol consumption (n=366,414), and physical activity (n=106,550). After all exclusion criteria were applied, data from 2,745,296 individuals were analyzed in this study.

**Page 9, line 12-17 (METHODS):**

“Third, we excluded participants who were diagnosed with CRC but had no confirmed treatment history. We defined colon resection (procedure code: K719), colorectal mucosal resection (procedure code: K721), rectal resection (procedure code: K740), and others (procedure codes: K726, K728, K732, and K736) as surgery for CRC, and the use of fluorouracil, irinotecan, oxaliplatin, or capecitabine as chemotherapy for CRC.”

**Page 11, line 6-12 (RESULTS):**

“The characteristics of the study participants (n=2,745,296) are shown in Table 1. The mean age was 45.3 ± 11.1 years, and 1,554,705 participants (56.6%) were men. Using the results of the urine dipstick test at the health check-up, participants were categorized as having no proteinuria (n=2,435,872), trace proteinuria (n=231,153), or positive proteinuria (n=78,271). Compared to those without proteinuria, participants with trace or positive proteinuria were more likely to be male. The prevalence of obesity, hypertension, diabetes mellitus, dyslipidemia, and cigarette smoking increased with proteinuria.”
"During a mean follow-up period of 1,189 ± 914 days, there were 10,615 incident CRC diagnoses (C18, C19, and C20 accounted for 8,591 [80.9%], 374 [3.5%], and 3,021 [28.5%], respectively, and 1,249 participants had two or more diagnoses). The incidence (95% confidence interval [CI]) of CRC was the lowest in participants without proteinuria (11.7 [95% CI, 11.5-11.9] per 10,000 person-years), followed by trace proteinuria (12.5 [95% CI, 11.7-13.3] per 10,000 person-years), and positive proteinuria (16.1 [95% CI, 14.6-17.7] per 10,000 person-years). In the unadjusted model (Model 1) and age- and sex-adjusted model (Model 2), trace proteinuria and positive proteinuria were associated with a higher risk of CRC events than no proteinuria. After multivariable adjustment for covariates, the HRs (95% CIs) for CRC events were 1.20 (95% CI, 1.12-1.29) for trace proteinuria and 1.23 (95% CI, 1.11-1.36) for positive proteinuria, respectively, compared to those without proteinuria (Figure 2)."
Page 12, line 16-Page 13, line 12 (RESULTS):

"First, we analyzed 3,403,240 participants after multiple imputations for missing data. During the follow-up period, 12,943 CRC events occurred. In these participants, trace proteinuria (HR, 1.19; 95% CI, 1.12-1.27) and positive proteinuria (HR, 1.25; 95% CI, 1.15-1.38) were both associated with an elevated incidence of CRC compared with no proteinuria (Supplemental table 1). Second, we excluded participants whose follow-up period for CRC was < 365 days and analyzed 2,263,006 participants with a follow-up period of ≥ 365 days. In these participants, compared with no proteinuria, trace proteinuria (HR, 1.20; 95% CI, 1.11-1.28) and positive proteinuria (HR, 1.24; 95% CI, 1.11-1.37) were associated with a greater risk of CRC (Supplemental table 2). Third, among 10,615 participants with a CRC diagnosis, we confirmed treatment for CRC (surgery or chemotherapy) in 7,927 participants (74.7%). We excluded 2,688 participants who were diagnosed with CRC but not confirmed as undergoing treatment for CRC. Even in this model, compared with no proteinuria, trace proteinuria (HR, 1.20; 95% CI, 1.10 to 1.30) and positive proteinuria (HR, 1.26; 95% CI, 1.12 to 1.41) were associated with a higher incidence of CRC (Table 2)."

---

**Table 1. The Frequency of Events, Corresponding Incidence Rates, and Hazard Ratios for Colorectal Cancer Events after Multiple Imputations for Missing Data**

| Proteinuria Category | No. (n=3,009,218) | Trace (n=294,120) | Positive (n=99,902) |
|----------------------|-------------------|------------------|---------------------|
| No. of Colorectal Cancer Events | 11,344 | 1,092 | 507 |
| Incidence Rate (per 10,000 person-years) | 11.2 (11.0-11.4) | 11.9 (11.2-12.7) | 15.7 (14.4-17.1) |
| Model 1 (Unadjusted) | [Reference] | 1.18 (1.11-1.26) | 1.50 (1.38-1.64) |
| Model 2 | [Reference] | 1.22 (1.15-1.30) | 1.38 (1.26-1.51) |
| Model 3 | [Reference] | 1.19 (1.12-1.27) | 1.25 (1.15-1.38) |

We analyzed 3,403,240 participants after multiple imputation for missing data. The incidence rate was per 10,000 person-years. Unadjusted and adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group are shown. Model 1 is unadjusted. Model 2 includes adjustment for age and sex. Model 3 includes adjustment for age, sex, obesity, hypertension, diabetes mellitus, dyslipidemia, cigarette smoking, alcohol consumption, and physical inactivity.

**Table 2. The Frequency of Events, Corresponding Incidence Rates, and Hazard Ratios for Colorectal Cancer Events in Participants with Follow-Up Period ≥ 365 Days**

| Proteinuria Category | No. (n=2,009,876) | Trace (n=189,269) | Positive (n=65,861) |
|----------------------|-------------------|------------------|---------------------|
| No. of Colorectal Cancer Events | 8,981 | 829 | 385 |
| Incidence Rate (per 10,000 person-years) | 15.6 (15.3-15.9) | 17.0 (15.9-18.2) | 21.7 (19.6-24.0) |
| Model 1 (Unadjusted) | [Reference] | 1.19 (1.10-1.27) | 1.45 (1.31-1.61) |
| Model 2 | [Reference] | 1.22 (1.14-1.32) | 1.35 (1.22-1.49) |
| Model 3 | [Reference] | 1.20 (1.11-1.28) | 1.24 (1.11-1.37) |

We analyzed 2,263,006 participants with follow-up period ≥ 365 days. The incidence rate was per 10,000 person-years. Unadjusted and adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group are shown. Model 1 is unadjusted. Model 2 includes adjustment for age and sex. Model 3 includes adjustment for age, sex, obesity, hypertension, diabetes mellitus, dyslipidemia, cigarette smoking, alcohol consumption, and physical inactivity.
4. The exclusion of several groups of participants in their sensitivity analyses may lose potentially important signal of the risk among them. I would instead suggest performing a subgroup analysis of participants with or without certain conditions (e.g., comorbidity, eGFR levels, mediation use).

Response:

Thank you for the important suggestion. We performed subgroup analyses as shown below. Our main results were consistent in all subgroups.

Changes:

Page 9, line 7-8 (METHODS):

“We conducted seven sensitivity analyses.”

Page 10, line 8-10 (METHODS):

“Sixth, we performed subgroup analyses stratified by age (≥ 50 years vs. < 50 years), sex, obesity, hypertension, diabetes mellitus, and eGFR (≥ 60ml/min/1.73m² vs. < 60ml/min/1.73m²).”

Page 13, line 15-16 (RESULTS):

“Sixth, the relationship between proteinuria and incident CRC was present in all subgroups (Figure 3).”
### Figure 3. Subgroup Analyses

| Proteinuria | Number | Colorectal Cancer | Hazard Ratio |
|-------------|--------|-------------------|--------------|
| Age ≥ 50 Years | | | |
| No | 862231 | 6154 | 1 [Reference] |
| Trace | 67625 | 549 | 1.17 (1.07-1.28) |
| Positive | 27979 | 282 | 1.23 (1.09-1.39) |
| Age < 50 Years | | | |
| No | 1573641 | 3190 | 1 [Reference] |
| Trace | 163528 | 322 | 1.25 (1.11-1.40) |
| Positive | 50292 | 118 | 1.18 (0.98-1.42) |
| Men | | | |
| No | 1360816 | 6288 | 1 [Reference] |
| Trace | 142049 | 649 | 1.20 (1.11-1.30) |
| Positive | 51840 | 318 | 1.22 (1.06-1.37) |
| Women | | | |
| No | 1075956 | 3926 | 1 [Reference] |
| Trace | 89104 | 222 | 1.19 (1.04-1.36) |
| Positive | 26431 | 82 | 1.23 (0.98-1.53) |
| Obesity (-) | | | |
| No | 1862241 | 6752 | 1 [Reference] |
| Trace | 166729 | 554 | 1.17 (1.07-1.27) |
| Positive | 47464 | 222 | 1.30 (1.14-1.49) |
| Obesity (+) | | | |
| No | 573031 | 2392 | 1 [Reference] |
| Trace | 64124 | 317 | 1.26 (1.12-1.42) |
| Positive | 30803 | 178 | 1.15 (0.98-1.34) |

Hazard Ratios (95% Confidence intervals)

### Figure 3. Subgroup Analyses (Continued)

| Proteinuria | Number | Colorectal Cancer | Hazard Ratio |
|-------------|--------|-------------------|--------------|
| Hypertension (-) | | | |
| No | 199403 | 6339 | 1 [Reference] |
| Trace | 185842 | 552 | 1.23 (1.13-1.35) |
| Positive | 50940 | 173 | 1.21 (1.04-1.40) |
| Hypertension (+) | | | |
| No | 441789 | 5005 | 1 [Reference] |
| Trace | 45511 | 319 | 1.14 (1.02-1.28) |
| Positive | 27331 | 227 | 1.25 (1.07-1.43) |
| Diabetes Mellitus (-) | | | |
| No | 2339525 | 8593 | 1 [Reference] |
| Trace | 217223 | 754 | 1.21 (1.12-1.30) |
| Positive | 64277 | 269 | 1.18 (1.05-1.33) |
| Diabetes Mellitus (+) | | | |
| No | 96147 | 751 | 1 [Reference] |
| Trace | 13930 | 117 | 1.18 (0.97-1.43) |
| Positive | 13994 | 131 | 1.34 (1.11-1.62) |
| eGFR ≥ 60 ml/min/1.73 m² | | | |
| No | 918625 | 2863 | 1 [Reference] |
| Trace | 90556 | 261 | 1.24 (1.08-1.41) |
| Positive | 28569 | 107 | 1.19 (0.98-1.45) |
| eGFR < 60 ml/min/1.73 m² | | | |
| No | 45943 | 273 | 1 [Reference] |
| Trace | 4554 | 37 | 1.69 (1.19-2.39) |
| Positive | 3216 | 22 | 1.20 (0.77-1.88) |

Hazard Ratios (95% Confidence intervals)
We performed subgroup analyses by age (≥ 50 years vs. < 50 years), sex, obesity, hypertension, diabetes mellitus, and eGFR (≥ 60ml/min/1.73m² vs. < 60ml/min/1.73m²) in the multivariable model. Adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group and the forest plot were shown.

Page 14, line 4-8 (DISCUSSION):

“In this nationwide analysis of a health claims database that had adults who underwent health check-ups, including the dipstick urine test, those with trace and positive proteinuria had a higher risk of incident CRC events than those with no proteinuria, even after adjusting for multiple potential confounders. This association was consistent in all subgroups stratified by age, sex, obesity, hypertension, diabetes mellitus, or eGFR.”

Page 15, line 9-13 (DISCUSSION):

“Interestingly, in a subgroup of participants with eGFR < 60 ml/min/1.73m², the HRs for both trace and positive proteinuria did not change. Therefore, eGFR did not affect the relationship between proteinuria and incident CRC, and proteinuria could influence the incidence of CRC independently of renal function.”

5. I wonder if the authors assessed the proportional hazard assumption?

Response:

We appreciate your erudite comment. We checked the proportional hazard assumption using Schoenfeld residual tests and there was no breach of this hypothesis.

Changes:

Page 9, line 6-7 (METHODS):

“We checked the proportional hazard assumption using Schoenfeld residual tests.”

Page 12, line 9-12 (RESULTS):

“We checked the proportional hazard assumption by Schoenfeld residual tests, and there was no breach of this hypothesis. The Schoenfeld residual global test also showed no breach of that hypothesis (p=0.66 for trace proteinuria and p=0.07 for positive proteinuria).”

Minor:

1. Please full spell out the JMDC at its first appearance in the manuscript.

Response:

JMDC is an abbreviation of “Japan Medical Data Center” previously. However, JMDC is currently an official name.
2. The number (or proportion) of missingness should be provided for each variable with missing data.

Response:
We appreciate your comment. We described it in the revised manuscript.

Changes:

Page 2, line 8-14 (ABSTRACT):

“PARTICIPANTS: We selected records of participants (n=3,543,705) who underwent health check-ups, including physical examinations, blood tests, and urine dipstick tests. We excluded participants who were aged < 20 years (n=25,577), had a history of CRC, colorectal disease, renal disease, and renal replacement therapy (n=114,888), or had missing data on medications (n=170,145), cigarette smoking (n=14,835), alcohol consumption (n=366,414), or physical activity (n=106,550). Finally, we analyzed 2,745,296 participants”

Page 6, line 11-17 (METHODS):

“For the current analyses, we selected records of participants (n=3,543,705) who underwent health check-ups, including physical examinations, blood tests, and urine dipstick test. We excluded participants who were aged < 20 years (n=25,577), had a history of CRC, colorectal disease, renal disease, and renal replacement therapy (n=114,888), or had missing data on medications for hypertension (n=169,239), diabetes mellitus (n=698), or dyslipidemia (n=208), cigarette smoking (n=14,835), alcohol consumption (n=366,414), or physical activity (n=106,550) (Figure 1).”

Figure 1. Flowchart

For the current analyses, we selected records of individuals (n=3,543,705) who underwent health check-up including physical examination, blood test, and urine dipstick test enrolled in the JMDC Claims Database between January 2005 and April 2020. We excluded individuals < 20 years of age (n=25,577), those with a history of CRC (ICD-10 codes C18, C19, C20), colorectal disease, renal disease, and renal replacement therapy (n=114,888), and missing data on medications for...
hypertension, diabetes mellitus, or dyslipidemia (n=170,145), cigarette smoking (n=14,835), alcohol consumption (n=366,414), and physical activity (n=106,550). After all exclusion criteria were applied, data from 2,745,296 individuals were analyzed in this study.
Response to the Reviewer 2

1. I find the term “negative proteinuria” to sound odd. Would it be better expressed as “no proteinuria/without proteinuria”? (Abstract)

Response/ Changes:
Thank you very much for this important comment. We agreed with the reviewer. We changed the description of “negative proteinuria” to “no proteinuria” or “without proteinuria” throughout the manuscript.

2. Line 90: for the known literature on CKD and CRC (refs 4-6), was CKD defined by the presence of proteinuria or using another clinical definition? It is not clear whether CKD and proteinuria can be considered interchangeable. Also, could the authors add a sentence of explanation on why these studies are conflicting and controversial? More detail on exactly what originality the study brings would be appreciated here. (Introduction)

Response:
We apologize for the lack of clarity. As mentioned below, previous reports (reference #4-6) were about the relationship between CKD (defined using eGFR value) and incident cancer. However, these did not explore the relationship between proteinuria to incident cancer. Our study's originality lies in using a large database to clinically clarify the relationship between proteinuria and incident cancer in a general population.

Changes:
Page 4, line 9-17 (INTRODUCTION):
“Although previous cohort studies reported the clinical between CKD (defined using estimated glomerular filtration rate [eGFR]) and incident cancer, the results of these studies lacked consistency.4 6 In the subgroup of older men, CKD was associated with a higher risk for incident lung and urinary tract cancers.4 Similarly, CKD was reported a higher renal and urothelial cancer risk.5 On the other hand, another cohort reported that CKD with diabetes mellitus was not related to incident overall cancer.6 Therefore, clinical evidence on the association between CKD and the risk of developing cancer is conflicting and controversial.4 6

Due to the above change, we switched the current #5 and the current #6 references in the revised manuscript.

3. Line 127 and line 150: “Physical inactivity” is mentioned in the exclusions but not defined until later in the methods section. Perhaps this could be changed to “physical activity” for the exclusions only for clarity. (Methods)
Response:
We agree with the reviewer’s comment. We changed “physical inactivity” to “physical activity” according to the reviewer’s comment.

4. The methodology lacks some detail. What clinical assays were used for the measurement of proteins and other bioassays? Please could the authors mention the platform or instrumentation used. (Methods)

Response:
We apologize for the lack of clarity. Unfortunately, information on clinical assays used for the measurement of proteinuria and laboratory tests are not available because the details of the testing methods used in health check-ups are left up to each institute. We described it as a study limitation.

Change:
Page 18, line 16-18 (DISCUSSION):
“Detailed information on clinical assays used for the measurement of proteinuria and laboratory tests are not available because the details of the testing methods used in health check-ups are left up to each institute.”

5. Line 165: “Numbers” should be changed to “proportions” or similar. (Methods)

Response:
Thank you very much for this important comment. We changed “percentages” to “proportions” in the text where it is applicable.

Change: (METHODS)
Page 8, line 16-17:
“Descriptive statistics are reported as medians (interquartile ranges) for continuous data and numbers (proportions) for categorical data.”

6. Line 171: Were the authors able to include non-steroidal anti-inflammatory drugs (NSAIDs) in this adjustment? Their use is a known risk factor for CRC. (Methods)

Response:
We appreciate this practical comment. We included the use of non-steroidal anti-inflammatory drugs (NSAIDs) to the multivariable model. Even after this adjustment, our main results did not change.
Changes:

Page 9, line 17- Page 10, line 1 (METHODS):

“Fourth, we included the use of non-steroidal anti-inflammatory drugs (NSAIDs) in the multivariable model. We defined the use of NSAIDs (ATC-code: M01A) within 90 days from the health check-up date.”

Page 13, line 12-13 (RESULTS):

“Fourth, even after adjustment for the use of NSAIDs, our primary results did not change (Supplemental table 3).”

| Proteinuria Category | No (n=2,435,872) | Trace (n=231,153) | Positive (n=78,271) |
|----------------------|------------------|-----------------|-------------------|
| No. of Colorectal Cancer Events | 9,344            | 871             | 400               |
| Incidence Rate (per 10,000 person-years) | 11.7 (11.5-11.9) | 12.5 (11.7-13.3) | 16.1 (14.6-17.7) |
| Model 3               | 1 [Reference]    | 1.20 (1.12-1.29) | 1.23 (1.11-1.36) |
| Model 3 + the use of NSAIDs | 1 [Reference]    | 1.20 (1.12-1.29) | 1.23 (1.11-1.36) |

Supplemental table 3. The Frequency of Events, Corresponding Incidence Rates, and Hazard Ratios for Colorectal Cancer Events Adjusted for Covariates Including the Use of Non Steroidal Anti-Inflammatory Drugs

We analyzed 2,745,296 participants. The incidence rate was per 10,000 person-years. Adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group are shown. Model 3 includes adjustment for age, sex, obesity, hypertension, diabetes mellitus, dyslipidemia, cigarette smoking, alcohol consumption, and physical inactivity.
7. Should models have additionally been adjusted for assessment region and socio-economic status? I presume this data would be easily available. (Methods)

Response:
Unfortunately, there is no information on region or socio-economic status in the JMDC Claim Database. We clearly described it as a study limitation.

Changes:
Page 17, line 17-Page 18, line 1 (DISCUSSION):
“The JMDC Claim Database does not include information on region or socioeconomic status. Further research is necessary to generalize our findings.”

8. Since ICD-10 codes C18, 19 or 20 were used, was information on tumour sub-site available? There is known heterogeneity between risk factors for colon and rectal tumours. (Methods)

Response:
Thank you for your helpful comment. Accordingly, we analyzed the association of proteinuria with the risk for C18, C19, or C20, separately.

Changes:
Page 10, line 10-11 (METHODS):
“Seventh, we analyzed the relationship of proteinuria with the risk for C18, C19, or C20, separately.”

Page 13, line 16-18 (RESULTS):
“Seventh, the association of proteinuria with a greater risk for CRC development was consistent in all CRC subsites separated by C18, C19, or C20 (Supplemental table 5).”
Supplemental table 5. The Frequency of Events, Corresponding Incidence Rates, and Hazard Ratios for Each Incident Diagnosis with C18, C19, or C20 Adjusted for Covariates

| Proteinuria Category | No. of Colon Cancer Events (ICD10 code: C18) | Trace | Positive |
|----------------------|--------------------------------------------|-------|----------|
|                      | (n=2,435,872)                              | (n=231,153) | (n=78,271) |
|                      | 7,562                                      | 702   | 327      |
| Incidence Rate (per 10,000 person-years) | 9.5 (9.3-9.7)                              | 10.1 (9.4-10.9) | 13.2 (11.8-14.7) |
| Model 1              | 1 [Reference]                             | 1.06 (1.00-1.14) | 1.38 (1.24-1.55) |
| Model 2              | 1 [Reference]                             | 1.12 (1.04-1.21) | 1.30 (1.16-1.45) |
| Model 3              | 1 [Reference]                             | 1.09 (1.01-1.18) | 1.19 (1.06-1.33) |

| No. of Rectosigmoid Junction Cancer Events (ICD10 code: C19) | 318 | 35 | 21 |
| Incidence Rate (per 10,000 person-years) | 0.4 (0.3-0.4) | 0.5 (0.4-0.7) | 0.8 (0.6-1.3) |
| Model 1 | 1 [Reference] | 1.26 (1.08-1.49) | 2.12 (1.36-3.30) |
| Model 2 | 1 [Reference] | 1.34 (0.95-1.90) | 1.96 (1.26-3.05) |
| Model 3 | 1 [Reference] | 1.28 (0.90-1.82) | 1.68 (1.07-2.64) |

| No. of Rectum Cancer Events (ICD10 code: C20) | 2,650 | 250 | 121 |
| Incidence Rate (per 10,000 person-years) | 3.3 (3.2-3.4) | 3.6 (3.2-4.1) | 4.9 (4.1-5.8) |
| Model 1 | 1 [Reference] | 1.08 (0.95-1.23) | 1.47 (1.22-1.76) |
| Model 2 | 1 [Reference] | 1.13 (0.99-1.28) | 1.35 (1.13-1.62) |
| Model 3 | 1 [Reference] | 1.10 (0.97-1.25) | 1.24 (1.03-1.49) |

We analyzed 2,745,296 participants for the hazard ratio for each incident diagnosis with C18, C19, or C20 among those with colorectal cancer. The incidence rate was per 10,000 person-years. Unadjusted and adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group are shown. Model 1 is unadjusted. Model 2 includes adjustment for age and sex. Model 3 includes adjustment for age, sex, obesity, hypertension, diabetes mellitus, dyslipidemia, cigarette smoking, alcohol consumption, and physical inactivity.
Given the large sample size and number of CRC cases recorded, it would be perhaps add novelty to perform another sensitivity analysis for early-onset CRC, i.e. before 55 years old. (Methods)

The data suggests an association of proteinuria with CRC regardless of kidney function. Since this suggests metabolic perturbation is the driver of increased CRC risk, why did the authors not perform a sex-stratified analysis, rather than a sex-adjusted model only? (Discussion)

Response:
Please kindly allow us to respond to these two comments simultaneously. We performed subgroup analyses by age and sex based on the reviewer’s feedback. Regarding the analysis for early-onset CRC, we divided study participants using the age at study enrollment.

Changes:

Page 10, line 8-10 (METHODS):
“Sixth, we performed subgroup analyses stratified by age (≥ 50 years vs. < 50 years), sex, obesity, hypertension, diabetes mellitus, and eGFR (≥ 60ml/min/1.73m² vs. < 60ml/min/1.73m²).”

Page 13, line 15-16 (RESULTS):
“Sixth, the relationship between proteinuria and incident CRC was present in all subgroups (Figure 3).”

Figure 3. Subgroup Analyses

| Proteinuria | Number | Colorectal Cancer | Hazard Ratio |
|-------------|--------|-------------------|--------------|
| Age ≥ 50 Years | No | 862231 | 6154 | 1 [Reference] |
| Trace | 67625 | 549 | 1.17 (1.07-1.28) |
| Positive | 27979 | 242 | 1.21 (1.09-1.35) |
| Age < 50 Years | No | 1575641 | 3190 | 1 [Reference] |
| Trace | 163528 | 322 | 1.25 (1.11-1.40) |
| Positive | 50292 | 118 | 1.18 (0.98-1.42) |
| Men | No | 1360816 | 6288 | 1 [Reference] |
| Trace | 142049 | 649 | 1.20 (1.11-1.30) |
| Positive | 51840 | 318 | 1.22 (1.06-1.41) |
| Women | No | 1075056 | 3456 | 1 [Reference] |
| Trace | 89104 | 222 | 1.19 (1.04-1.36) |
| Positive | 26431 | 82 | 1.23 (0.98-1.55) |
| Obesity (-) | No | 1862241 | 6752 | 1 [Reference] |
| Trace | 164729 | 554 | 1.17 (1.07-1.27) |
| Positive | 47468 | 222 | 1.30 (1.14-1.49) |
| Obesity (+) | No | 573031 | 2392 | 1 [Reference] |
| Trace | 64124 | 317 | 1.26 (1.12-1.42) |
| Positive | 30803 | 178 | 1.15 (0.98-1.34) |

Hazard Ratios (95% Confidence intervals)
We performed subgroup analyses by age (≥ 50 years vs. < 50 years), sex, obesity, hypertension, diabetes mellitus, and eGFR (≥ 60ml/min/1.73m² vs. < 60ml/min/1.73m²) in the multivariable model. Adjusted hazard ratios (95% confidence intervals) associated with the proteinuria group and the forest plot were shown.

10. Line 212: The follow-up is quite short compared to the range of dates over which the data were collected. Were most medical checks taken towards the end of the period? (Results)

Response:
We apologize for the lack of clarity. As the reviewer pointed out, the JMDC Claims Database collected data between 2005 and 2020. Study participants are continuously enrolled in this database, and not all participants have been followed since 2005. Therefore, the longest observation period is 4,774 days, but the shortest observation period is only 1 day.

11. Line 219-221: It should be mentioned that the negative proteinuria group was the reference group. (Results)

Response:
We apologize for the lack of clarity. We revised this sentence as shown below.

Change

Page 12, line 6-9 (RESULTS):

“After multivariable adjustment for covariates, the HRs (95% CIs) for CRC events were 1.20 (95% CI, 1.12-1.29) for trace proteinuria and 1.23 (95% CI, 1.11-1.36) for positive proteinuria, respectively, compared to those without proteinuria (Figure 2).”

12. Is there any evidence from mendelian randomisation studies on the link between proteinuria and CRC to support the association found in the study? (Discussion)

Response:

We agree with the reviewer that mendelian randomization studies are important to determine the association of proteinuria with incident CRC. As far as we know, there has been no mendelian randomization studies on the relationship between proteinuria and CRC.

VERSION 2 – REVIEW

| REVIEWER | Sumida, Keiichi  
| University of Tennessee Health Science Center, Division of Nephrology, Department of Medicine |
| REVIEW RETURNED | 27-Jan-2022 |

GENERAL COMMENTS | The manuscript has been improved after revision. I have no further comments. |

| REVIEWER | Rothwell, Joseph  
| INSERM U1018 Equipe 11, Exposome and Heredity team |
| REVIEW RETURNED | 01-Feb-2022 |

GENERAL COMMENTS | The authors have rigorously addressed all the issues raised and I have no further comments. |