

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	The Association of Dietary Quality with Colorectal Cancer among Normal Weight, Overweight, and Obese Men and Women: A Prospective Longitudinal Study in the United States
<b>AUTHORS</b>	Torres Stone, Rosalie; Waring, Molly; Cutrona, Sarah; Kiefe, Catarina; Allison, Jeroan; Doubeni, Chyke A

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Gertraud Maskarinec University of Hawaii Cancer Center Honolulu, HI, USA
<b>REVIEW RETURNED</b>	12-Jan-2017

<b>GENERAL COMMENTS</b>	<p>This manuscript addresses an important study question, the potential influence of diet quality on the development of colorectal cancer. The analysis is based on a very large well defined study population with more than 6,000 incident cases, the statistical methods are appropriate, and the findings are interesting. However, the data presentation and the text requires some improvement.</p> <p>More detailed comments:</p> <ul style="list-style-type: none"><li>• Given the observational study design, it would be more appropriate to say "is associated" than "protects".</li><li>• In the Abstract, please add the total number of CRC cases by sex and BMI status and the mean follow-up time to the Results. Instead of stating that more research is needed, the conclusions could emphasize the differential relation of diet quality by BMI status.</li><li>• In the Public Health Implications, the reference to food environment is inappropriate as the current analysis did not examine any features of food environment, just individual dietary patterns. The last sentence about more research is needed can also be removed.</li><li>• The Introduction is very lengthy and part of the presentation about previous research can be moved to the Discussion. The sentence about interventions is not directly relevant to this analysis.</li><li>• Please add to the Methods why underweight participants and those with a family history of CRC were excluded. The sentence about health status is not needed as that variable was not included in the models.</li><li>• The description of the Cox regression is unclear. Do the authors mean that length of follow-up was the underlying time metric and not</li></ul>
-------------------------	---

	<p>age?</p> <ul style="list-style-type: none"> <li>• Was anything known about comorbidities, such as diabetes and hypertension among the study participants? If yes, did they affect CRC risk?</li> <li>• The Results section needs to report the total number of CRC cases by stage at diagnosis, mean follow-up time, overall hazard ratios for the total study population, and p-values for the interaction terms of BMI with dietary indices. It may also be useful to stratify the models by early vs. late stage at diagnosis.</li> <li>• In the interest of space, Table 2 could be deleted. Tables 3a and 3b as well as 4a and 4b could be combined.</li> <li>• In the Discussion, any reference to the food environment can be removed as this paper addresses eating behavior as assessed by dietary patterns based on the assessment of individual dietary intakes.</li> <li>• Hormonal factors may also responsible for sex differences in CRC etiology. Previous research has reported association with hormone treatment.</li> <li>• In the conclusions, it would help to disentangle whether the authors think that dietary quality is important as a risk factor for obesity, which increases CRC risk, or whether dietary quality may be an independent predictor of CRC risk.</li> </ul> <p>Minor comments:</p> <ul style="list-style-type: none"> <li>• In the article summary, the number of participants was typed incorrectly.</li> <li>• The reference style in the text is incorrect and the formatting of the bibliography is off in several places.</li> </ul>
--	--

<b>REVIEWER</b>	Teresa Fung Simmons College, USA
<b>REVIEW RETURNED</b>	18-Jan-2017

<b>GENERAL COMMENTS</b>	<p>Overall comments:</p> <ol style="list-style-type: none"> <li>1. Please avoid using causal language such as "lowers", "protect" and "effect" in the entire manuscript, especially in the introduction and discussion when referring to observational studies.</li> <li>2. Given the key feature of this manuscript is weight specific analysis, p value for interactions should be presented in table 3a-b.</li> <li>3. It would be informative to conduct subsite analysis if sample size allows.</li> </ol> <p>Specific comments:</p> <p>Abstract: "conclusion" and "public health implication" -- Please write the conclusion specific to your study, ie focus on results specific to weight status. As is, both subsections over state the results given your study and the ones cited are observational studies.</p>
-------------------------	---

	<p>Introduction: Not clear where the authors got the CRC as 2nd highest cancer death and 49,700. In <a href="http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2016/cancer-facts-and-figures-2016.pdf">http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2016/cancer-facts-and-figures-2016.pdf</a>, CRC is the 3rd highest cancer death in both men and women. And 49190 deaths is ESTIMATED for 2016.</p> <p>Table 1: Need to add the diet scores and key foods and nutrients intake.</p> <p>Table 2: Is that multivariable adjusted? This is not key to your aims, if the results are not multivariable adjusted, presenting mean diet scores for each category would be sufficient.</p> <p>Tables 3a-b: Please also present age and energy adjusted results, this allows the readers to get a sense on the magnitude of confounding. Final multivariable models (in all tables) should be adjusted for energy intake and diet scores are usually correlated with energy intake.</p> <p>Tables 3a-b, 4a-b: The authors are fortunate to have a large sample size, it is unclear why they need to dichotomize the score. There is likely enough individuals to do a quartile if not quintile categorization. Please also run regression for a 5-point increase of HEI, and 1 point increase for DASH and the Mediterranean score.</p> <p>Tables 4a-b: The difference in risk is very small, no more than 0.6%. Given that this population all had AARP membership, they tended to be more affluent, therefore generalizability could also be limited. Would suggest eliminating these tables.</p> <p>Discussion: Your sample included individuals in their 50's, they are middle-aged, not older. Please discuss and provide some potential mechanisms on why the association was observed only in non-obese individuals. The last paragraph on page 18 is not central to your aim, would suggest removing it. In the last paragraph on page 19, please note that although your sample was drawn from the entire nation, it is not representative of adults in that age group because individuals of low SES are not included. This should be discussed as a limitation as well.</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer 1 stated that this was important study in addressing the potential influence of diet quality on the development of colorectal cancer and made constructive comments to improve data presentation and text improvements.

Reviewer 2 provided detailed methodological and analytic suggestions to improve the manuscript overall.

<b>I.</b>	<b>Response to Comments from Reviewer #1</b>		
	<b>Comment</b>	<b>Response</b>	<b>Location in Revised</b>

			Paper
1	Given the observational study design, it would be more appropriate to say “is associated” than “protects”.	We have replaced the word “protects” with “is associated” throughout the manuscript.	
2	In the Abstract, please add the total number of CRC cases by sex and BMI status and the mean follow-up time to the Results. Instead of stating that more research is needed, the conclusions could emphasize the differential relation of diet quality by BMI status.	<p>We agree with this point and have now added this sentence as the first sentence in the abstract: “Over a mean duration of 123 months follow-up, there were 6,515 new diagnoses of colorectal cancer (1,953 among the normal weight, 2,924 among the overweight, and 1,638 among the obese; 4,483 among men and 2,032 among women).”</p> <p>The Conclusion in the abstract now emphasizes the differential relation of diet quality by BMI status.</p> <p>“These findings illustrate the value of healthy eating habits among men who are of normal weight and provide evidence to inform new strategies for cancer prevention.”</p>	p.2
3	In the Public Health Implications, the reference to food environment is inappropriate as the current analysis did not examine any features of food environment, just individual dietary patterns. The last sentence about more research is needed can also be removed.	We have removed any reference to food environment and have deleted more research is needed (the last sentence in public health implications).	p.2
4	The Introduction is very lengthy and part of the presentation about previous research can be moved to the Discussion. The sentence about interventions is not directly relevant to this analysis.	<p>We shifted sections of previous research from the Introduction to the Discussion section. We removed the sentence about interventions.</p> <p>We moved the following sentence from Intro to the Limitations section (bottom of p.19)</p> <p>“This is important because despite steady improvements in healthy eating patterns among US adults the overall dietary quality remains poor particularly in low income</p>	p.18, p.16

		<p>populations.” – “despite steady improvements in healthy eating patterns among US adults the overall dietary quality remains poor particularly in low income populations.”</p> <p>We moved the following sentences from Intro to the Discussion section p.17</p> <p>“A recent narrative review of publications using the Nurses’ Health Study (1976-2016) identified red and processed meat, alcohol, smoking and obesity as factors that increase the risk of CRC. An ecological study suggested that 76% of the inter-country variation in colorectal cancer incidence was explained by meat, fish, and olive oil intake, with olive oil intake being associated with reduced risk.”</p>	
5	Please add to the Methods why underweight participants and those with a family history of CRC were excluded.	For this paper, we were interested in those who were at average risk because of the concern that those with family history, genetic factors might overwhelm nutritional factors. We have added this sentence to the limitations, “Our analytic dataset excluded those with family history of colorectal cancer and are therefore only generalize to those who are of average risk.”	p.19, first sentence in the Limitations section
6	The sentence about health status is not needed as that variable was not included in the models.	We have removed the sentence on health status on p.9.	p.9
7	The description of the Cox regression is unclear. Do the authors mean that length of follow-up was the underlying time metric and not age?	We agree with this comment and have changed “person-years” to “duration of observation” in the first sentence of the third paragraph of the Statistical Analysis section.	p.11
8	Was anything known about comorbidities, such as diabetes and hypertension among the study participants? If yes, did they affect CRC risk?	These data were not requested as part of the analytic dataset. Incorporating these data into a new analytic dataset would create a delay of several weeks. It is our opinion that the findings of the paper are still relevant without a consideration of comorbidities. However, at the discretion of the editors, we are willing to pursue this additional information. Meanwhile, we have added this sentence to the limitation section, “Medical co-morbidity was not included as a covariate in the multivariable models.”	p.19, second sentence in the Limitations section
9	The Results section needs to report the total number of CRC	We agree and have added the additional information to the results section.	p.11, pp.15-16 Tables

<p>cases by stage at diagnosis, mean follow-up time, overall hazard ratios for the total study population, and p-values for the interaction terms of BMI with dietary indices. It may also be useful to stratify the models by early vs. late stage at diagnosis.</p>	<p>For the first sentence of the second paragraph of the Results section we changed “During 10 years of follow-up” to “Over a mean follow-up duration of 123 months,” ...</p> <p>We added the following sentence as a second sentence into the above-mentioned paragraph:</p> <p>Over a mean duration of 123 months follow-up, there were 6,515 new diagnoses of colorectal cancer (1,953 among the normal weight, 2,924 among the overweight, and 1,638 among the obese; 4,483 among men and 2,032 among women). Of all new diagnoses, 9.7% were Stage 0; 38.4% were Stage 1; 14.0% were Stage 2; 22.7% were Stage 3; and 15.3% were Stage 4.</p> <p>We added this sentence as a final sentence in the above-mentioned paragraph:</p> <p>For the overall population, the hazard of incident colorectal cancer diagnosis was 33.3% less for women compared to men. Compared to those who had normal weight, the hazard of incident colorectal cancer diagnosis was 13.1% greater for those who were overweight and 30.6% greater for those who obese.p.11</p> <p><b><i>p-values for the interaction terms of BMI with dietary indices. It may also be useful to stratify the models by early vs. late stage at diagnosis.</i></b></p> <p>It is our opinion that p-values for the interaction would be a distraction to the main message of this paper which hinges on the lack of interaction, rather than the presence of an interaction. Instead, the main message of the paper is that there are similar gradients related to dietary adherence across weight strata. In other words, it’s the absence of an interaction that makes this paper interesting. As such, we found that all p-values for the interaction terms were not significant.</p> <p>However, we would like to note that reviewing this comment, along with a comment below from Reviewer 2, made the team realize that it is important to actually demonstrate the gradients that are associated with increasing dietary adherence within the weight strata. Therefore, Tables 4a and 4b have been revised based on quintiles of dietary adherence, rather than a dichotomous measures pp.15-16. In addition, we now</p>	<p>4a, 4b</p>
---	--	---------------

		<p>include p-values for trend across the gradients of dietary adherence within each weight strata. Details are provided below in our responses to the critiques of the second reviewer.</p> <p>That said, we wish to note that at the discretion of the editor, we would certainly be willing to add in the p-values for the interaction terms.</p> <p><b><i>It may also be useful to stratify the models by early vs. late stage at diagnosis.</i></b></p> <p>After careful consideration, we have chosen not to present this data; however, as for other situations, we would be willing to do so at the discretion of the editor. It is our opinion that the stage of diagnosis is not central to the main message of this paper. Furthermore, the paper presents a large amount of complex data and adding this additional analysis would further increase the complexity and density of the manuscript, without contributing substantively to the main message.</p>	
10	<p>In the interest of space, Table 2 could be deleted. Tables 3a and 3b as well as 4a and 4b could be combined.</p>	<p>It would be our preference to retain Table 2. We believe that it is important to ground the context of this manuscript with the predictors of dietary intake. Our main message is that dietary quality is related to cancer risk for each of our three weight strata, and including data about the predictors of dietary quality <i>from this data set</i> is important to complete the picture. However, at the discretion of the editor, we would certainly be willing to remove Table 2.</p> <p>We would also prefer to keep the tables for men (3a, 4a) and women (3b, 4b) separate. In accordance with standard practice in the field, we performed stratified analyses for men and women because predictors of colorectal cancer have been shown to differ by sex. Having separate tables emphasizes this distinction, while the parallel formatting of the tables (between Tables 3a and 3b and between Tables 4a and 4b) makes it easy to understand the direct correspondence. However, at the discretion of the editor, we would certainly be willing to combine these tables.</p>	
11	<p>In the Discussion, any reference to the food environment can be removed as this paper addresses eating behavior as assessed by dietary patterns based on the</p>	<p>We have removed any reference to the food environment in the abstract and in the Discussion section of the paper.</p>	

	assessment of individual dietary intakes.		
12	Hormonal factors may also be responsible for sex differences in CRC etiology. Previous research has reported association with hormone treatment.	We thank the reviewer for bringing this to our attention. We have added two studies on the effect of postmenopausal hormone therapy and CRC and added two sentences in the Discussion section.  "Studies of postmenopausal hormone therapy and colorectal cancer report a reduction in risk of colon cancer and a decrease in the risk of rectal cancer for postmenopausal women who had ever taken hormone therapy compared with women who never used hormones. The CRC risk reduction appears to be stronger for current and long term hormone users."	p.18
13	In the conclusions, it would help to disentangle whether the authors think that dietary quality is important as a risk factor for obesity, which increases CRC risk, or whether dietary quality may be an independent predictor of CRC risk.	We appreciate the reviewer's suggestion. Unfortunately, our data do not allow us to disentangle these two constructs.	
14	Minor comments: In the article summary, the number of participants was typed incorrectly.	We included the correct number of participants in the article summary, i.e., 398,458.	p.3
15	The reference style in the text is incorrect and the formatting of the bibliography is off in several places.	We have carefully reviewed the style in the text and formatted the bibliography. We use BMJ Endnotes.	

<b>Response to Comments from Reviewer #2</b>			
	<b>Comment</b>	<b>Response</b>	<b>Location in Revised Paper</b>
1	Overall comments: Please avoid using causal language such as "lowers", "protect" and "effect" in the entire manuscript, especially in the introduction and discussion when referring to observational	We have removed the language "lowers" "protect" and "effect" throughout the manuscript.	



	studies.		
2	Given the key feature of this manuscript is weight specific analysis, p value for interactions should be presented in table 3a-b.	Please see comment #9 in Response to Review Comments from Reviewer 1 above.	pp.14-15
3	It would be informative to conduct subsite analysis if sample size allows.	Given available data, this analysis is not possible.	
4	Abstract: "conclusion" and "public health implication" -- Please write the conclusion specific to your study, ie focus on results specific to weight status. As is, both subsections over state the results given your study and the ones cited are observational studies.	Conclusion: These findings illustrate the value of healthy eating habits among men who normal weight and provide evidence to inform new strategies for cancer prevention.  Public Health Implications: The findings accentuate the need to establish strategies to improve diet quality and prevent obesity as a cancer prevention strategy.	p.2  p.2
5	Introduction: Not clear where the authors got the CRC as 2nd highest cancer death and 49,700. In <a href="http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2016/cancer-facts-and-figures-2016.pdf">http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2016/cancer-facts-and-figures-2016.pdf</a> , CRC is the 3rd highest cancer death in both men and women. And 49190 deaths is ESTIMATED for 2016.	We appreciate the 2016 updated figures the reviewer provided with the associated reference. We have made the changes to the first sentence of the Intro. Colorectal cancer (CRC) is the third leading cause of cancer-related deaths in the United States, claiming over 49,190 lives in 2016.	p.4
6	Table 1: Need to add the diet scores and key foods and nutrients intake.	Mean dietary scores and standard deviations are added to Table 1. It is our opinion that adding in food components would create an over-burdened table without any additional benefit to the reader of this manuscript. However, at the discretion of the editor, this information could be added as an appendix.	p.10
7	Table 2: Is that multivariable adjusted? This is not key to your aims, if the results are not multivariable adjusted, presenting mean diet scores for each category would be sufficient.	The results in Table 2 are multivariable adjusted, and the title has been revised to reflect this.	P.12

8	<p>Tables 3a-b: Please also present age and energy adjusted results, this allows the readers to get a sense on the magnitude of confounding. Final multivariable models (in all tables) should be adjusted for energy intake and diet scores are usually correlated with energy intake.</p>	<p>Age is included as a confounder for each model presented in Tables 3a and 3b. p.13-14</p> <p>The scores for the Mediterranean Diet and the DASH diet are energy adjusted. Therefore, including energy adjustment in the models would not be appropriate. However, in the methods section, we did not indicate that the DASH score were energy adjusted.</p> <p>The scores for the HEIX are not energy adjusted. However, our calculation of these scores is based on a well-validated algorithm published by the National Cancer Institute. To obtain the necessary data to add energy adjustment in the models for HEIX would be very difficult at this time.</p> <p>It is instructive to note that in general, the results for the three dietary measures were very similar, suggesting that not much is lost because the HEIX scores are not energy adjusted. p.7</p> <p>We have added this sentence to the end of the second paragraph of the Determinants sub-section of the Methods Section: "Mediterranean Diet Scores were energy adjusted." p.7</p> <p>We have added this sentence to the third paragraph of the Determinants sub-section of the Methods Section: "HEIX scores were not energy adjusted." p.7</p> <p>We have added this sentence to the end of the fourth paragraph of the Determinants sub-section of the Methods Section: "DASH score were energy adjusted." p.7</p>	<p>p.14,</p> <p>p.7</p> <p>p.7</p> <p>p.7</p> <p>p.7</p> <p>p.7</p>
---	---	--	---

			p.7
9	<p>Tables 3a-b, 4a-b: The authors are fortunate to have a large sample size, it is unclear why they need to dichotomize the score. There is likely enough individuals to do a quartile if not quintile categorization. Please also run regression for a 5-point increase of HEI, and 1 point increase for DASH and the Mediterranean score.</p>	<p>We are very thankful for this comment because it caused our team to reflect on the main message of this manuscript. Upon reflection, we recognized that it is the gradient effect of each dietary measure within each weight strata that is most important. Dichotomizing the dietary measures as we did does not allow this gradient effect to be manifested. Nonetheless, the dichotomized approach is helpful to provide an easily digestible message that is consistent with more granular analyses.</p> <p>Therefore, we have now based Tables 4a and 4b on quintiles of dietary scores, as suggested by Reviewer 2, and the results very nicely illustrate the gradient effect. We have included p-values for trend based on linear polynomial contrast as implemented by Stata.</p> <p>As such, we have updated Table 4. Pp.15-16</p> <p>We have inserted this sentences after the third sentence in the final paragraph of the Results section: "For this second set of models, dietary scores were entered as quintiles." P.14</p> <p>We have added this sentence as the antepenultimate sentence in the final paragraph of the Results section: "Statistical significance was based on linear trend across dietary quintile." P.14</p> <p>To describe the resulting findings, we have inserted the following as a fifth paragraph in the Results Section:</p> <p>"Finally, based on the multivariable model Cox regression models, we predicted incidence of new colorectal cancer at 10 years. As shown in Table 4a, we found statistically significant linear trends, suggesting a gradient affect associating increasing adherence to high-quality dietary patterns with decreasing incidence of colorectal cancer at 10 years. Gradient effects were strongest for men who were of normal weight or overweight, and less strong for men who were obese. The findings were more mixed for Women (Table 4b). For both men and women, absolute predicted rates of colorectal cancer were consistently less than 2.5%."</p> <p>p.14</p>	<p>pp.15-16</p> <p>p.14</p> <p>p.14</p>

		<p>We have also added the following sentence as the next-to-last sentence in the first paragraph of the discussion: "We also found an important gradient effect linking improving dietary quality with lower incident colorectal cancer for men."p.16</p>	
--	--	---	--

10	Tables 4a-b: The difference in risk is very small, no more than 0.6%.	We disagree with this suggestion. Hazard ratios only provide estimates of relative effect. When the overall risk is low, the importance of relative effect can easily be overestimated if measures of absolute effect are not considered, as presented in Tables 4a and 4b. In addition, the proper interpretation of the hazard ratio is complex and not intuitive to those without significant statistical training. In contrast, the intuitive meaning of predicted probabilities as presented in Tables 4a and 4b is more easily grasped.	pp.15-16
11	Discussion: Your sample included individuals in their 50's, they are middle-aged, not older. Please discuss and provide some potential mechanisms on why the association as observed only in non-obese individuals.	We changed the language to reflect the age distribution of the sample population from older adults to "middle aged and older adults."  Please discuss and provide some potential	p.16

	<p>The last paragraph on page 18 is not central to your aim, would suggest removing it. In the last paragraph on page 19, please note that although your sample was drawn from the entire nation, it is not representative of adults in that age group because individuals of low SES are not included. This should be discussed as a limitation as well.</p>	<p>mechanisms on why the association as observed only in non-obese individuals. We added this paragraph to the manuscript "Diet quality was not associated with cancer risk among obese adults. It is plausible that the beneficial effects of a healthy diet are attenuated by the inflammatory, hormonal, and other metabolic changes induced by obesity that promote colorectal carcinogenesis.(41) For example, the gut microbiome that provides important metabolic capabilities, is responsive to alterations of diet (42), and has been shown in obese people to be different from, and less diverse than, those of the non-obese (43)." Pp.18-19</p> <p>We have removed the last paragraph (on food environment) and added in the last paragraph on page 19 and in the limitations section "Our study population was relatively homogenous with upper-to-middle class Americans in urban centers: non-whites comprised a relatively small proportion of our sample."p.19</p>	<p>pp.18-19</p> <p>P.19</p>
--	---	--	-----------------------------

**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Teresa Fung Simmons College, USA
<b>REVIEW RETURNED</b>	09-Mar-2017

<b>GENERAL COMMENTS</b>	<p>The authors have addressed many of my previous comments. However, I still have some significant concerns.</p> <p>Comments on response to reviewer comments (the numbers correspond to the table that the authors submitted that addressed reviewer 2 comments:</p> <p>2. I understand the authors' concern on the interpretation of the p values for interactions. Please include those p values and let readers make their own interpretations.</p> <p>4. The conclusion as written in the abstract is more appropriate for the public health implication section. The conclusion should be specific to the results your observed.</p> <p>5. It is critical to state that the number of deaths in 2016 is the ESTIMATED number death (see page 4 of the document I cited in my original comments). It is inaccurate to state that CRC is "claiming" that number of deaths.</p> <p>6. It is important to include consumption quantity of key food groups in table 1. This would greatly facilitates comparison between studies for readers. When different studies find different results, it is important to examine intake difference between studies.</p>
-------------------------	---

	<p>8. I am not certain that the authors have done energy adjustment the appropriate way. Multiplying the diet score by a "correction factor" derived from standard energy level/actual energy intake is not a standard energy adjustment method. None of the 3 studies the authors cited (references 13, 29, 31) used that method to adjust for energy intake. Since this is the NIH-AARP dataset, please use an energy adjustment method consistent with other papers that used these data. For a classic reference on energy adjustment, please see Nutritional Epidemiology by Walter Willett (3rd edition, Chapter 11).</p> <p>9. Regarding dicotomizing the diet scores, the high vs low adherence is not based on any meaningful clinical cutoffs, moreover the readers of this paper would be able to understand methodology (and results) that examines "dose response" (ie quintile analysis). Therefore, please run the analysis in table 3a-b in quintiles.</p> <p>Specific comments:  Abstract, conclusion -- I'd like to suggest something in this style as the conclusion "We observed a lower risk of colorectal cancer among ____ with higher adherence to ____."  P4, L12 -- "estimated" deaths.  P11, L34-41 -- Please state if these are multivariable adjusted results. Please also provide confidence interval to show variability of results.</p>
--	---

## VERSION 2 – AUTHOR RESPONSE

Dear Reviewers:

First, we'd like to thank you for the excellent and careful review of the manuscript. We really appreciate it. Please find out response to your review below.

Best regards.

2. I understand the authors' concern on the interpretation of the p values for interactions. Please include those p values and let readers make their own interpretations.

The requested p-values for the interactions are now included in Tables 4a and 4b and are referenced in the text. Please note that Tables 3a and 3b report results from models that are stratified by weight category, and, therefore, interaction coefficients were not included. A improved explanation of the modeling strategy has also been provided in the text.

4. The conclusion as written in the abstract is more appropriate for the public health implication section. The conclusion should be specific to the results your observed.

The conclusion for the abstract has been revised as requested.

5. It is critical to state that the number of deaths in 2016 is the ESTIMATED number death (see page 4 of the document I cited in my original comments). It is inaccurate to state that CRC is "claiming" that number of deaths.

The requested change has been made.

6. It is important to include consumption quantity of key food groups in table 1. This would greatly facilitates comparison between studies for readers. When different studies find different results, it is

important to examine intake difference between studies.

The requested results have been added to Table 1 and referenced in the text.

8. I am not certain that the authors have done energy adjustment the appropriate way. Multiplying the diet score by a "correction factor" derived from standard energy level/actual energy intake is not a standard energy adjustment method. None of the 3 studies the authors cited (references 13, 29, 31) used that method to adjust for energy intake. Since this is the NIH-AARP dataset, please use an energy adjustment method consistent with other papers that used these data. For a classic reference on energy adjustment, please see Nutritional Epidemiology by Walter Willett (3rd edition, Chapter 11).

We have revised our approach to be in compliance with this request. All multivariable models now contain explicit adjustment for energy intake. After making these changes, the main results have remained essentially unchanged. However, we have improved the language describing the results thought the abstract and the paper.

9. Regarding dichotomizing the diet scores, the high vs low adherence is not based on any meaningful clinical cutoffs, moreover the readers of this paper would be able to understand methodology (and results) that examines "dose response" (ie quintile analysis). Therefore, please run the analysis in table 3a-b in quintiles.

The requested changes have been made. All multivariable models with colorectal cancer as the outcome now enter dietary scores as quintiles.

Specific comments:

Abstract, conclusion -- I'd like to suggest something in this style as the conclusion "We observed a lower risk of colorectal cancer among \_\_\_\_ with higher adherence to \_\_\_\_\_."

The requested change has been made.

P4, L12 -- "estimated" deaths.

The requested change has been made.

P11, L34-41 -- Please state if these are multivariable adjusted results. Please also provide confidence interval to show variability of results.

The requested change has been made

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Teresa Fung Simmons College, USA
<b>REVIEW RETURNED</b>	11-May-2017

<b>GENERAL COMMENTS</b>	The authors have addressed my comments adequately. I have no further comments.
-------------------------	--