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

TITLE (PROVISIONAL)	Assessing the Impact of Colonoscopy Complications on use of	
	Colonoscopy among Primary Care Physicians and Other Connected	
	Physicians: An Observational Study of Older Americans	
AUTHORS	Keating, Nancy; O'Malley, Alistair; Onnela, Jukka Pekka; Landon,	
	Bruce	

VERSION 1 - REVIEW

REVIEWER	Stacey Fedewa American Cancer Society, United States
REVIEW RETURNED	22-Sep-2016

GENERAL COMMENTS	Keating and colleagues used a random sample of Medicare claims
	to examine whether primary care physicians' CRC screening rates
	(with colonoscopy) changed if one of their patients experienced a
	major complication during a colonoscopy. Authors hypothesized that
	screening rates for physicians' patients would decrease in a period
	following an experience with a patient having a serious adverse
	event. Authors also examined whether potential changes in
	screening patterns varied on patient age and whether a serious
	adverse event would alter patients' colonoscopy rates among
	primary care physicians in the same practice. The paper poses a
	unique and interesting question, with public health relevance as
	physician recommendation is one of the most important predictors of
	CRC screening in the United States.
	The authors correctly identify the major limitations of the study,
	including the lack of direct information on referral patterns and
	awareness of adverse events, and information regarding the use of
	FOBT. Additionally, authors assumed that the number of people who
	should be removed from the denominator (eg: people with CRC or
	had previous colonoscopy) was stable over time. It is worth noting
	that colonoscopy use has been increasing over time (by year) and
	authors could conduct a sensitivity analyses adjusting for year to
	help address this point. I believe only seasonal, but not annual,
	variations were accounted for in the model.
	People may receive a colonoscopy for diagnostic reasons, are
	authors hypothesizing that they expect all colonoscopies to
	decrease following an adverse event or just screening
	colonoscopies? I would presume the latter. If so, a recent algorithm
	had been developed to predict screening versus non-screening
	colonoscopies in claims data (see: PMC4537082).
	I was a little confused in the analyses section of the methods. Was
	an interaction term used to measure the differences in differences

(DID) or was this captured in the indicator variable with 4 categories? If so, could authors clarify the use of this variable for DID and what these 4 categories were and how they were constructed.
Was adverse event occurrence in-line with previous estimates? On page 13, lines 42-47: could it also be that older patients receive more colonoscopies related to symptoms?
Minor comment-there's an extra period on page 13 line 42

REVIEWER	Seung-Hwa Lee
	Seo-Hae Hospital
	South Korea
REVIEW RETURNED	21-Oct-2016

GENERAL COMMENTS	The report describes the assessment of the Impact of Colonoscopy Complications on use of Colonoscopy . The quality of manuscript is very good. This manuscript is so good that it deserves publication. It is useful to many gastroenterologists in their practice. In order to improve the quality of this manuscript, I recommend some revisions. 1. You would like to add additional reference in introduction section as follows. 1) Lee SH, Park YK, Lee DJ, Kim KM, Colonoscopy procedural skills and training for new bagingers. World L Controenterol2014
	 December 7; 20(45): 16984-16995. 2. For ethical reason, you would like to add the IRB (institutional review board) number of this study in method section. 3. The manuscript may be published after correcting grammatical and typing errors. I truly enjoyed reviewing this manuscript. I think that the article will be widely cited many investigators in the gastroenterology. Lastly, I appreciate the opportunity of review of your manuscript.

REVIEWER	Li Yang Department of Gastroenterology and Hepatology, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China
REVIEW RETURNED	17-Nov-2016

GENERAL COMMENTS	The study has several limits, first of all the decision making process of physicians varied, nevertheless the findings could give rise to understanding of real word practices. Large data from the U.S.
	Medicare program and construction of Poisson regression could be convincing.

REVIEWER	Karen Barclay The University of Melbourne Australia
REVIEW RETURNED	25-Dec-2016

GENERAL COMMENTS	This is a well-written paper from an Institute of Excellence with readability and interest. I congratulate the authors on their efforts. The paper uses large data sets to extrapolate the patterns of colonoscopy requesting of primary care physicians after a major coded complication of a patient following a colonoscopy. The main finding is of a statistical reduction in colonoscopies requested in the second quarter after an event. There is little information on this topic in the literature so the paper is additive to current knowledge. The design is robust and the analysis complete. The paper has few grammatical errors that detract from its message. Strengths include the quality and design and addition to the known information in the literature. Weaknesses are the use of assumption and inference, exclusions, statistical v clinical relevance and the ability to appeal to the reader.
	Grammar and Tables/Figures:
	P13 line 15 ??completed?? Use of physician's, physicians and physicians' needs to be reviewed throughout the paper
	P 15 line 33 ??? ensure rather than assure
	Although there is only one, the use of Figure without a number 1 is odd to read
	General comments:
	The racial constitution of the study demographics does not reflect general US demographic data (eg 85% white) – what are your comments on the general applicability of the results as the literature suggests ethnicity is important in determining health decisions.
	Age-group sub-analysis in the more elderly showed reductions in three of four quarters – what is your impression about this? How does this relate to the comment about why only Quarter 2 was affected in the overall analysis?
	The exclusions are not addressed as to their possible effect.
	Although there is a statistically significant effect, can you comment on whether this is actually clinically relevant in terms of the small numbers of requests for a colonoscopy by each physician in each quarter?
	The use of decision making tools has been shown of limited efficacy in the literature – do you have any novel or implementation strategies? How do you know decision-making tools were not used?
	If there is considered to be a relevant clinical effect, can you offer

suggestions as to what clinicians could do with respect to their own behaviour modification? This is a separate issue to the use of decision-making tools.
Did the multiple complication physicians have further reduced requesting v one patient affected? It would be logical to think that more complications led to a bigger effect – was this assessed?
The message of your paper may be better received by your target audience by:
 Acknowledging the complexity of decision making, even given the use of decision making tools and guideline adherence of which I am supportive Acknowledging those for whom the memory of serious adverse events is relevant for longer periods and offering them insight Addressing the inexperienced clinician who may take at face value the short memory impact suggested and minimise the implications of longer impact of adverse clinical outcomes Reducing the strength of some of your statements eg the mention of reflex requesting behaviour, clinicians having cognitive bias and inaccurate interpretation of information
The conclusions need to reflect the findings of the paper. The results demonstrate statistical differences in requesting of a small proportion of physicians over a period of time which may be reflective of the influence of an adverse clinical outcome on physician behaviour.
Thank you for this interesting and informative paper.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer N ame: Stacey Fedewa Institution a nd Countr y: American

Cancer S ciety, Unit ed States P lease state any comp eting

interests: None

Keating and colleagues used a random sam ple of Medi care claims to examin e whether p rimary are physicians' CRC screening rates (with c olonoscop y) changed if one of t heir patients xperience d a major c omplication during a colonoscopy. Authors hypothesized that scr eening rates for physicians' patients wou ld decrease in a period following an experie nce with a patient having a serious adver se event. A uthors als o examined whether potential ch anges in screening patterns va ried on patient age an d whether a serious ad verse even t would alter patients' olonoscop y rates am ong primar y care phys icians in the same pr ctice. The paper pos es a unique and interesting question, with public health rele vance as p hysician recommendation is one of the ost import ant predictors of CRC screening in the Unit ed States.

We apprecia te these co mments.

The authors correctly i dentify the major limitations of th e study, including the lack of direct information on referral patterns a nd awarene ss of adve rse events, and inform ation regar ding the use of F OBT. Additionally, authors assumed that th e number of people who should be removed fr m the den ominator (eg: people w ith CRC or had previo us colono scopy) was stable over time. It is worth noting that colonoscopy use has b een increasing over t ime (by yea r) and

authors could conduct a sensitivity analyses adjusting for year to help address this point. I believe only seasonal, but not annual, variations were accounted for in the model.

As described in the "Analyses" section (page 10-11), we included indicator variables for study month to address effects of time as well as potential seasonal differences in colonoscopy use. In the revised manuscript, we clarified that we included indicators for each of the 60 study months and that this would adjust for differences over time (page 11, first paragraph): "The models included fixed effects for each physician as well as indicator variables *for each of the 60* study months (which adjusts for *differences over time* and/or seasonal differences in colonoscopy use)..."

People may receive a colonoscopy for diagnostic reasons, are authors hypothesizing that they expect all colonoscopies to decrease following an adverse event or just screening colonoscopies? I would presume the latter. If so, a recent algorithm had been developed to predict screening versus non-screening colonoscopies in claims data (see: PMC4537082).

The reviewer raises an interesting question. While we might expect to see a greater decrease in screening colonoscopies following an adverse colonoscopy event because these may be less necessary, we might also see a decline in diagnostic colonoscopies, which have higher baseline rates of adverse events. We appreciate the Adams et al reference about the algorithm to identify screening vs. diagnostic colonoscopies. We find this promising, but it has been developed and tested using data from a case control study that included just 493 enrollees in one of four private health plans that were part of the Cancer Research Network. The study showed the algorithm to have excellent classification accuracy in internal validation. However, the authors recommended external validation using data from other sources, and we find no evidence that has been studied in other populations such as the Medicare population we study. We have added text to the limitations (page 17, first paragraph): "Finally, we did not attempt to distinguish between screening and diagnostic colonoscopies. While we might expect to see a greater decrease in screening colonoscopies following an adverse colonoscopy event because these may be less necessary, we might also see a decline in diagnostic colonoscopies, which have higher baseline rates of adverse events. A new algorithm for identifying screening colonoscopies using claims data²¹ may allow for such distinctions once externally validated."

I was a little confused in the analyses section of the methods. Was an interaction term used to measure the differences in differences (DID) or was this captured in the indicator variable with 4 categories? If so, could authors clarify the use of this variable for DID and what these 4 categories were and how they were constructed.

The statistical model is a longitudinal model in which the key predictors are the time-varying indicators of whether a patient in the physician's cohort experienced and adverse event. We include physician indicators (as fixed effects) and month indicators (as fixed effects). The time-varying adverse event indicators can be thought of as interactions between physician type (ever adverse event versus never adverse event) and time (in proximal period versus not) and so in this way the model might be viewed as an extension of the standard difference -in-difference model. Crucially, each physician is serving as their own control in the same way that the units in a prepost difference-in-differences analysis serve as their own control. However, because of the confusion the specific reference to a difference-in-differences design might introduce, we have revised the text to simply refer to our design as a longitudinal design in which the exposures of interest (occurrence of an adverse event k time periods ago) are time-varying predictors. Please see page 10, last paragraph through page 11, first paragraph.

Was adverse event occurrence in-line with previous estimates?

A previous study in the Medicare population showed that rates of perforation within 30 days of colonoscopy was 6 of 10,000 patients and rates of gastrointestinal bleeding or transfusion within 30 days of colonoscopy was 64 per 10,000 patients. Because we wanted to be sure to identify adverse events that primary care physicians

would be aware of, we focused on perforations or gastrointestinal bleeding that resulted in hospital admission or death within 14 days, and identified a rate of 10 per 10,000, which we believe is consistent with these rates. We have added text to clarify that the rates were similar to prior studies (page 16, last paragraph): "...the relatively few serious adverse events observed, despite being consistent with prior studies,⁸..."

On page 13, lines 42-47: could it also be that older patients receive more colonoscopies related to symptoms?

We agree that older patients may be relatively more likely to receive diagnostic vs. screening colonoscopy than younger patients. We clarified that the lower number of colonoscopies may be related to fewer screening colonoscopies (page 15, paragraph 1): "Nevertheless, fewer colonoscopies were performed overall among the older versus younger patients, which may reflect physicians' appreciation of the lower benefit *of screening colonoscopy* in this group."

Minor comment-there's an extra period on page 13, line 42.

Thank you, we have corrected this error.

<u>Reviewer: 2</u> Reviewer Name: Seung-Hwa Lee

Institution and Country: Seo-Hae Hospital, South Korea Please state any competing interests: None declared

The report describes the assessment of the Impact of Colonoscopy Complications on use of Colonoscopy . The quality of manuscript is very good. This manuscript is so good that it deserves publication. It is useful to many gastroenterologists in their practice. In order to improve the quality of this manuscript, I recommend some revisions.

We appreciate the reviewer's positive comments about our manuscript.

1. You would like to add additional reference in introduction section as follows.

1) Lee SH, Park YK, Lee DJ, Kim KM, Colonoscopy procedural skills and training for new beginners. World J Gastroenterol2014 December 7; 20(45): 16984-16995.

We appreciate this recommendation. This paper provides recommendations on how to perform colonoscopy for those who may be new to performing this procedure. Because we do not believe that this is directly relevant to our manuscript, which focuses on the potential for psychological biases to influence future care decisions, we have not yet added it to the revised manuscript. We would be willing to work with the Editor to find a suitable location for the reference if the Editor requests.

2. For ethical reason, you would like to add the IRB (institutional review board) number of this study in method section.

We have included the study number for this approved protocol in the revised manuscript on page 10 (patient involvement section).

3. The manuscript may be published after correcting grammatical and typing errors.

We regret the typographical errors and have corrected them.

I truly enjoyed reviewing this manuscript. I think that the article will be widely cited many investigators in the gastroenterology. Lastly, I appreciate the opportunity of review of your manuscript.

We appreciate the reviewer's positive comments.

Reviewer: 3

Reviewer Name: Li Yang

Institution and Country: Department of Gastroenterology and Hepatology, West China Hospital, Sichuan University, Chengdu, Sichuan, People's Republic of China Please state any competing interests: None

The study has several limits, first of all the decision making process of physicians varied, nevertheless the findings could give rise to understanding of real word practices. Large data from the U.S. Medicare program and construction of Poisson regression could be convincing.

We agree with the reviewer that despite the limitations, which we outline in the discussion section, that our use of a large, national dataset that allowed us to study a large representative cohort of patients and their physicians and our rigorous difference-in-differences study design allow us to provide some empirical evidence for the influence of notable adverse events on care.

Reviewer: 4

Reviewer Name: Karen Barclay Institution and Country: The University of Melbourne, Australia Please state any competing interests: None declared

This is a well -written paper from an Institute of Excellence with readability and interest. I congratulate the authors on their efforts. The paper uses large data sets to extrapolate the patterns of colonoscopy requesting of primary care physicians after a major coded complication of a patient following a colonoscopy. The main finding is of a statistical reduction in colonoscopies requested in the second quarter after an event. There is little information on this topic in the literature so the paper is additive to current knowledge. The design is robust and the analysis complete. The paper has few grammatical errors that detract from its message. Strengths include the quality and design and addition to the known information in the literature. Weaknesses are the use of assumption and inference, exclusions, statistical v clinical relevance and the ability to appeal to the reader. We appreciate the reviewer's summary and positive comments about the paper. We address the reviewer's comments about weaknesses in the responses below.

Grammar and Tables/Figures:

P13 line 15 ??completed??

Use of physician's, physicians and physicians' needs to be reviewed throughout the paper P 15 line 33 ??? ensure rather than assure

Although there is only one, the use of Figure without a number 1 is odd to read

We have corrected the spelling of "completed" on page 14 (paragraph 3). We have reviewed our use of "physician's", "physicians" and "physicians". We used "physician's" when referring to the patients of a single physician, and in the revised manuscript, we corrected one instance where we used the pleural possessive mistakenly and another where we maintained the plural, but were referring to "physicians' practices",

clarifying that "practice" should have been "practices". We have also changed "assure" to "ensure" on page 17 (last paragraph).

Because we only had one figure, we used "Figure" instead of "Figure 1", which is a convention for many journals. Based on the reviewer's suggestion, we now refer to "Figure 1" (page 13) and will defer to the Editor's wishes on this issue.

General comments:

The racial constitution of the study demographics does not reflect general US demographic data (eg 85% white) – what are your comments on the general applicability of the results as the literature suggests ethnicity is important in determining health decisions.

The racial constitution of the study population reflects that of the US population over the age of 65, which has much less racial/ethnic diversity than the US population under the age of 65.

Age-group sub-analysis in the more elderly showed reductions in three of four quarters – what is your impression about this? How does this relate to the comment about why only Quarter 2 was affected in the overall analysis?

Because the test of the overall interaction of age by quarter was not statistically significant (P=0.15, as reported in Table 2 and in the text on page 13, last paragraph), in order to avoid inflating type I errors we do not draw conclusions about sub-group effects, including the observation that the differences in quarters 1 and 3 were statistically significant for older patients and not younger patients. That is, we comply with our pre-specified study plan of looking at subgroup effects if the omnibus test of the interaction effect is statistically significant, which controls for type I error in the face of multiple testing as we are at no risk of incurring a type I error whenever the omnibus interaction test is not significant.

The exclusions are not addressed as to their possible effect.

We excluded physicians with fewer than 25 Medicare patients assigned to them in any month, and the patients of these physicians. We have added text to the limitations to clarify that our conclusions do not generalize to such very-low volume physicians (page 16, paragraph 2): "We also studied only physicians caring for at least 25 Medicare beneficiaries, thus our findings may not generalize to very-low-volume physicians."

Although there is a statistically significant effect, can you comment on whether this is actually clinically relevant in terms of the small numbers of requests for a colonoscopy by each physician in each quarter?

We agree with the reviewer that the effect is small, and were careful to describe it as such throughout the manuscript. Nevertheless, as colonoscopy is such a frequently used procedure, even small effects may be clinically meaningful. As we highlight in the first paragraph of the discussion (page 14): "The negative impact is relatively modest for this clinical condition, wherein screening generally is supported by strong evidence; effects could be larger for other clinical conditions."

The use of decision making tools has been shown of limited efficacy in the literature – do you have any novel or implementation strategies? How do you know decision-making tools were not used?

As we comment in the manuscript, the Cochrane Database Systematic Review by Stacey et al suggests that "decision aids can help with making such information easily accessible to patients and their physicians¹⁸" (page 15, second paragraph). In that review, only 8 of the 115 studies of decisions aids for screening decisions addressed colorectal cancer screening, consistent with the relatively few decision tools available. Nevertheless, we cannot be certain that decision tools were not used. We have added text to the discussion (page 16, paragraph 2): "Second, our evidence is indirect; we had no information about the physician's decision-making process *(including the possible use of decision aids), …*"

If there is considered to be a relevant clinical effect, can you offer suggestions as to what clinicians could do with respect to their own behaviour modification? This is a separate issue to the use of decision-making tools.

We believe that a good grasp of the actual benefits and harms and their expected probabilities have the greatest potential to help physicians avoid cognitive biases. As we state in the discussion (page 16, first paragraph): "If physicians and patients routinely discuss or review the benefits and harms of tests, procedures, and treatments, then the associated probabilities and their expected implications will remain familiar to them."

Did the multiple complication physicians have further reduced requesting v. one patient affected? It would be logical to think that more complications led to a bigger effect – was this assessed?

We conducted additional analyses to assess if there was a stronger association of adverse events with fewer future colonoscopies among the small number of physicians who experienced more than one adverse event (overall 4864 of the 30,704 physicians had at least one patient with an adverse event, but only 951 of these physicians had more than one patient with an event). We found some evidence of a stronger effect among physicians with more than one adverse event (see Table below for the change in quarterly number of colonoscopies among physicians with more than one adverse event). Because the number of physicians with multiple adverse events is so small, our study is not sufficiently powered to draw conclusions from these findings for physicians with more than a single adverse event. Due to concern about limited power as well as the addition of other analyses to the revised manuscript, we have not added this analysis to the revised manuscript; however we are willing to do so if the Editor requests. Instead, we added text to the limitations (page 16, last paragraph): "Also, the relatively few serious adverse events observed, despite being consistent with prior studies,⁸ limited our power to assess for differences among physicians experiencing multiple adverse events."

	% Change (95% Cl)	P value for interaction
All physicians with an		
adverse event		
Quarter 1	-2.2	.007
Quarter 2	-4.0	<.001
Quarter 3	-3.5	<.001
Quarter 4	-2.9	.001
Physicians with >1		
adverse event		
Quarter 1	-6.5	.60
Quarter 2	-7.9	.43
Quarter 3	-5.3	.04
Quarter 4	-3.9	.01

Table. Change in quarterly number of colonoscopies among physicians' patients following an adverse colonoscopy event among a physician's patient for physicians with more than one adverse event

The message of your paper may be better received by your target audience by:

- Acknowledging the complexity of decision making, even given the use of decision making tools and guideline adherence of which I am supportive

- Acknowledging those for whom the memory of serious adverse events is relevant for longer periods and offering them insight

Please see response to next comment.

- Addressing the inexp erienced clinician who may take at face value the short memory impact sug gested and minimise the implications of lon ger impact of adverse clinical outcomes

W ith this an the prior c omment, we believe that the review er is questioning wheth er certain ty pes f physician s, for example, those with less experience, are more impacted than other physician s when a patient has a serious adverse event, such that an a dverse event in one of their patients might lead t a stronger or more las ting effect on colonoscopies for their other patients. We have c onducted an additional analysis to assess the possibility that inexperienced physic ians might be more likely t o be impact ed more by a patient's experience of a serious adverse event. We found that less exp erienced (y ounger) physicians wer more impacted than o lder physici ns. This

nalysis has been adde d to the intro duction (page 7, paragr aph 2), methods (page 11, paragraph), results (page 13, last paragraph and Table) and discuss ion (page 1, paragraph 1 and last aragraph).

- Reducing the strength of some f your statements eg the mentio n of reflex r equesting behaviour, clinicians having cogn itive bias and inaccurate interpre tation of in formation

In the revised manuscrip t, we have added a sta ement unde rscoring th complexity of decision making (page 15, parag raph 3). We have also attempted to moderate some of our text. For xample, we removed th e word "reflexively" (page 15, para graph 1). W e do not bel eve that the mention of clinicians hav ing cognitive biases is likely to be p oorly received, becaus e we believe that

there is sufficient evidence in both t e lay and th e medical literature to s upport this, which we have c ited, and so mething tha t we hope o ur paper will remind ph ysicians abo ut.

The conclu sions need to reflect t e findings of the pap r. The resu Its demon strate statis tical differences in requesting of a sm all proporti on of physicians over a period of time which may be refl ective of th e influence of an adverse clinical outcome on physicia n behaviou r. We believe that we have underscord the small effects. In our concluding paragraph, we state: "In c onclusion, a physician's experienc of a patient having a serious adverse event from colonosc opy was

ssociated w ith a small and tempor ry decline in rates of co lonoscopy among that physician's other atients that did not vary by the bas eline risk of the physicia n's patients based on age."

Thank you f or this interesting and informativ e paper.

We apprecia te the revie wer's helpful comments.

We apprecia te the opportunity to rev ise our man uscript, and we believe that the ma nuscript has been

s trengthene d by the changes. We h ope that you est to keep the manusc ript brief, d spite the a I ength is 3,341 words.

find it acceptable for p ublication. e hav ditions requ ested by reviewers; the current

e have don e our current manuscript

VERSION 2 – REVIEW

REVIEWER	Stacey Fedewa American Cancer Society, United States
REVIEW RETURNED	09-Feb-2017

GENERAL COMMENTS	The authors have done a nice job addressing my concerns. I thank
	them for writing this manuscript and the opportunity to review it.