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)	The health system costs of potentially inappropriate prescribing in Ontario, Canada: a protocol for a population-based cohort study
AUTHORS	Black, Cody; Thavorn, Kednapa; Coyle, Douglas; Smith, Glenys; Bjerre, Lise

VERSION 1 – REVIEW

REVIEWER	Tom Fahey HRB Centre for Primary Care Research & RCSI Medical School Dublin, Ireland
REVIEW RETURNED	02-Feb-2018

GENERAL COMMENTS	Reviewer's comments: Manuscript ID bmjopen-2018-021727, entitled "The health system costs of potentially inappropriate prescribing in Ontario: a population-based cohort study." Thank you for asking me to review this paper. It is a study protocol of a retrospective cohort study using Ontario's health administrative databases. The overall aim will be to assess the burden of potentially inappropriate prescribing (PIP) and incremental costs (medication, hospitalisation and emergency department visit costs) attributable to PIP overall and with individual PIP drugs classes. The protocol is clearly presented. Statistical analysis will be in relation to 1) assignment of time-to-PIP; 2) costs of PIP (medication use, ED visits and hospitalization). Subgroup analysis is proposed for PIP by age, and for STOPP and START criteria separately. I only have minor comments: I am not familiar with the coding of the Ontario system and whether morbidity codes are sufficient in terms of completeness and validity to cover all START criteria which relate to medication errors of omission, rather than medication errors of commission (START). The authors have previously published a protocol in BMJ Open (their reference #13).[1] Having looked at this published protocol, a large part covers the measurement of STOPP and START. It is really an editorial decision as to whether this submission that adds enough material about the economic evaluation of PIP to merit a separate protocol publication. Reference Bjerre L, Ramsay T, Cahir C et al. Assessing potentially inappropriate prescribing (PIP) and predicting patient outcomes in Ontario's older population: a population-based cohort study applying subsets of the STOPP/START and Beers' criteria in large health administrative databases. http://bmjopen.bmj.com/content/5/11/e010146

REVIEWER	Andreas D. Meid (supported by Carmen Ruff) Dr. sc. hum. Andreas Meid Carmen Ruff Heidelberg University Hospital Dept. of Clinical Pharmacology & Pharmacoepidemiology Im Neuenheimer Feld 410 69120 Heidelberg Germany Homepage: http://www.klinikum.uni-heidelberg.de/clinpharm
REVIEW RETURNED	08-Feb-2018

GENERAL COMMENTS	Short summary of the study In this study, the authors aim to analyze costs that are related to potentially inappropriate prescribing of drugs to patients > 65 years in the province of Ontario in Canada. For this purpose a retrospective cohort study in administrative health data will be conducted applying the STOPP/START criteria version 2 by O'Mahony et al. An overall economic burden associated with inappropriate prescribing (STOPP-criteria) will be calculated and a subset of STOPP/START criteria will be applied to investigate their economic burden. In general, the authors would like to show the potentially avoidable costs of inappropriate prescribing. General comment We support the idea of this study and acknowledge the need to fill the evidence gap of (economic) consequences of potentially inappropriate prescribing, e.g., as defined by the STOPP/START list. There are nevertheless some crucial issues to be resolved in order to ensure an appropriate and transparent study conduct, which should be the aim of publishing a study protocol. Abstract (Strengths and limitations of this study): 1. We would suggest the use of the term "older people" or "older person" rather than "elderly". There is a clear movement to refer to older people, see for example, http://www.bmj.com/content/334/7588/316 2. Methods and analysis: is "[…] to aid in prioritizing targets […]" more clear, if it is actually meant? 3. Please mind the spelling of "strengths" 4. The last point mentioned in the "Strengths and limitation" part is rather an outlook than a strength or a limitation. Please consider to revise this section. Introduction: Background 5. The abbreviation PIO is not exclusive; other abbreviations that might be more familiar exist, as well. Please consider using "PPO" (potential prescribing omissions) [1,2]; at least the references should be cited to provide a good background for the reader. References: [1] Hill-Taylor B, Walsh KA, Stewart S, Hayden J, Byrne S, Sketris IS. Effectiveness of the STOPP/START (Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert doctors to the Right Treatment) criteria: systematic review and meta-analysis of randomized controlled studies. J Clin Pharm Ther 2016;41:158-69 [2] Meid AD, Lampert A, Burnett A, Seidling HM, Haefeli WE.

	The impact of pharmaceutical care interventions for medication underuse in older people: a systematic review and meta-analysis. Br J Clin Pharmacol 2015;80:768-76 6. The authors claim that PIO/PPO is associated with more drugs prescribed. This is rather controversial (e.g., [3-5]) References: [3] Gorup EC, Šter MP. Number of medications or number of diseases: what influences underprescribing? Eur J Clin Pharmacol 2017;73:1673-1679 [4] Meid AD, Quinzler R, Groll A, Wild B, Saum KU, Schöttker B, Heider D, König HH, Brenner H, Haefeli WE. Longitudinal evaluation of medication underuse in older outpatients and its association with quality of life. Eur J Clin Pharmacol 2016;72:877-85 [5] Meid AD, Quinzler R, Freigofas J, Saum KU, Schöttker B, Holleczek B, Heider D, König HH, Brenner H, Haefeli WE. Medication Underuse in Aging Outpatients with Cardiovascular Disease: Prevalence, Determinants, and Outcomes in a Prospective Cohort Study. PLoS One 2015;10:e0136339 7. Please give a reference for your definition of polypharmacy or revise the last sentence of the first section. Polypharmacy may refer to the number of drugs and not to the clinical necessity of the drugs prescribed [6]. References: [6] Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatr 2017;17:230. Methods: Study design 8. Please change the word “form” into “from” (“A population-based, […] identical to that from previous publications form a larger retrospective, population-based cohort study […]”.) 9. Please make sure to be consistent in giving the dates of your study period and patient accrual period. In figure 1, there seem to be other dates given than in the methods part. Beyond, it could be worth specifying the reason for one year before the accrual period (e.g., “run-in period to obtain baseline information”) and one year after the accrual period (e.g., “follow-up for long-term outcomes”). 10. The authors strongly claim that a fixed observation window of 90 days is appropriate, because the PIP influence is not expected to go beyond this period. This is speculative and not supported by clinical evidence. We consider the 90-day-period as reasonable, but we do not know the time effect and such an assumption should never been introduced with “since”. Please replace “since it is […]” by “because we do not assume that the potential […]”. In the literature, a 90-day observation window is well known regarding patient’s outcomes such as admissions to hospitals [7,8], but it is also possible that an outcome of a PIP could be observed later than 90 days after the initiation of the PIP, e.g. adverse drug reactions which occur “delayed” (time-related) [9]. If you have only analyzed the following 90 days after a PIP was prescribed the study period could have been ended 90 days after the accrual period. Why did you then choose a study period including the accrual period plus another year? References: [7] Wang L, Porter B, Maynard C, Evans G, Bryson C, Sun H, Gupta

	I, Lowy E, McDonell M, Frisbee K, Nielson C, Kirkland F, Fihn SD. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. Med Care 2013;51:368-73. [8] Bernabeu-Mora R, García-Guillamón G, Valera-Novella E, Giménez-Giménez LM, Escolar-Reina P, Medina-Mirapeix F. Frailty is a predictive factor of readmission within 90 days of hospitalization for acute exacerbations of chronic obstructive pulmonary disease: a longitudinal study. Ther Adv Respir Dis 2017;11:383-392 [9] Edwards IR, Aronson JK: Adverse drug reactions: definitions, diagnosis, and management. Lancet 2000; 356: 1255-1259 11. The authors mention five ICES-derived cohorts. Please give more details: What does case ascertainment actually mean, how will it be achieved, how did you identify the respective populations, which codes (e.g. ICD-10 codes, …) were used for the respective cohort? Is it possible that some patients occur in multiple cohorts (for example in the hypertension cohort AND in the congestive heart failure cohort) or do the cohorts only consist of patients from only one disease (for example patients in the hypertension cohort suffer only from hypertension)? Please indicate in the part describing Objective 2 for what kind of analysis did you use the additional derived cohorts. Objective 1 – Overall health system costs due to PIP 12. The authors mention an unpublished manuscript about the coding process. Wouldn't this be the actually informative part of such a study protocol? Not that such interesting information is left out of the submitted protocol, it is not available at all. (Supplementary) Tables could/should supply this useful information. 13. Please give first the written out word before you use an abbreviation. What does "HAD" mean? 14. Please mind to put blanks between two words ("clinicalevents", "daysspent" "covariateand"). 15. How did you identify clinical events that were linked to PIP? As we understood, you did not include every hospitalization or emergency department visit occurring after a PIP in your analysis. Under what kind of circumstances did you include them? Did you use any probability scales (e.g., Naranjo Adverse Drug Reaction Probability Scale for example)? How can you be sure that the outcome was related to the PIP? What kind of decision rules were applied and by whom (what kind of profession?)? 16. Please consider to clarify that the omission of a medication would only be "negative" (medication) costs at a first glance. These sign of the initially negative costs change at a later time, for instance after the deterioration of the patient's health state. 17. The index date assignment to the control patients appears to be a reasonable approach. However, there are alternatives to think of. Couldn't it be possible to rather "predict" an index date by exploiting the underlying distribution and individual covariates (from PIP users)? 18. The authors firmly state the "matching cannot be conducted". We have some doubts as long as this cannot be

	empirically proven. Well-performing software solutions for propensity-score matching exist that allow efficient usage also in situation with largely different group sizes. Such a sentence could be written more cautiously, e.g., "we do not consider matching as an option because …" 19. The authors mention confounding (or rather bias) arising from the situation if the entry date of control subjects would have been used as the index date. Does this refer to "immortal time bias"? Objective 2 – Incremental costs of specific PIP criteria 20. 2nd section of "Exposure": Please consider giving information on the codes that were used to identify the listed criteria and the outcomes/diagnoses instead of referring to the publications where they already had been applied (especially if they are not published yet). It is not possible to understand and assess this method completely by not knowing what exactly was done. 21. Table 1. Please consider to define thresholds regarding the estimation of PIP frequency and costs (e.g. costs < 1000 $CAD were classified as low costs). Please consider to be consistent by giving the definitions of the criteria (either way they were published by O'Mahony et al. or as a description of each criterion). 22. Please mind to put blanks between two words ("criteriondescribed"). Discussion 23. Please consider that studies modeling the impact of, for instance, medication underuse as defined by the START criteria exist [10]. This would emphasize the study's value in providing empirical evidence for this topic. References [10] Meid AD, Haefeli WE. Age-Dependent Impact of Medication Underuse and Strategies for Improvement. Gerontology 2016;62:491-9 Ethics and Dissemination Anticipated limitations 24. Please consider to revise the sentence about assessing the adherence in claims data (e.g. "It is difficult to establish adherence in claims data, several approaches are known from literature [Ref]. Each of them has their own limitations such as…").

VERSION 1 – AUTHOR RESPONSE

Table 1. Response to reviewer comments

	Comment	**Response and Action Taken**
	Editorial Requests	
1	In light of reviewer 1's comments, can you please work on providing a stronger justification for publishing this protocol separately from reference #13? There	This protocol presents the details of a study that cannot be found anywhere else. It contains additional details of the economic analysis component of the PIP STOPP study that did not be provided in reference 13. This

	seems to be strong overlap between the two papers.	study does share some similarities with the protocol in reference #13 in terms of data source used, cohort identification and exposure of interest, but the submitted protocol differs greatly in a number of important areas, including the description of the outcomes to be studied, the observation window for outcome ascertainment, as well as the statistical methodology to be used to assess the association between our exposure and outcomes of interest. These sections as described within our submitted protocol under review at BMJ Open. Publication of this protocol will highlight the need to assess the health system impact of potentially inappropriate prescribing. The research methods we have proposed may be adopted by fellow researchers in other jurisdictions thus increasing the potential positive impact of our research
2	Please revise the title to clarify where Ontario is and make it clear this is a protocol	Thank you for this suggestion, as it adds needed clarity to our study title. We have altered the title of the protocol form the submitted version to "The health system costs of potentially inappropriate prescribing in Ontario, Canada: a protocol for a population-based cohort study"
	Reviewer 1	
1	I am not familiar with the coding of the Ontario system and whether morbidity codes are sufficient in terms of completeness and validity to cover all START criteria which relate to medication errors of omission, rather than medication errors of commission (START).	Thank you for your review and helpful feedback. Our team has previously applied the STOPP/START criteria in health administrative data in Ontario, where we were able to apply 64% of STOPP criteria (potential errors of commission) versus 27% of START criteria (potential errors of omission). The process of identifying the criteria applicable to health administrative data focused on identifying criteria where ICD codes would be valid and sufficient to identify the diseases. We currently do not have lab data required for some criteria, and there are other criteria that cannot be coded due to various difficulties, including a lack of codes necessary to identify certain diseases that are part of various STOPP or START criteria. The full coding paper is to be published soon and is referenced as a manuscript in publication. We have included in the exposure section of our protocol additional information on the number of STOPP/START criteria applicable to health

		administrative data in Ontario. We have also altered the text at the end of the first exposure section paragraph to say that the manuscript "is in preparation" instead of "will soon be available".
2	The authors have previously published a protocol in BMJ Open (their reference #13). [1] Having looked at this published protocol, a large part covers the measurement of STOPP and START. It is really an editorial decision as to whether this submission that adds enough material about the economic evaluation of PIP to merit a separate protocol publication.	Please see our response to the first editorial request.
	Reviewer 2	
1	We would suggest the use of the term "older people" or "older person" rather than "elderly". There is a clear movement to refer to older people, see for example, http://www.bmj.com/content/334/7588/316	Thank you for your review and valuable comments. Thank you for bringing this to our attention. We have altered our abstract and introduction of our protocol where "elderly" appeared to "older person(s)".
2	Methods and analysis: is "[…] to aid in prioritizing targets […]" more clear, if it is actually meant?	There are multiple potential targets that PIP policy could address, and the research will allow for prioritization of such targets by their health system impact.
3	Please mind the spelling of "strengths"	Thank you for alerting us to this error. We have changed the heading of the Strengths and Limitations section to reflect the correct spelling of "Strengths"
4	The last point mentioned in the "Strengths and limitation" part is rather an outlook than a strength or a limitation. Please consider to revise this section	We would argue that the fact that this study has the potential to impact medication policy directly given the outcomes is a strength given that not all studies have the prospect of such a direct impact on policy. Part of the data steward's (Institute for Clinical Evaluative Sciences) mission statement is to generate trusted evidence that makes policy better. (www.ices.on.ca/About-ICES/Mission-vision-and-values) Therefore we do not feel this warrants further modification, but would be willing to discuss with editors.

5	The abbreviation PIO is not exclusive; other abbreviations that might be more familiar exist, as well. Please consider using “PPO” (potential prescribing omissions) [1,2]; at least the references should be cited to provide a good background for the reader. References: [1] Hill-Taylor B, Walsh KA, Stewart S, Hayden J, Byrne S, Sketris IS. Effectiveness of the STOPP/START (Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert doctors to the Right Treatment) criteria: systematic review and meta-analysis of randomized controlled studies. J Clin Pharm Ther 2016;41:158-69 [2] Meid AD, Lampert A, Burnett A, Seidling HM, Haefeli WE. The impact of pharmaceutical care interventions for medication underuse in older people: a systematic review and meta-analysis. Br J Clin Pharmacol 2015;80:768-76	Our team has chosen PIO to match the vernacular of PIP, such that the two terms match with regards to the “Potentially Inappropriate” terminology, and only differ between prescription in omission. Nonetheless, we recognize that PPO is the terminology used by the STOPP/START criterion and will include it as well. We have included PPO as a possible alternative in the first paragraph of the introduction.
6	The authors claim that PIO/PPO is associated with more drugs prescribed. This is rather controversial (e.g., [3-5]) References: [3] Gorup EC, Šter MP. Number of medications or number of diseases: what influences underprescribing? Eur J Clin Pharmacol 2017;73:1673-1679 [4] Meid AD, Quinzler R, Groll A, Wild B, Saum KU, Schöttker B, Heider D, König HH, Brenner H, Haefeli WE. Longitudinal evaluation of medication underuse in older outpatients and its association with quality of life. Eur J Clin Pharmacol 2016;72:877-85 [5] Meid AD, Quinzler R, Freigofas J, Saum KU, Schöttker B, Holleczek B, Heider D, König HH, Brenner H, Haefeli WE. Medication Underuse in Aging Outpatients with Cardiovascular Disease: Prevalence, Determinants, and Outcomes in a Prospective Cohort Study. PLoS One 2015;10:e0136339	Thank you for pointing this out and providing us with these references. The inclusion of PIO (PPO) in this statement is an oversight and we will correct this to make sure it specifies that more clearly that PIP is associated with medication overuse and polypharmacy, not PIO. We have added “PIP” to the last sentence in place of “its” in the first paragraph of the introduction so that it now reads “…and the likelihood of PIP increases…”
7	Please give a reference for your definition of polypharmacy or revise the last sentence of the first section. Polypharmacy may refer to the number	Members of our group have conducted polypharmacy research in the past and have preferred the definition of polypharmacy that incorporates the clinical necessity of the medications prescribed in the definition. The

	of drugs and not to the clinical necessity of the drugs prescribed [6]. References: [6] Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatr 2017;17:230.	systematic review the reviewer points to include such a definition, and the conclusion points to a need to move towards a definition that incorporates the appropriateness of medications and not just numerical cut-offs. We have respectfully decided to continue with such a definition and have included a reference to our preferred definition. (Fulton MM, Allen ER. Polypharmacy in the elderly: a literature review. J Am Acad Nurse Pract. 2005 Apr;17(4):123-32.)
8	Please change the word "form" into "from" ("A population-based, […] identical to that from previous publications form a larger retrospective, population-based cohort study […]".)	Thank you for bringing this typo to our attention. We have altered the text as suggested in the first sentence of the study design section.
9	Please make sure to be consistent in giving the dates of your study period and patient accrual period. In figure 1, there seem to be other dates given than in the methods part. Beyond, it could be worth specifying the reason for one year before the accrual period (e.g., "run-in period to obtain baseline information") and one year after the accrual period (e.g., "follow-up for long-term outcomes").	Thank you for bringing these discrepancies regarding dates to our attention. The correct study period end date is March 31 2015, and the correct accrual period end date is December 31 2014 to allow for a 90 day follow-up period for outcomes between the last possible PIP date and the study end period. Within the definition of observation periods section we have changed the accrual period end date to reflect the correct date of December 31 2014, as well as the one-year follow-up window from "one-year" to 90-day. The correct accrual end date was also added to the second paragraph of the exposure section for objective 1. The correct accrual period end date was also inserted into Figure 1.
10	The authors strongly claim that a fixed observation window of 90 days is appropriate, because the PIP influence is not expected to go beyond this period. This is speculative and not supported by clinical evidence. We consider the 90-day-period as reasonable, but we do not know the time effect and such an assumption should never been introduced with "since". Please replace "since it is […]" by "because we do not assume that the potential […]". In the literature, a 90-day observation window is well known regarding patient's outcomes such as admissions to	Thank you for the suggested wording change. This will allow us to more clearly describe this as an assumption and not a fact. As we have described in the previous comment, the 1-year follow-up period was an error, and the corrected accrual period allows for a maximum follow-up of 90 days from end of accrual to end of follow-up for outcome ascertainment. We have changed the last sentence in the definition of observation periods section to reflect the reviewer's suggestion.

	hospitals [7,8], but it is also possible that an outcome of a PIP could be observed later than 90 days after the initiation of the PIP, e.g. adverse drug reactions which occur “delayed” (time-related) [9]. If you have only analyzed the following 90 days after a PIP was prescribed the study period could have been ended 90 days after the accrual period. Why did you then choose a study period including the accrual period plus another year? References: [7] Wang L, Porter B, Maynard C, Evans G, Bryson C, Sun H, Gupta I, Lowy E, McDonell M, Frisbee K, Nielson C, Kirkland F, Fihn SD. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. Med Care 2013;51:368-73. [8] Bernabeu-Mora R, García-Guillamón G, Valera-Novella E, Giménez-Giménez LM, Escolar-Reina P, Medina-Mirapeix F. Frailty is a predictive factor of readmission within 90 days of hospitalization for acute exacerbations of chronic obstructive pulmonary disease: a longitudinal study. Ther Adv Respir Dis 2017;11:383-392 [9] Edwards IR, Aronson JK: Adverse drug reactions: definitions, diagnosis, and management. Lancet 2000; 356: 1255-1259	
11	The authors mention five ICES-derived cohorts. Please give more details: What does case ascertainment actually mean, how will it be achieved, how did you identify the respective populations, which codes (e.g. ICD-10 codes, …) were used for the respective cohort? Is it possible that some patients occur in multiple cohorts (for example in the hypertension cohort AND in the congestive heart failure cohort) or do the cohorts only consist of patients from only one disease (for example patients in the hypertension cohort suffer only from hypertension)? Please indicate in the part describing Objective 2 for what kind of analysis did you use the	The cohorts described are already created, maintained cohorts that use validated approaches to identify patients with the particular comorbidity that the cohort is for, and can be linked to other health administrative databases. It is possible for patients to be in multiple registries. We have no control over the creation of these cohorts and their validated definitions for identifying persons within Ontario with the particular comorbidity of interest. More information on these cohorts and the linked ICES databases can be found in the referred documents and these details the reviewer seeks were not included as they are beyond the scope of the protocol and can be found within existing literature. 1. Institute for Clinical Evaluative Sciences. Datasets available through Data & Analytic

	additional derived cohorts.	Services [Internet]. [cited 2018 Feb 2]. Available from: https://datadictionary.ices.on.ca/applications/datadictionary/Default.aspx?viewmode=DataServices 2. Institute for Clinical Evaluative Sciences. ICES Data Dictionary [Internet]. 2016. Available from: https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx 3. Bronskill S, Carter M, Costa A et al. Aging in Ontario : An ICES Chartbook of Health Service Use by Older Adults [Internet]. 2010. Available from: https://www.ices.on.ca/Publications/Atlases-and-Reports/2010/Aging-in-Ontario The details requested are out of scope, and as is indicated within the protocol, the ICES-derived cohorts were not used in the analysis of objective 2, and were only used as a means to identify patients with particular diseases to help with identification of our exposure, PIP, which was also conducted in objective 1.
12	The authors mention an unpublished manuscript about the coding process. Wouldn't this be the actually informative part of such a study protocol? Not that such interesting information is left out of the submitted protocol, it is not available at all. (Supplementary) Tables could/should supply this useful information.	The coding manuscript we refer to in our protocol will be published shortly and include all the details the reviewer seeks. Their inclusion here may preclude publication in another journal due to copyright issues.
13	Please give first the written out word before you use an abbreviation. What does "HAD" mean?	Thank you for bringing this to our attention. We have spelled out HAD (health administrative database) instead of using the acronym in the instances where it appeared.
14	Please mind to put blanks between two words ("clinicalevents", "daysspent" "covariateand").	This space issue is not present in the Microsoft Word version of our submitted manuscript. No changes were necessary on our end.
15	How did you identify clinical events that were linked to PIP? As we understood, you did not include every hospitalization or emergency department visit occurring after a PIP in your analysis. Under what kind of circumstances did you include them? Did you use any probability scales (e.g., Naranjo Adverse Drug Reaction Probability Scale for example)? How can you be sure that the	Only hospitalizations and emergency department visits occurring within 90 days after PIP will be included, as this is the defined outcome ascertainment period and any events beyond such a time cannot not be reasonably expected to be associated with PIP based on our team's clinical experts. No other restrictions or decisions rules will be placed.

	outcome was related to the PIP? What kind of decision rules were applied and by whom (what kind of profession?)?	
16	Please consider to clarify that the omission of a medication would only be "negative" (medication) costs at a first glance. These sign of the initially negative costs change at a later time, for instance after the deterioration of the patient's health state.	Thank you for this point. Our intention was to remain conservative with regards to statements on costs to the health system, and analysis of the impact of PIP, but we do accept that it is important to acknowledge the potential for increased downstream costs of medication omissions, and not just the reduced medication costs. We are concerned that this may cause more confusion, as it goes beyond rationale and enters into discussion territory and may distract from our proposed methods. We will make sure to add this note as a discussion point to the manuscript for the objective 1 results. However, we have changed the wording from "negative costs" to a "reduction in costs".
17	The index date assignment to the control patients appears to be a reasonable approach. However, there are alternatives to think of. Couldn't it be possible to rather "predict" an index date by exploiting the underlying distribution and individual covariates (from PIP users)?	While the approach suggested by the reviewer might be a viable alternative, the implementation would be overly intensive given the objective of our study. Our intended goal is to identify the impacts of PIP itself. Creating a predictive model to determine when a PIP might occur given a set of covariates would be an interesting endeavour, but beyond the scope of our study.
18	The authors firmly state the "matching cannot be conducted". We have some doubts as long as this cannot be empirically proven. Well-performing software solutions for propensity-score matching exist that allow efficient usage also in situation with largely different group sizes. Such a sentence could be written more cautiously, e.g., "we do not consider matching as an option because …"	Thank you for providing us with this comment and the suggested wording. We have given in-depth consideration to using matching as an approach, however decided not to proceed along this path because we were unable to suitably and efficiently match exposed to unexposed subjects in prior studies with a similar cohort. We recognize that our prior attempts in previous studies at matching with a similar cohort may not have been completely exhaustive of the available approaches in the literature, though they did incorporate the collective expertise of our experienced data analysts at ICES. Additionally, the approach suggested within our protocol suits our stated objectives without compromising the validity of our study. We have changed the sentence on matching in the second paragraph of the assignment of time-to-PIP for unexposed patients section to reflect the reviewer's suggestion of more cautious wording. It now reads: "We do not consider matching as an option…" instead of "Matching cannot be conducted…".

19	The authors mention confounding (or rather bias) arising from the situation if the entry date of control subjects would have been used as the index date. Does this refer to “immortal time bias”?	Based on the definition for immortal time bias, it does not appear as though it would be applicable. The bias we refer to would be due to the fact that the control group would be systematically younger than the exposed group, thus introducing further confounding by age that we are concerned would not be completely addressed by the inclusion of the age covariate within our models.
20	2nd section of “Exposure”: Please consider giving information on the codes that were used to identify the listed criteria and the outcomes/diagnoses instead of referring to the publications where they already had been applied (especially if they are not published yet). It is not possible to understand and assess this method completely by not knowing what exactly was done.	As we have mentioned in response to comment 12, these details will soon be available in an upcoming publication we have referred to within our protocol as a manuscript in preparation. While these details would be helpful, they are far too elaborate to be included in this protocol. The coding manuscript we refer to in our protocol will be published shortly and include all the details the reviewer seeks. Their inclusion here may preclude publication in another journal due to copyright issues.
21	Table 1. Please consider to define thresholds regarding the estimation of PIP frequency and costs (e.g. costs < 1000 $CAD were classified as low costs). Please consider to be consistent by giving the definitions of the criteria (either way they were published by O’Mahony et al. or as a description of each criterion).	There were no defined thresholds for the ranking of frequency and costs ascribed to the selected PIP for the second objective. The ranking of high, mid and low were approximated based on the relative position of each criteria on the scatterplot of figure 2. The statement regarding being consistent with providing definitions of the criteria is unclear. Within the text we have provided paraphrased definitions of the criteria for brevity and readability in a paragraph format, and we have also provided their full definition as they appear in the O’Mahony publication in table format.
22	Please mind to put blanks between two words (“criteriondescribed”).	This space issue is not present in the Microsoft Word version of our submitted manuscript.
23	Please consider that studies modeling the impact of, for instance, medication underuse as defined by the START criteria exist [10]. This would emphasize the study’s value in providing empirical evidence for this topic. References [10] Meid AD, Haefeli WE. Age-Dependent Impact of Medication Underuse and Strategies for Improvement. Gerontology 2016;62:491-	We thank the reviewer for bringing this recent publication to our attention, as this will help contextualize our results once we have conducted our study and prepare to publish them and describe them in the context of available literature. This literature will be will be consulted for the discussion of our results manuscripts

	9	
24	Please consider to revise the sentence about assessing the adherence in claims data (e.g. “It is difficult to establish adherence in claims data, several approaches are known from literature [Ref]. Each of them has their own limitations such as…”).	Thank you for this comment. Within ICES housed data our options are limited with regards to efficiently measuring adherence to medication and it is typically limited to comparing the date when an original prescription was scheduled to expire with the dispensation date of the renewal prescription, which has been described within our anticipated limitations section.

VERSION 2 – REVIEW

REVIEWER	Tom Fahey RCSI Medical School Dublin
REVIEW RETURNED	16-Mar-2018

GENERAL COMMENTS	The reviewers have responded to my two concerns from my initial review. I have not read their initial protocol, also published in BMJ Open. I feel it is an editorial decision whether or not this submission is sufficiently different to merit separate publication. All other aspects of their protocol look good to me.

REVIEWER	Andreas D. Meid (supported by Carmen Ruff) Heidelberg University Hospital Dept. of Clinical Pharmacology & Pharmacoepidemiology Im Neuenheimer Feld 410 69120 Heidelberg Germany
REVIEW RETURNED	21-Mar-2018

GENERAL COMMENTS	The authors have adequately answered the raised points and improved their manuscript. There are nevertheless some remaining issues concerning contents exclusively reserved for another paper in progress. This applies to how STOPP/START criteria are to be operationalized. We still hold the opinion that such information would make a study protocol richer. Thus, we agree with the other reviewer that this is an editorial decision as to whether this submission adds enough material.