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)	Prevalence of potentially inappropriate prescribing in a subpopulation of older European clinical trial participants: a cross-sectional study
AUTHORS	O Riordan, David; Aubert, Carole; Walsh, KA; Van Dorland, Anette; Rodondi, Nicolas; Du Puy, Robert; Poortvliet, Rosalinde K. E.; Gussekloo, Jacobijn; Sinnott, Carol; Byrne, Stephen; Galvin, Rose; Jukema, J. Wouter; Mooijaart, SP; Baumgartner, Christine; Mc Carthy, Vera; Walsh, Elaine; Collet, Tinh-Hai; Dekkers, Olaf; Blum, Manuel; Kearney, Patricia

VERSION 1 – REVIEW

REVIEWER	Gulistan Bahat Istanbul University, Istanbul Medical Scool, Turkia
REVIEW RETURNED	05-Sep-2017

GENERAL COMMENTS	Title: The article is on the prevalence and associated factors of PIPs with STOPP-START v2. The title is not totally in line with the content. Thus it should be modified, maybe "Potentially inappropriate prescribing among older European adults by a subset of STOPP START version 2 criteria" or other else. Strengths and limitations The authors stated that "This is the first study to estimate and compare the prevalence and type of PIP and PPOs using the STOPP/START V2 criteria in community-dwelling older adults enrolled to a clinical trial in three European countries." This is an over-forcing statement: designating the study as the first one that estimate and compare the prevalence and type of PIP and PPOs using the STOPP/START V2 criteria in community-dwelling older adults. The first study is the study noted by the authors as reference 19 which then followed by the reference 18. The main difference is this study is performed among a clinical trial patients. The clinical trial itself has no effect on PIPs or PPOs making this statement inappropriate regarding specifically the inclusion methodology of the TRUST data set. Another point is that the authors used only a subset of the criteria set which makes this statement –again-overforcing. The authors should give up to claim this study as the first one in view of these. Nevertheless, the study is important as there is very scarce data on STOPP/START v2 criteria set and this study makes an important contribution.

	The authors noted “Some countries may have specific guidelines for the optimal treatment of conditions, therefore these guidelines could differ from the recommendations in the STOPP/START criteria and could explain why some PIP and PPOs were identified in one population and not in other” as a strength/limitation. This is just an explanation which should be placed to Discussion. This should be omitted from this section. Introduction The authors wrote so much about STOPP START v1, both on its content and prevalence of PIPs detected by it. However, this is unnecessary as this report is not on version 1 but v2 of the criteria set. So, these should be omitted. The authors may consider to give some literature knowledge on the prevalence of PIPs detected by the version 2. Methods A subset of 51/80 PIP and 22/34 PPOs indicators from the STOPP/START V2 criteria were applied to the participants which is considerably lower than the originally proposed criteria. This is a major limitation and may explain the lower prevalence of PIP in this study. This should be integrated into the limitations section located at the beginning of the article. This point was touched in discussion section but should be more detailed. The authors used some drugs as the indicators of some co-morbidities. However, as an example a dopamine agonist may be used for restless leg syndrome or colchisine may be used for FMF. Thus, this should also be considered in the discussion and comments. Statistical Analysis Did the authors looked for multicollinearity between the independent variables? Polypharmacy and multimorbidity may have a close interaction. This should be clarified. RESULTS The authors stated that “overall prevalence of PIPs were similar while PPOs were lower in Irish group”. However, when considering the figures, the Irish participants also have lower PIP, about ½ of the other groups. Hence, it is better to state that Irish group had both lower PIP and PPO but this was only significant for PPO in the analysis. Why the % of females lower in Ireland and Switzerland??: This is an unexpected finding as there is female dominance among older adults, and also among patients with hypothyroidism. This should be explained. Is it sthg related to sampling? This point may suggest that this report may not represent community dwelling elderly. This point should be clarified and if so, should be included in limitation section DISCUSSION Why did the authors give subheadings? The editor in chief would decide whether this is appropriate. No need to again write the finding. The article is long enough and the reader already read the finding at the abstract and the Results section. The authors did not consider the disadvantage of the explicit criteria that they used in this study over the implicit evaluation of PIP esp. the CGA.

	This should be noted in the discussion and commented. MINOR POINTS Instead of comorbidity, multi-morbidity term would be better as authors try to indicate the higher number of comorbidities in a given patient.

REVIEWER	Marie Bradley Food and Drugs Administration, United States
REVIEW RETURNED	06-Sep-2017

GENERAL COMMENTS	This is a cross sectional study comparing the prevalence of potentially inappropriate prescribing and prescribing omissions as defined by the STOPP/START version 2 criteria among participants of a randomized controlled trial on thyroid hormone supplementation conducted in three European countries. This study provides an interesting insight into the variations in PIP and PPO across Ireland, the Netherlands and Switzerland and makes use of the recently updated version of the STOPP/START criteria. While few studies have applied the recently updated STOPP/START criteria there have been a plethora of studies using STOPP/START version one and other explicit prescribing criteria such as Beers to estimate PIP/PPO prevalence worldwide. Findings from these studies assure us that high levels of PIP and often PPO are common in most countries and as a result it would appear that repeating more of the same studies may not be as helpful as focusing on interventions to reduce the issues. My comments are as follows: 1. The study population is small (n=532) and only n=115 in Ireland compared to the larger size of previous population based studies. The population enrolled in a clinical trial are likely to be highly selected and may be less likely to be representative of the population in each of the countries. It is therefore difficult to see how the findings of this study could accurately reflect what is occurring at the population level and also to see how any inferences could be made about required interventions to address these findings at the population level 2. There have been plethora of studies investigating PIP and PPO using version one of the STOPP/START criteria across various countries in both Europe and worldwide. I would like more justification as to why it is important to use the updated version of the STOPP criteria given that only 15 additional criteria have been added for PIP. I would also like to know how many of these additional criteria were actually applied in this study as only 51/80 PIP indicators and 22/32 PPO indicators could be applied. For example it would appear that many of the newer criteria such as those in “section E Renal system” could not be applied to this dataset in the absence of GFR data. How will applying these additional criteria change what we already know about the prevalence of PIP and PPO and what we should do to help address these prescribing issues? My concerns are really about what impact this study might have above and beyond the many previous studies that have already firmly established the high prevalence of PIP/PPO in many countries. Is this just more of the same?

	3. It is somewhat concerning that medication use was self-reported (at two sites) rather than being ascertained from electronic medical and prescription records, as it was in many of the previous population based studies on PIP. While self-report can be useful to capture use of over the counter (OTC) medications, there are various limitations associated with this method and it is highly dependent on the patient disclosing information on all medications. For various reasons patients may choose to withhold this information. Granted those patients enrolled in clinical trials are usually more compliant with these types of activities but this is not guaranteed especially in an older population where memory or forgetfulness could be an issue. In Switzerland a list of medications was obtained from the participants GP but as the authors have stated in the discussion GPs are not gatekeepers to care in Switzerland and so they may not have comprehensive information on all medications being taken especially those initiated by a specialist. Also in Switzerland it is unlikely that the GP would have information on OTC medications and some of the STOPP drugs such as aspirin and NSAIDs are available OTC. 4. In previous studies of PIP, at least in the UK and Ireland, the criterion relating to the use of maxiumum dose PPIs for more than 8 weeks has been a major contributor to the higher prevalence of PIP seen. Given the nature of the data available in this study this criterion could not be applied and as such, especially in Ireland, this will have resulted in an under-estimation of PIP. Minor points The authors state there was a similar prevalence of PIP in the three regions in the results section of the abstract but then in the conclusion they say it varies considerably? Methods It seems like a lot of work to manually estimate these criteria for 500 patients. Was some sort of computerized algorithm used? That should be made more explicit. Results Please include P values in Table 1. The tables show the results from the regression analyses not the actual analyses, please correct. Discussion It would be nice to have the authors opinion on why they think PPI and PPO was highest in Switzerland? The authors cite a few prior studies that have used the STOPP/START version two to assess either PIP or PPO or both in Spain and Turkey. I have also identified a study in Albania by Hudhra et al, 2016 and one in Ethiopia Getachew H et al, 2016 and wondered why these were not also cited. The authors claim that the lower prevalence of benzodiazepine PIP in Ireland may have been related to more careful auditing by GPs could it also have been related to the very small sample size in this study Ireland n= 115? The sentence on line 454 on inclusion of older people in clinical trials seems somewhat out of place- can you justify or remove. That whole section on trials in multimorbidity seems somewhat irrelevant to the current study. The sentence on sensitivity analysis should be moved to the results and mentioned in the methods.

REVIEWER	Marcela Jirón University of Chile. Chile
REVIEW RETURNED	14-Sep-2017

GENERAL COMMENTS	This is an interesting topic. It has been limited described in a cross-national study and using the updated version of STOPP/START criteria. Understanding the difficulties to put in place this version of the criteria on a dataset, the results would be an interesting approach if authors are willing to modify the title and objectives of this manuscript. This study has important selection biases. SCH patients may have health problems such as more nervousness, anxiety, constipation, significantly increased concentrations of serum total cholesterol, LDL and triglyceride, among others. Further SCH patients are more likely to have some diagnoses included in the definition of PIM and PPO. The validity of prevalence estimates is limited due to authors included a subset of STOPP/START criteria. I assume this selection was made based on the lack of clinical and drug use information available. Approximately a 40% of STOPP/START were lost by methods. The small sample size and sampling are an important limitation for the generalizability, but also over the odds to have a PIP and PPOs. It is unclear to me how the sample size was estimated and how representative it is for each country. In my opinion each comorbidity or Charlson Index need to be shown in details and included in the multivariable analysis. Conclusions need to be wrote based on results and main findings.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gulistan Bahat
Institution and Country: Istanbul University, Istanbul Medical School, Turkia:
Title:
The article is on the prevalence and associated factors of PIPs with STOPP-START v2. The title is not totally in line with the content. Thus it should be modified, maybe "Potentially inappropriate prescribing among older European adults by a subset of STOPP START version 2 criteria" or other else.

Response: Thank you for this comment. We have now suggested the following title:
"Prevalence of potentially inappropriate prescribing in a sub-population of older European clinical trial participants"

Strengths and limitations:
The authors stated that "This is the first study to estimate and compare the prevalence and type of PIP and PPOs using the STOPP/START V2 criteria in community-dwelling older adults enrolled to a clinical trial in three European countries." This is an over-forcing statement: designating the study as the first one that estimate and compare the prevalence and type of PIP and PPOs using the STOPP/START V2 criteria in community-dwelling older adults.

The first study is the study noted by the authors as reference 19 which then followed by the reference 18. The main difference is this study is performed among a clinical trial patients. The clinical trial itself has no effect on PIPs or PPOs making this statement inappropriate regarding specifically the inclusion methodology of the TRUST data set.

Response: Thank you for this comment. We believe this is the first comparative study estimating the prevalence and type of PIP and PPOs in community-dwelling older adults across multiple European countries. The studies carried out by Blanco-Reina et al. and Bahat et al. examined the prevalence in one country e.g. Spain and Turkey respectively. We have revised this sentence and it now reads as follows:
"To the best of our knowledge, this is the first study to estimate and compare the prevalence and type of PIP and PPOs using a subset of the STOPP/START V2 criteria in community-dwelling older adults across three different European populations."

Comment: Another point is that the authors used only a subset of the criteria set which makes this statement –again- over forcing. The authors should give up to claim this study as the first one in view of these. Nevertheless, the study is important as there is very scarce data on STOPP/START v2 criteria set and this study makes an important contribution.

Response: Thank you for this comment.

Comment: The authors noted "Some countries may have specific guidelines for the optimal treatment of conditions, therefore these guidelines could differ from the recommendations in the STOPP/START criteria and could explain why some PIP and PPOs were identified in one population and not in other" as a strength/limitation. This is just an explanation which should be placed to Discussion. This should be omitted from this section.

Response: Thank you for this recommendation, we have removed the following sentence
"Finally, some countries may have specific guidelines for the optimal treatment of conditions. These guidelines could differ from the recommendations in the STOPP/START criteria and could explain why some PIP and PPOs were identified in one population and not in others." from the Strengths and Limitations section and placed in the Discussion.

Comment: Introduction
The authors wrote so much about STOPP START v1, both on its content and prevalence of PIPs detected by it. However, this is unnecessary as this report is not on version 1 but v2 of the criteria set. So, these should be omitted. The authors may consider to give some literature knowledge on the prevalence of PIPs detected by the version 2.

Response: Thank you for this suggestion. We have now shortened the Introduction regarding the STOPP/START V1 criteria. It now reads as follows:
"In recent years, the STOPP/START (Screening Tool of Older Persons Prescriptions, Screening Tool to Alert doctors to Right Treatment) criteria were developed and validated as an explicit measure of PIP and potential prescribing omissions (PPOs) for use in older adults (≥ 65 years) in European countries. All criteria are organised according to physiological systems for ease of use. In 2014, the STOPP/START criteria were revised and adapted to new evidence-based guidelines, STOPP/START version 2 (STOPP/START V2), comprising 80 STOPP and 34 START criteria. Several new STOPP categories created in V2 include antiplatelet/anticoagulant drugs, drugs affecting, or affected by, renal function and drugs that increase anticholinergic burden. New START categories include urogenital system drugs, analgesics and vaccines. A number of criteria from V1 were removed in V2 due to a lack of evidence from the published literature.

A number of studies have reported the prevalence of PIP/PPOs in large populations of older adults using subsets of the STOPP/START V1 criteria. Cahir et al. estimated the prevalence of PIP was 36% among adults ≥70 years in a primary care population in Ireland. In a similar study, Bradley et al. reported a prevalence of PIP of 29% among older adults in primary care in the UK. However, there is a lack of research exploring the prevalence of PIP and PPOs in community-dwelling older adults using the updated criteria. Blanco-Reina et al. reported a prevalence of PIP and PPOs of 40.4% and 21.8% respectively among older adults in Spain. ”

Comment: Methods
A subset of 51/80 PIP and 22/34 PPOs indicators from the STOPP/START V2 criteria were applied to the participants which is considerably lower than the originally proposed criteria. This is a major limitation and may explain the lower prevalence of PIP in this study. This should be integrated into the limitations section located at the beginning of the article. This point was touched in discussion section but should be more detailed.

Response: It was only possible to apply a subset of the STOPP/START version 2 criteria, as information on drug strength, dose and duration of prescriptions was not available in the TRUST dataset. Also, to maximise the validity of our results, we felt it was prudent to only apply the criteria that we were confident we had full information on. This lack of information proved difficult in trying to capture all the criteria.
Previous studies have also experienced similar issues in applying the full set of the criteria e.g. Galvin et al “Prevalence of potentially inappropriate prescribing and prescribing omissions in older Irish adults: findings from The Irish LongituDinal Study on Ageing study (TILDA)”. European Journal of Clinical Pharmacology 2014; 70: 599-606. The authors could only apply a subset of 26 out of 65 STOPP and 10 out of 22 START version 1 criteria due to a lack of prescribing information.

While Verdoorn et al “Majority of drug-related problems identified during medication review are not associated with STOPP/START criteria” European Journal of Clinical Pharmacology 2015; 71: 1255-1262. The authors applied 25 out of 65 STOPP and 18 out of 22 START version 1 criteria.
Therefore, to address your comment we have included the following sentence in the “Strengths and Limitations” section at the beginning of the article:
“It was only possible to apply a subset of the criteria to the database due to a lack of information on drug strength, dose and duration of prescriptions and this may explain the low prevalence of PIP and PPOs in the study.”

Comment: The authors used some drugs as the indicators of some co-morbidities. However, as an example a dopamine agonist may be used for restless leg syndrome or colchicine may be used for FMF. Thus, this should also be considered in the discussion and comments.

Response: Thank you for this suggestion. This reproducible method for ascertaining diagnoses has been used in previous studies e.g.
Galvin et al “Prevalence of potentially inappropriate prescribing and prescribing omissions in older Irish adults: findings from The Irish LongituDinal Study on Ageing study (TILDA)”. European Journal of Clinical Pharmacology 2014; 70: 599-606.
However, to address your comment we have now included the following sentence in the limitations section in the Discussion:
“Although prescription drugs were used as proxies to indicate diagnoses, the possibility that these drugs may have been used to treat other conditions cannot be excluded.”

Comment: Statistical Analysis
Did the authors looked for multicollinearity between the independent variables? Polypharmacy and multimorbidity may have a close interaction. This should be clarified.

Response: Yes, we looked for multicollinearity between the independent variables polypharmacy and comorbidity. We calculated the variance inflation factor (VIF) as it quantifies the severity of multicollinearity. A rule of thumb indicates that if VIF > 10 then multicollinearity is high.
The VIF for comorbidity and polypharmacy were 1.82 and 2.13 respectively.

The following sentence has been included in the methods section:
"Multicollinearity between the independent variables polypharmacy and comorbidity was assessed by calculating the variance inflation factor (VIF)."
The following sentence has been included in the results section:
"The VIF for comorbidity and polypharmacy were 1.82 and 2.13 respectively."

Comment: Results
The authors stated that "overall prevalence of PIPs were similar while PPOs were lower in Irish group". However, when considering the figures, the Irish participants also have lower PIP, about ½ of the other groups. Hence, it is better to state that Irish group had both lower PIP and PPO but this was only significant for PPO in the analysis.

Response: Thank you for this suggestion. The results in the abstract now reads as follows:
"The overall prevalence of PIP was lower in the Irish participants (8.7%) compared to the Swiss (16.7%) and Dutch (12.5%) participants (p=0.15) and was not statistically significant. The overall prevalence of PPOs was approximately one-quarter in the Swiss (25.3%) and Dutch (24%) participants and lower in the Irish (14%) participants (p=0.04) and the difference was statistically significant."

Comment: Why the % of females lower in Ireland and Switzerland??: This is an unexpected finding as there is female dominance among older adults, and also among patients with hypothyroidism. This should be explained. Is it sthg related to sampling? This point may suggest that this report may not represent community dwelling elderly. This point should be clarified and if so, should be included in limitation section

Response: Thank you for this suggestion. The following paragraph has been included in the limitations section.
"The TRUST trial concerned patients with subclinical hypothyroidism (SCH). It is possible that women with SCH were more likely than men to have been treated by doctors and therefore not eligible for the trial, as doctors tend to associate thyroid disease more with women. Also, SCH symptoms can overlap with post-menopausal symptoms that women report (i.e. tiredness, low mood etc.) therefore pushing doctors to treat this condition."

Comment: Discussion
Why did the authors give subheadings? The editor in chief would decide whether this is appropriate.

Response: Thank you for this comment. Having reviewed previous papers published in the BMJ Open that included subheadings in the Discussion, we decided to follow the same format. However, we are happy to remove them if the Editor in Chief decides this is appropriate.

Comment: No need to again write the finding. The article is long enough and the reader already read the finding at the abstract and the Results section.

Response: Thank you for this suggestion. We have now removed the short paragraph on the statement of principal findings.

Comment: The authors did not consider the disadvantage of the explicit criteria that they used in this study over the implicit evaluation of PIP esp. the CGA. This should be noted in the discussion and commented.

Response: Thank you for this suggestion. We have now included the following paragraph in the Strengths and Limitations section of the Discussion:
"A number of different approaches for optimising prescribing appropriateness have been published. For example, comprehensive geriatric assessment (CGA) is a time consuming and resource intensive strategy to deploy and is more commonly used for intervention rather than prevalence studies. Therefore, STOPP/START was considered the most appropriate and feasible tool for this study."

Comment: Minor points
Instead of comorbidity, multi-morbidity term would be better as authors try to indicate the higher number of comorbidities in a given patient.

Response: Thank you for this suggestion. We have now used the term multimorbidity instead of comorbidity in the manuscript.

Reviewer: 2
Reviewer Name: Marie Bradley
Institution and Country: Food and Drugs Administration, United States
This is a cross sectional study comparing the prevalence of potentially inappropriate prescribing and prescribing omissions as defined by the STOPP/START version 2 criteria among participants of a randomized controlled trial on thyroid hormone supplementation conducted in three European countries. This study provides an interesting insight into the variations in PIP and PPO across Ireland, the Netherlands and Switzerland and makes use of the recently updated version of the STOPP/START criteria. While few studies have applied the recently updated STOPP/START criteria there have been a plethora of studies using STOPP/START version one and other explicit prescribing criteria such as Beers to estimate PIP/PPO prevalence worldwide. Findings from these studies assure us that high levels of PIP and often PPO are common in most countries and as a result it would appear that repeating more of the same studies may not be as helpful as focusing on interventions to reduce the issues.
Thank you for these comments. We agree there are a plethora of published studies that have used explicit and implicit tools to measure prescribing appropriateness. However, this study has some novel aspects:
It is one of the first to use the updated STOPP/START criteria.
The comparisons across countries regarding the role of the GP, policies on prescribing medicines e.g. benzodiazepines, characteristics of participants and prescribing practices in different countries are also interesting.
There is a lack of international comparisons in this field, and despite the limitations of the sample we believe that this study will add to the research on the important issue about prescribing practices among older adults. The study findings could be tested in future larger scale studies by adopting policies on prescribing from one country in another.

Comment: 1. The study population is small (n=532) and only n=115 in Ireland compared to the larger size of previous population based studies. The population enrolled in a clinical trial are likely to be highly selected and may be less likely to be representative of the population in each of the countries. It is therefore difficult to see how the findings of this study could accurately reflect what is occurring at the population level and also to see how any inferences could be made about required interventions to address these findings at the population level.

Response: Thank you allowing us to explain this process further. We have now included the following paragraph in the Limitations section of the Discussion.

"It is acknowledged that the sample size (n=532) is relatively small, however, the aim was to estimate and compare the prevalence and type of PIP and PPOs in a sample of patients from three European countries. The study population was based on participants enrolled to a clinical trial and may be somewhat different from the general population. However, the main inclusion criteria for the TRUST trial are quite broad. Secondly, although the data is based on a population of patients with SCH, there is no evidence to suggest that this would influence their chance of having a PIP or PPO."

Comment: 2. There have been plethora of studies investigating PIP and PPO using version one of the STOPP/START criteria across various countries in both Europe and worldwide. I would like more justification as to why it is important to use the updated version of the STOPP criteria given that only 15 additional criteria have been added for PIP.
I would also like to know how many of these additional criteria were actually applied in this study as only 51/80 PIP indicators and 22/32 PPO indicators could be applied. For example it would appear that many of the newer criteria such as those in "section E Renal system" could not be applied to this dataset in the absence of GFR data. How will applying these additional criteria change what we already know about the prevalence of PIP and PPO and what we should do to help address these prescribing issues?

Response: We have included the following paragraph in the Methods section:
"There have been significant changes to the updated criteria. Firstly there are more criteria in V2 (80 STOPP and 34 START compared 65 STOPP and 22 START in V1). Secondly, new drug groups have been included in the updated criteria e.g. sulphonylureas with a long duration of action. Thirdly, a number of criteria from V1 were removed from V2 due to a lack of evidence from the published literature. The extra criteria included in V2 arose from new clinical trial information, new systematic information and expert panel suggestions for new criteria. This highlights the need to update and revise the criteria on a regular basis as some criteria can become outdated or obsolete. Also, new drugs have entered the market since the V1 criteria were validated in 2008."
The number of additional STOPP and START criteria applied in this study were 22 and 9 respectively. The following paragraph has been included in the Clinical and Policy implications section in the Discussion.
"Screening tools such as the STOPP/START criteria have proven to be very beneficial not only in identifying the prevalence of PIP/PPOs in studies but also in intervention studies to improve medication appropriateness and reduce the risk of ADRs in older people. The updated version with the additional criteria will help to identify a larger number of PIP and PPO instances and therefore has a greater potential to reduce ADRs and improve other relevant patient outcomes."

Comment: My concerns are really about what impact this study might have above and beyond the many previous studies that have already firmly established the high prevalence of PIP/PPO in many countries. Is this just more of the same

Response: Please see the response to your first point.

Comment: 3. It is somewhat concerning that medication use was self-reported (at two sites) rather than being ascertained from electronic medical and prescription records, as it was in many of the previous population based studies on PIP. While self-report can be useful to capture use of over the counter (OTC) medications, there are various limitations associated with this method and it is highly dependent on the patient disclosing information on all medications. For various reasons patients may choose to withhold this information. Granted those patients enrolled in clinical trials are usually more compliant with these types of activities but this is not guaranteed especially in an older population where memory or forgetfulness could be an issue.

In Switzerland a list of medications was obtained from the participants GP but as the authors have stated in the discussion GPs are not gatekeepers to care in Switzerland and so they may not have comprehensive information on all medications being taken especially those initiated by a specialist. Also in Switzerland it is unlikely that the GP would have information on OTC medications and some of the STOPP drugs such as aspirin and NSAIDs are available OTC.

Response: Thank you for this comment. We have now included the following paragraph in the Limitations section of the Discussion.
"Although different approaches to the collection of medication data were used in each country, the authors (and the TRUST consortium with regards to safety) believe that all methods are thorough enough to capture all medication. For example, studies have highlighted that self-report medications are most likely to be congruent with patient records as a measure of current medications."
Reference used: Caskie G, Willis S. Congruence of Self-Reported Medications With Pharmacy Prescription Records In Low-Income Older Adults. Gerontologist. 2004; 44(2): 176–185.

Comment: 4. In previous studies of PIP, at least in the UK and Ireland, the criterion relating to the use of maximum dose PPIs for more than 8 weeks has been a major contributor to the higher prevalence of PIP seen. Given the nature of the data available in this study this criterion could not be applied and as such, especially in Ireland, this will have resulted in an under-estimation of PIP.
Response: Thank you for your comment. We have acknowledged this a limitation in the study with the following sentence:
"It was only possible to apply a subset of the STOPP/START V2 criteria, as information required for some criteria (i.e. drug strength, dose and duration of prescriptions) was not available in the TRUST dataset. For example, the prescribing of PPIs at full therapeutic dose for more than eight weeks was not reported. This may have contributed to an underestimation of the real prevalence of PIP in the study."

Minor points
The authors state there was a similar prevalence of PIP in the three regions in the results section of the abstract but then in the conclusion they say it varies considerably?

Response: Thank you for highlighting this. The conclusion now reads as follows:
"This study has estimated and compared the prevalence and type of PIP and PPOs among this cohort of community dwelling older people. It demonstrated a significant difference in the prevalence of PPOs between the three populations. Further research is urgently needed into the impact of system level factors as this has important implications for patient safety, healthcare provision and economic costs."

Comment: Methods
It seems like a lot of work to manually estimate these criteria for 500 patients. Was some sort of computerized algorithm used? That should be made more explicit.

Response: Thank you for this comment. We have now revised the following paragraph in the methods section. It now reads as follows:
"It was agreed a priori by the authors that D.O.R. (research pharmacist) would manually apply the criteria to all the Irish, Swiss and Dutch data. For validation purposes, the criteria were applied independently by a second member of the research team. K.W. (research pharmacist) applied the criteria to a random 10% sample of the Irish and Dutch data. C.E.A. (research medical doctor) applied the criteria to a random 10% sample of the Swiss data. There are two studies (OPERAM, SENATOR) currently assessing the automatization of the STOPP/START criteria to identify PIP and PPOs in older people.

The results from both studies should inform on the best method of automatizing screening tools to identify PIP and PPOs in this group of people. Therefore, the method used in this study for assessing the STOPP/START criteria should be considered as valid."

Comment: Results
Please include P values in Table 1.
The tables show the results from the regression analyses not the actual analyses, please correct.

Response: Thank you first this suggestion we have now included p-values in Table 1.
We have now changed the title of Table 2. It now reads as follows:
"Results of the univariable and multivariable logistic regression analyses for the association between age, sex, multimorbidity and polypharmacy with potentially inappropriate prescribing (PIP)."
We have also changed the title of Table 3: It now reads as follows:
"Results of the univariable and multivariable logistic regression analyses for the association between age, sex, multimorbidity and polypharmacy with potential prescribing omissions (PPOs)".

Comment: Discussion
It would be nice to have the author's opinion on why they think PPI and PPO was highest in Switzerland?

Response: The following sentence is included in the Discussion.
"The screening process and identification of potential participants for this clinical trial differed between countries and may explain some of the differences in the prevalence of PIP and PPOs."
The following paragraph has been included in the Discussion to expand on the comment.
"As Swiss patients can visit medical specialists directly if necessary they may receive more non-essential medicines. The prescribing process is further complicated if patients attend several specialists. Also, if there is a lack of collaborative decision making between the patient's GP and medical specialists this could result in a higher prevalence of PIP/PPOs among Swiss participants."

Comment: The authors cite a few prior studies that have used the STOPP/START version two to assess either PIP or PPO or both in Spain and Turkey. I have also identified a study in Albania by Hudhra et al, 2016 and one in Ethiopia Getachew H et al, 2016 and wondered why these were not also cited.

Response: Thank you for highlighting this.
Getachew, H et al. "Inappropriate prescribing of antithrombotic therapy in Ethiopian elderly population using updated 2015 STOPP/START criteria: a cross sectional study." Clinical Interventions in Aging. 2016: 11, 819-827.
Hudhra et al. "Prevalence and factors associated with potentially inappropriate prescriptions among older patients at hospital discharge." Journal of Evaluation in Clinical Practice.2016: 22, 707-713.
We have now cited both references and included the following text in the Discussion:
"A study conducted among 319 older patients discharged from a hospital in Albania identified that 63% received at least one PIP. In another study carried out in Ethiopia, the prevalence of inappropriate prescribing of antithrombotic therapy among 156 hospitalised elderly patients was assessed. The prevalence of PIP and PPOs were 51.4% and 48.6% respectively."

Comment: The authors claim that the lower prevalence of benzodiazepine PIP in Ireland may have been related to more careful auditing by GPs could it also have been related to the very small sample size in this study Ireland n= 115?

Response: Thank you for this comment. We have now included the following sentence in the Discussion.

“This low prevalence could also be due to a difference in the sampling approach in Ireland or it may have occurred by chance.”

Comment: The sentence on line 454 on inclusion of older people in clinical trials seems somewhat out of place- can you justify or remove. That whole section on trials in multimorbidity seems somewhat irrelevant to the current study.

Response: Thank you for this suggestion. We have now removed the section on multimorbidity in older people from the Discussion.

Comment: The sentence on sensitivity analysis should be moved to the results and mentioned in the methods.

Response: The following sentence has been mentioned in the methods:
“Sensitivity analysis excluding criteria triggered by combination of more than one drug was also performed.”
The following sentence has now been moved to the results:
“Sensitivity analysis excluding criteria triggered by combination of more than one drug had no effect on the results.”

Reviewer: 3
Reviewer Name: Marcela Jirón
Institution and Country: University of Chile. Chile
This is an interesting topic. It has been limited described in a cross-national study and using the updated version of STOPP/START criteria. Understanding the difficulties to put in place this version of the criteria on a dataset, the results would be an interesting approach if authors are willing to modify the title and objectives of this manuscript.

Response: Thank you for this suggestion. We have now changed the title of the study to
“Prevalence of potentially inappropriate prescribing in a sub-population of older European clinical trial participants".
Thank you for this suggestion. We have modified the objectives. It now reads as follows.
“To estimate and compare the prevalence and type of potentially inappropriate prescribing (PIP) and potential prescribing omissions (PPOs) amongst community-dwelling older adults (≥65 years) enrolled to a clinical trial in three European countries.”

Comment: This study has important selection biases. SCH patients may have health problems such as more nervousness, anxiety, constipation, significantly increased concentrations of serum total cholesterol, LDL and triglyceride, among others. Further SCH patients are more likely to have some diagnoses included in the definition of PIM and PPO.

Response: Thank you for highlighting this. We have now included the following paragraph in the Limitations section of the Discussion.
“It is acknowledged that the sample size (n=532) is relatively small, however, the aim was to estimate and compare the prevalence and type of PIP and PPOs in a sample of patients from three European countries. The study population was based on participants enrolled to a clinical trial and may be somewhat different from the general population. However, the main inclusion criteria for the TRUST trial are quite broad. Secondly, although the data is based on a population of patients with SCH, there is no evidence to suggest that this would influence their chance of having a PIP or PPO.”
SCH is a subclinical condition and there is ongoing certainty as to whether it is truly associated with health problems.
As stated in the Methods section.

“SCH was defined in the TRUST study as persistently elevated thyroid stimulating hormone (TSH) levels (4.6-19.9 mU/L) with free thyroxine (T4) within the local laboratory reference range.”

Comment: The validity of prevalence estimates is limited due to authors included a subset of STOPP/START criteria. I assume this selection was made based on the lack of clinical and drug use information available. Approximately a 40% of STOPP/START were lost by methods.

Response: Thank you for this comment. It was only possible to apply a subset of the STOPP/START version 2 criteria, as information on drug strength, dose and duration of prescriptions was not available in the TRUST dataset. Also, to maximise the validity of our results, we felt it was prudent to only apply the criteria that we were confident we had full information on. This lack of information proved difficult in trying to capture all the criteria.
Previous studies have also experienced similar issues in applying the full set of the criteria e.g. Galvin et al “Prevalence of potentially inappropriate prescribing and prescribing omissions in older Irish adults: findings from The Irish LongituDinal Study on Ageing study (TILDA)”. European Journal of Clinical Pharmacology 2014; 70: 599-606. The authors could only apply a subset of 26 out of 65 STOPP and 10 out of 22 START version 1 criteria due to a lack of prescribing information.
While Verdoorn et al “Majority of drug-related problems identified during medication review are not associated with STOPP/START criteria” European Journal of Clinical Pharmacology 2015; 71: 1255-1262. The authors applied 25 out of 65 STOPP and 18 out of 22 START version 1 criteria.
Therefore, to address your comment we have included the following sentence in the “Strengths and Limitations” section at the beginning of the article:
“It was only possible to apply a subset of the criteria to the database due to a lack of information on drug strength, dose and duration of prescriptions and this may explain the low prevalence of PIP and PPOs in the study.”

Comment: The small sample size and sampling are an important limitation for the generalizability, but also over the odds to have a PIP and PPOs.

Response: Please see the response to your second comment.

Comment: It is unclear to me how the sample size was estimated and how representative it is for each country.

Response: Thank you for this comment. A sample size was not estimated for this cross-sectional study as we wanted to estimate the prevalence of PIP and PPOs in a sample of patients from three European countries. Therefore, all patients recruited to the TRUST trial in Ireland (n=115), Switzerland (n= 162) and the Netherlands (n=255) were included in the study.

Comment: In my opinion each comorbidity or Charlson Index need to be shown in detail and included in the multivariable analysis.

Response: Thank you for this suggestion, however we believe that it would not be feasible to include each comorbidity in the multivariable analysis as some patients reported up to 13 comorbidities in the study.
However, we have now provided an appendix with the prevalence of the most common comorbidities in each country.

Comment: Conclusions need to be wrote based on results and main findings.

Response: Thank you for this suggestion.

The main study conclusion is structured into three parts:
The first part summarises the main study findings:
“These study findings highlight that PIP and PPOs are prevalent among a sample of community-dwelling older people enrolled to a clinical trial in three European countries. “
The second part briefly explains the possible reasons for the variation in the prevalence of PIP and PPOs in the study:
“The screening process and identification of potential participants for this clinical trial differed between the countries and may explain some variation in the populations recruited and prevalence of PIP and PPOs. Guidelines for chronic disease, policies on costs and accessibility of medications influence prescribing practices and may impact on individual doctor’s prescribing decisions. “
The third part highlights areas of possible research:
“Further research is urgently needed into the impact of system level factors as this has important implications for patient safety, healthcare provision and economic costs. “

VERSION 2 – REVIEW

REVIEWER	Gulistan Bahat Istanbul University Istanbul Medical Faculty Turkia
REVIEW RETURNED	10-Nov-2017

GENERAL COMMENTS	I congratulate the authors for their successfull revision. Some minor points are needed to be revised. Introduction: 1) The authors have given PIP prevalnece by STOPP v1 in 2 countries: UK and Ireland. This stuıdy was erformed in Ireland, Swiss and Holland. So, they may check the literature if there is a prevalence study for Swiss and Holland and give their figures, instead of the UK figure. 2) Instead the authors should give the so far reported PIP prevalence by STOPP v2 in the literature as they are few and this stduy is all about the STOPP v2, not the v1.

REVIEWER	Marie Bradley FDA, US
REVIEW RETURNED	07-Nov-2017

GENERAL COMMENTS	1.While I am satisfied with the authors responses to most of my comments my issue with this paper remains and that is studies on PIP/PPO prevalence in order to be informative, in terms of being useful for designing interventions or informing policy to reduce these practices, need to be large and well designed. This study does not meet these criteria and is fraught with potential limitations which threaten its validity and so irrespective of the international comparison component this study is not really fit for purpose. Ideally in order to get a true indication of the extent of PIP/PPO studies need to be conducted among large numbers of patients, in electronic medical record databases or cohort studies with comprehensive clinical and prescribing information, not just a convenient sample, to get an accurate estimate of prevalence.

	Only then can this information be used to make changes to prescribing practices on a national or international level. The conclusion of this study is “These study findings highlight that PIP and PPOs are prevalent among a sample of 580 community-dwelling older people enrolled to a clinical trial in three European countries”. I am just not convinced at how useful the information obtained in this study is at all. 2.The authors state “There is a lack of international comparisons in this field, and despite the limitations of the sample we believe that this study will add to the research on the important issue about prescribing practices among older adults.”- I would like to know exactly what the authors feel it adds to the literature- this is vague. Why is it important to have international comparisons beyond being interesting? 3.The number of additional STOPP and START criteria applied in this study were 22 and 9 respectively.If there were only 15 new STOPP criteria added in version 2 how were 22 additional criteria studied. This is not clear? 4.In Switzerland a list of medications was obtained from the participants GP but as the authors have stated in the discussion GPs are not gatekeepers to care in Switzerland and so they may not have comprehensive information on all medications being taken especially those initiated by a specialist. Also in Switzerland it is unlikely that the GP would have information on OTC medications and some of the STOPP drugs such as aspirin and NSAIDs are available OTC- this question has not been addressed? 5.The conclusion states “Guidelines for chronic disease, policies on costs 584 and accessibility of medications influence prescribing practices and may impact on individual 585 doctor’s prescribing decisions.” Where in the discussion has this been mentioned?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gulistan Bahat
Institution and Country: Istanbul University, Istanbul Medical Faculty, Turkia.
I congratulate the authors for their successful revision. Some minor points are needed to be revised.
Introduction:

1) The authors have given PIP prevalence by STOPP v1 in 2 countries: UK and Ireland. This study was performed in Ireland, Swiss and Holland. So, they may check the literature if there is a prevalence study for Swiss and Holland and give their figures, instead of the UK figure.

Thank you for this suggestion. We have included a prevalence study for the Netherlands and a study carried out in a hospital setting in Switzerland. We would prefer to include the study from the Bradley et al. as it is based on PIP prevalence with a national study population in a European country. However, we are happy to remove this paper if necessary. The following section has been included in the Introduction:
“Bruin-Huisman et al. estimated the prevalence of potentially inappropriate medicines (PIMs) and PPOs among older patients in primary care in the Netherlands. In this retrospective longitudinal study the mean prevalence of ≥1 PIMs and PPOs were 34.7% and 84.8% respectively.

Urfer et al. assessed the efficacy and safety of a prescriber checklist for reducing inappropriate prescribing among 900 patients ≥65 years admitted to an internal ward of a Swiss hospital. The study reported that 37% of patients had ≥ 1 PIM while 25% had ≥ 1 PPO."
Page 7, Lines 160-166 in the revised track changes document
References: Bruin-Huisman et al. Potentially inappropriate prescribing to older patients in primary care in the Netherlands: a retrospective longitudinal study. Age and Ageing, Volume 46, Issue 4, 1 July 2017, p614–619.
Urfer et al. Intervention to Improve Appropriate Prescribing and Reduce Polypharmacy in Elderly Patients Admitted to an Internal Medicine Unit. PLOS ONE, Nov 2016; 11(11), p1-15.
2) Instead the authors should give the so far reported PIP prevalence by STOPP v2 in the literature as they are few and this study is all about the STOPP v2, not the v1.

Thank you for this suggestion. We have moved the studies that used STOPP/START V2 from the Discussion section and included them in the Introduction. It now reads as follows:
"However, there is a lack of research exploring the prevalence of PIP and PPOs in community-dwelling older adults using the updated criteria. For example, Blanco-Reina et al. reported a prevalence of PIP and PPOs of 40.4% and 21.8%, respectively among older adults in Spain. In a study conducted in Turkey, 667 participants aged ≥65 years were admitted to an outpatient clinic of a university hospital. The prevalence of PIP reported was 39.1%. A study conducted among 319 older patients discharged from a hospital in Albania identified that 63% received at least one PIP. In another study carried out in Ethiopia, the prevalence of inappropriate prescribing of antithrombotic therapy among 156 hospitalised elderly patients was assessed. The prevalence of PIP and PPOs were 51.4% and 48.6% respectively."
Page 7, Lines 166-174 in the revised track changes document.

Reviewer: 2
Reviewer Name: Marie Bradley
Institution and Country: FDA, US

1. While I am satisfied with the authors responses to most of my comments my issue with this paper remains and that is studies on PIP/PPO prevalence in order to be informative, in terms of being useful for designing interventions or informing policy to reduce these practices, need to be large and well designed. This study does not meet these criteria and is fraught with potential limitations which threaten its validity and so irrespective of the international comparison component this study is not really fit for purpose. Ideally in order to get a true indication of the extent of PIP/PPO studies need to be conducted among large numbers of patients, in electronic medical record databases or cohort studies with comprehensive clinical and prescribing information, not just a convenient sample, to get an accurate estimate of prevalence. Only then can this information be used to make changes to prescribing practices on a national or international level. The conclusion of this study is "These study findings highlight that PIP and PPOs are prevalent among a sample of 580 community-dwelling older people enrolled to a clinical trial in three European countries". I am just not convinced at how useful the information obtained in this study is at all.

Resposne: Thank you for this comment. We believe that this study is an important first step to justify the need for large comparative studies using routine data. This can then help to inform policy or the development of appropriate interventions on optimising prescribing practices in older adults at a national or international level. The conclusion now reads:

"These study findings highlight that PIP and PPOs are prevalent among a sample of community-dwelling older people enrolled to a clinical trial in three European countries. The screening process and identification of potential participants for this clinical trial differed between the countries and may explain some variation in the populations recruited and prevalence of PIP and PPOs.

This study is an important first step to justify the need for large comparative studies using routine data. This can then help to inform policy or the development of appropriate interventions on optimising prescribing practices in older adults at a national or international level. Further research is urgently needed into the impact of system level factors as this has important implications for patient safety, healthcare provision and economic costs."

Response: Page 26, lines 544-554 in the revised track changes document

2. The authors state "There is a lack of international comparisons in this field, and despite the limitations of the sample we believe that this study will add to the research on the important issue about prescribing practices among older adults."- I would like to know exactly what the authors feel it adds to the literature- this is vague. Why is it important to have international comparisons beyond being interesting?

Response: Thank you for this comment. The following section has been included in the "Strengths and Limitations" section of the Discussion:

"International comparisons can support or refute arguments for change in healthcare, serve as an additional lens on the state of the quality of care provided nationally, and can help build the evidence base necessary to identify problems and understand changes in the quality of care between countries."

Page 24, Lines 495-498 in the revised track changes document

3. The number of additional STOPP and START criteria applied in this study were 22 and 9 respectively. If there were only 15 new STOPP criteria added in version 2 how were 22 additional criteria studied. This is not clear?

Response: Thank you. To clarify your query the following section has been included in the Methods section:

"Twelve STOPP V1 criteria were removed from V2. Twenty seven new criteria were introduced in STOPP V2 and 22 of these criteria were applied to the TRUST dataset. Three START V1 criteria were removed from V2, while 15 new criteria were introduced. Nine of these new criteria were applied to the TRUST dataset."

Response: Page 9, Lines 212-215 in the revised track changes document

Reference: O' Mahony, D et al. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. Age and Aging, Vol 44, Issue 2, March 2015, p213-218.

4. In Switzerland a list of medications was obtained from the participants GP but as the authors have stated in the discussion GPs are not gatekeepers to care in Switzerland and so they may not have comprehensive information on all medications being taken especially those initiated by a specialist. Also in Switzerland it is unlikely that the GP would have information on OTC medications and some of the STOPP drugs such as aspirin and NSAIDs are available OTC- this question has not been addressed?

Response: Thank you for this comment. We have included the following section under "Strengths and Limitations" to clarify this:

“Although Swiss patients can avail of a health care plan that requires them to visit the GPs first, the GP may not have comprehensive information on all the patients’ medicines. This includes information on OTC medicines such as aspirin and NSAIDs which are included in the STOPP criteria and may explain the difference in PIP prevalence in these patients.”

Response: Page 25, Lines 530-534 in the revised track changes document

5. The conclusion states “Guidelines for chronic disease, policies on costs 584 and accessibility of medications influence prescribing practices and may impact on individual 585 doctor’s prescribing decisions.” Where in the discussion has this been mentioned?

Response: Thank you for this suggestion. We have now decided to remove this sentence as it does not fit with the revised conclusion.