

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)	Drivers and strategies for avoiding overuse. A cross-sectional study to explore the experience of Spanish primary care providers handling uncertainty and patients' requests
AUTHORS	Mira, José; Carrillo, Irene; Silvestre, Carmen; Pérez Pérez, Pastora; Nebot, Cristina; Olivera, Guadalupe; Gonzalez de Dios, Javier; ARANAZ ANDRÉS, JESÚS MARIA

VERSION 1 – REVIEW

REVIEWER	Kevin Selby Kaiser Permanente Division of Research, Oakland, CA, USA
REVIEW RETURNED	19-Jan-2018

GENERAL COMMENTS	<p>Thank you for the opportunity to review this interesting manuscript. Overuse is a topic of concern, and as the authors correctly state, little is known about the views of gatekeeper primary care providers on overuse. Such information could help guide interventions to address overuse.</p> <p>My main critique of this paper is that it focuses far too much on p-values of comparisons of questionable importance, presents an overwhelming number of results and comparisons, and needs significant editing. I would recommend focusing the article more, limiting the number of results presented, which will make it easier to follow and interpret. We would benefit greatly from seeing an English translation of the actual questionnaire. And finally, the limitations are inadequately addressed.</p> <p>Specific feedback:</p> <ul style="list-style-type: none">- consider having a title that includes a description of this as a cross-sectional survey of primary care providers- Abstract: please include sampling strategy and response rates. The meaning of the p-values and what comparisons are being made are unclear throughout the paper, but especially in the abstract. When you say higher pressure increased overuse, by how much and compared with what? I would consider presenting fewer results but making them clearer. Finally, I didn't understand how exactly "printed and digital media" could contribute to an inability to counter requests.- Introduction: good broad coverage of the topic and clearly stated objectives. What are "Questions to be replayed"?- Methods: big problem is not knowing how questions were asked and a clearer idea of the sampling strategy. Easiest would be to provide an English translation of the actual survey. How many emails were sent? What was done with incomplete questionnaires?- Any cognitive theory behind the questionnaire?
-------------------------	--

	<p>Results:</p> <ul style="list-style-type: none"> - please provide the response rate - when presenting comparisons, please provide the percentage difference, and not simply the test-statistic and p-value. With your large data set, p-values are less important than seeing the actual amount of difference. - Table 2, why have the p-values? there are far too many comparisons in this study, such that a p-value <0.05 begins to lose its meaning. Are we really interested in comparisons of question responses between GPs and pediatricians? that wasn't one of the objectives - Table 3: consider ordering based on overall score with maximum responsibility highest? very unclear. Again, why the p-values? - Several times, the authors present the profile of patient who requests certain things. Very interesting, but unclear how those results were obtained or quantified. - Several of the tables could benefit from adding "Total" columns and removing the p-values. <p>Conclusions:</p> <ul style="list-style-type: none"> - the authors state several times that defensive medicine is an important driver. However, "avoid a claim" was chosen relatively low. - what targets do you suggest for campaigns to reduce overuse? <p>For limitations:</p> <ul style="list-style-type: none"> - how does your sample compare to the overall composition of GPs and pediatricians in these organizations? - how does the use of email sampling and an electronic questionnaire bias your sample? - physicians might not admit to overuse, how would that bias your results? - this question might be better addressed by other qualitative studies such as focus groups - the questionnaire has not been previously validated
--	--

REVIEWER	Carissa Bonner University of Sydney, Australia
REVIEW RETURNED	23-Jan-2018

GENERAL COMMENTS	<p>Thank you for the opportunity to review this interesting paper reporting a large survey of Spanish primary care providers' perceptions about the overuse of medicines/tests/referrals. The paper is generally well written and describes the methods and results clearly. The paper could be improved with some additional explanation of the planned analyses and a more focused discussion. Specific suggestions/comments are below.</p> <p>INTRODUCTION</p> <p>Generally clear with a good rationale for the study Page 4 line 26 – remove “they” after [9, 18-19] Page 4 line 50 – remove “the” before Spanish primary care Page 4 line 52 – replace facing with face</p> <p>METHODS</p> <p>Well explained context, materials, recruitment Need more explanation of the planned analyses under statistical methods – What were your specific hypotheses? Did you plan to</p>
-------------------------	---

	<p>compare males/females, public/private and the three health professional groups on all survey responses or only the ones reported? Was any adjustment made for conducting a large number of statistical tests to reduce the error rate?</p> <p>Page 5 line 8 – need references for first paragraph under setting</p> <p>Page 4 line 8 - remove “the” after In Spain</p> <p>Page 5 line 26 – replace vary with varies</p> <p>Page 5 line 26 – swap order of “overuse unnecessary”</p> <p>Page 5 line 56 – replace replayed with answered</p> <p>Page 6 line 6 – replace validate with validated</p> <p>Page 6 line 7 – not sure what is meant by “although some participants encourages answers expected from the researcher”, can you clarify this please</p> <p>RESULTS</p> <p>The descriptions of significant effects are clear, although it is not clear what else was tested (see comment on methods). The tables are very helpful but would be easier to follow if they all showed GPs, pediatricians and nurses in the same order (e.g. table 1 has nurses first but table 4 has them last; and table 2 doesn’t report nurses when all the others do). It would also be helpful to show items from highest to lowest for GPs since they are the biggest group, to allow easier identification of the main drivers and differences between groups.</p> <p>DISCUSSION</p> <p>Could you start the discussion with an overall summary of the key drivers for each question in the results, rather than jumping straight to the last result on patient characteristics? Comparing each question to other research would be helpful, as you have done for frequency of requests on page 14 as this is useful for the reader.</p> <p>Page 12 line 40 – remove “the” and reverse order of “medicine defensive”</p> <p>Page 13 – be careful with wording to say that these findings are health professional perceptions rather than what is definitely happening in practice, e.g. page 13 line 49 “receive more aggressive responses” should really be “perceive more aggressive responses” since we are relying on self report</p> <p>Page 13 – also need to be careful about causation statements e.g. did you actually test this – line 21 “as pressure from patients becomes more insistent” and line 50 “as requests from patients become more misguided”? This wording implies a relationship based on a continuous data model.</p> <p>Page 14 - For practical implications, the idea of including arguments for patients in assistance algorithms (line 47) is a great idea and could be expanded with reference to the decision aid literature (where decision aids outlining evidence of benefit/harm have been shown to reduce more aggressive choices)</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Kevin Selby

Institution and Country: 1) University of Lausanne, Department of Ambulatory Care and Community Medicine 2) Kaiser Permanente Division of Research, Oakland, CA, USA

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Thank you for the opportunity to review this interesting manuscript. Overuse is a topic of concern, and as the authors correctly state, little is known about the views of gatekeeper primary care providers on overuse. Such information could help guide interventions to address overuse.

My main critique of this paper is that it focuses far too much on p-values of comparisons of questionable importance, presents an overwhelming number of results and comparisons, and needs significant editing. I would recommend focusing the article more, limiting the number of results presented, which will make it easier to follow and interpret. We would benefit greatly from seeing an English translation of the actual questionnaire. And finally, the limitations are inadequately addressed.

Many thanks for your valuable comments. We have taken the opportunity to improve our original from them. The English translation has been reviewed and the questionnaire is included now. The limitations have been reviewed.

Specific feedback:

- consider having a title that includes a description of this as a cross-sectional survey of primary care providers

The title has been changed to the following: "Drivers and strategies for avoiding overuse. A cross-sectional study to explore the experience of Spanish primary care providers handling uncertainty and patients' requests"

-Abstract: please include sampling strategy and response rates.

Both are now included. Sorry we did not include them before.

The meaning of the p-values and what comparisons are being made are unclear throughout the paper, but especially in the abstract.

We have tried to express ourselves in a much clearer way than before.

When you say higher pressure increased overuse, by how much and compared with what? I would consider presenting fewer results but making them clearer.

Sorry, let me try to explain. *Persistence* has been used in place of *pressure*.

Finally, I didn't understand how exactly "printed and digital media" could contribute to an inability to counter requests.

New medical advances and treatments that patients find on these types of media. This is now explained.

- Introduction: good broad coverage of the topic and clearly stated objectives. What are "Questions to be replayed"?

Thanks. We have tried to improve our original including the key questions explored.

- Methods: big problem is not knowing how questions were asked and a clearer idea of the sampling strategy. Easiest would be to provide an English translation of the actual survey.

The survey has been included.

How many emails were sent?

Our estimation is that 12787 emails were sent. From experience, we calculated that 88% were opened. We used this recruitment procedure because of the Spanish Personal Data Protection Law,

which prohibits us from directly sending the invitations to the professionals unless they have authorized us beforehand to write to them. Although we received excellent collaboration from the collaborating entities, we did not have complete control of this recruiting procedure.

What was done with incomplete questionnaires?

Incomplete questionnaires were not considered.

- Any cognitive theory behind the questionnaire?

No, there is not any cognitive theory behind the instrument. The report of the questionnaire designed by the ABIM Foundations did not inform about it.

Results:

- please provide the response rate

This is now included.

- when presenting comparisons, please provide the percentage difference, and not simply the test-statistic and p-value. With your large data set, p-values are less important than seeing the actual amount of difference.

The percentage difference and 95% interval confidants are now included.

- Table 2, why have the p-values? there are far too many comparisons in this study, such that a p-value <0.05 begins to lose its meaning. Are we really interested in comparisons of question responses between GPs and pediatricians? that wasn't one of the objectives

The two last columns of this table have been erased. These comparisons do not introduce relevant information.

- Table 3: consider ordering based on overall score with maximum responsibility highest? very unclear. Again, why the p-values?

The tables have been reviewed and ordered. In some cases, the p-values have been erased.

- Several times, the authors present the profile of patient who requests certain things. Very interesting, but unclear how those results were obtained or quantified.

Instruments used are now included as an annex.

- Several of the tables could benefit from adding "Total" columns and removing the p-values.

This suggestion has been applied.

Conclusions:

- the authors state several times that defensive medicine is an important driver. However, "avoid a claim" was chosen relatively low.

Yes, this is true, However, we have considered a set of data. Avoiding a claim was chosen by 18.9% and 11.7% of GP and pediatricians, respectively, but avoiding a future demand, by 21.4% and 13.5%; due to following the standard (avoiding a claim), by 11.9% and 3.1%; and, finally, to gain greater control and safety, by 38.4% and 42.7%.

- what targets do you suggest for campaigns to reduce overuse?

We probably need to research about if Do Not Do has a higher chance of producing an adverse event.

For limitations:

- how does your sample compare to the overall composition of GPs and pediatricians in these organizations?

Information about this has been introduced.

- how does the use of email sampling and an electronic questionnaire bias your sample?

- physicians might not admit to overuse, how would that bias your results?

- this question might be better addressed by other qualitative studies such as focus groups

- the questionnaire has not been previously validated

These valuable ideas have been introduced as a limitation.

Reviewer: 2

Reviewer Name: Carissa Bonner

Institution and Country: University of Sydney, Australia

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Thank you for the opportunity to review this interesting paper reporting a large survey of Spanish primary care providers' perceptions about the overuse of medicines/tests/referrals. The paper is generally well written and describes the methods and results clearly. The paper could be improved with some additional explanation of the planned analyses and a more focused discussion. Specific suggestions/comments are below.

Thanks for all your work reviewing this original. Your valuable comments have been very useful for us because they improved this article.

INTRODUCTION

Generally clear with a good rationale for the study

Page 4 line 26 – remove “they” after [9, 18-19]

Page 4 line 50 – remove “the” before Spanish primary care

Page 4 line 52 – replace facing with face

Thanks. Changes have been introduced.

METHODS

Well explained context, materials, recruitment

Need more explanation of the planned analyses under statistical methods – What were your specific hypotheses? Did you plan to compare males/females, public/private and the three health professional

groups on all survey responses or only the ones reported? Was any adjustment made for conducting a large number of statistical tests to reduce the error rate?

Information describing the comparisons is now included.

Page 5 line 8 – need references for first paragraph under setting

Page 4 line 8 - remove “the” after In Spain

Page 5 line 26 – replace vary with varies

Page 5 line 26 – swap order of “overuse unnecessary”

Page 5 line 56 – replace replayed with answered

Page 6 line 6 – replace validate with validated

Page 6 line 7 – not sure what is meant by “although some participants encourages answers expected from the researcher”, can you clarify this please

All these mistakes have been resolved. Pag6 line 7 has been explained. Thanks for all this.

RESULTS

The descriptions of significant effects are clear, although it is not clear what else was tested (see comment on methods). The tables are very helpful but would be easier to follow if they all showed GPs, pediatricians and nurses in the same order (e.g. table 1 has nurses first but table 4 has them last; and table 2 doesn't report nurses when all the others do). It would also be helpful to show items from highest to lowest for GPs since they are the biggest group, to allow easier identification of the main drivers and differences between groups.

Yes, this data is clearer and easier to understand.

Same order has been introduced.

Tables are now showing items from highest to lowest.

Table 2 does not include nurses because they cannot order tests. The title has been modified because the beginning text could introduce confusion.

DISCUSSION

Could you start the discussion with an overall summary of the key drivers for each question in the results, rather than jumping straight to the last result on patient characteristics? Comparing each question to other research would be helpful, as you have done for frequency of requests on page 14 as this is useful for the reader.

We have tried to introduce changes to explain the results in a good way. Discussion has been reviewed.

Page 12 line 40 – remove “the” and reverse order of “medicine defensive”

This suggestion has been introduced.

Page 13 – be careful with wording to say that these findings are health professional perceptions rather than what is definitely happening in practice, e.g. page 13 line 49 “receive more aggressive responses” should really be “perceive more aggressive responses” since we are relying on self report

This is true. A change has been introduced.

Page 13 – also need to be careful about causation statements e.g. did you actually test this – line 21 “as pressure from patients becomes more insistent” and line 50 “as requests from patients become more misguided”? This wording implies a relationship based on a continuous data model.

In both sentences, *insistent* has been used.

Page 14 - For practical implications, the idea of including arguments for patients in assistance algorithms (line 47) is a great idea and could be expanded with reference to the decision aid literature (where decision aids outlining evidence of benefit/harm have been shown to reduce more aggressive choices)

Thanks. This is an excellent proposal. We have included a comment and a new reference.

VERSION 2 – REVIEW

REVIEWER	Kevin Selby University of Lausanne, Switzerland
REVIEW RETURNED	08-Mar-2018

GENERAL COMMENTS	<p>This manuscript has improved greatly, and I congratulate the authors on their hard work. Some small adjustments should still be made prior to publication.</p> <p>Notably, in the abstract: Results paragraph:</p> <ol style="list-style-type: none"> 1. "Persistence and aggressiveness by patients increased overuse ($p < 0.001$)." -> this statement implies that the authors have before-after results showing that persistence resulted in overuse. I'm not sure exactly what the authors are referring to and what comparison the p-value refers to. 2. "Evidence and clinical safety were the arguments that dissuade patients from their requests the most ($p < 0.001$). Cost savings are not a convincing argument, above all for pediatricians ($p < 0.001$)." --> It is unclear to me what comparisons are being represented by the p-values. I'm assuming they are for comparisons between professionals? Please make clear comparison statements, and consider presenting the proportion in each group to make the results clearer. 3. I would consider removing the p-values for comparison between professional groups completely, given that that is not listed as an objective in the abstract. I would continue giving the number of providers who responded in each category to show which answers were most frequent. <p>The abstract conclusion: "...explains overuse" suggests that the elements listed explain all overuse. I would consider changing the wording.</p>
-------------------------	---

REVIEWER	Carissa Bonner University of Sydney, Australia
REVIEW RETURNED	19-Mar-2018

GENERAL COMMENTS	Thank you for addressing my previous comments, which have mostly been addressed well. A few clarifications are still needed for the measures and new sections of the discussion.
-------------------------	--

	<p>Page 5 - I suggest removing this whole section: "A leading questions bias was controlled using validated questions, although this bias could not have been controlled at all, and some participants encouraged answers expected from the researcher." This is still very unclear - was the leading question biased controlled or not? How did participants encourage certain answers? If you used validated questions, you would need some more detail on this with references and reliability/validity statistics for the measure/s used. But in the discussion you say "Although the questionnaire was used in a previously study,[12] it has not been previously validated."</p> <p>Discussion – this now gives a better overview of findings and comparison to other research, but some parts need to be reworded to better reflect the method of self-report of perceived drivers from health professionals:</p> <p>Page 11 - "In this study, female patients and patients who suffer an unspecified or yet-undiagnosed pathology exerted greater pressure upon the professionals." change to "In this study, health professionals reported greater pressure from female patients and patients who suffer an unspecified or yet-undiagnosed pathology."</p> <p>Page 12 – "This study's findings reveal that as pressure from patients becomes more insistent... General practitioners are pressured more by their patients than pediatricians or nurses..." change to "This study's findings reveal that as perceived pressure from patients becomes more insistent... General practitioners perceive more pressure from their patients than pediatricians or nurses..."</p> <p>These parts need revised wording to clarify issues or explain what you mean:</p> <p>Page 12 – clarify what you mean by the gatekeepers implication in "and suggests the gatekeeper's implication could not prevent (or at least reduce) overuse."</p> <p>Page 12 – need to revise wording to reflect that PSA tests do actually carry risks such as overdiagnosis and overtreatment of inconsequential cancers with no mortality benefit to treatment "It is unlikely that ordering a test such as the prostate-specific antigen test (PSA) in an asymptomatic male who insists so he can "rest easy" is the same as initiating a totally contraindicated treatment and one that poses risks for the patient." – suggest changing "poses risks" to "poses immediate risks", or using a different example.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

On behalf of the authors of the original title "Drivers and strategies for avoiding overuse. A cross-sectional study to explore the experience of Spanish primary care providers handling uncertainty and patients' requests" we appreciate very much all the word done by the reviewers. Their suggestions have improved this paper. We have introduced changes in red to facilitate its identification.

Reviewer: 1

This manuscript has improved greatly, and I congratulate the authors on their hard work. Some small adjustments should still be made prior to publication.

Many thanks for your valuable comments.

Notably, in the abstract:

Results paragraph:

1. "Persistence and aggressiveness by patients increased overuse ($p < 0.001$)."

-> this statement implies that the authors have before-after results showing that persistence resulted in overuse. I'm not sure exactly what the authors are referring to and what comparison the p-value refers to.

This statement has been removed to avoid confusion.

2. "Evidence and clinical safety were the arguments that dissuade patients from their requests the most ($p < 0.001$). Cost savings are not a convincing argument, above all for pediatricians ($p < 0.001$)."

--> It is unclear to me what comparisons are being represented by the p-values. I'm assuming they are for comparisons between professionals? Please make clear comparison statements, and consider presenting the proportion in each group to make the results clearer.

The p-values have been eliminated and replaced by the proportions data. Only in one case, we have kept the p-value but included the chi-square and percentage of difference data.

3. I would consider removing the p-values for comparison between professional groups completely, given that that is not listed as an objective in the abstract. I would continue giving the number of providers who responded in each category to show which answers were most frequent.

We have considered your appreciation and made the changes indicated.

The abstract conclusion:

"...explains overuse" suggests that the elements listed explain all overuse. I would consider changing the wording.

We have qualified this statement and now suggest that the elements listed explain only part of the overuse because the referee point out a right observation.

Reviewer: 2

Thank you for addressing my previous comments, which have mostly been addressed well. A few clarifications are still needed for the measures and new sections of the discussion.

Thanks for all your work. Your valuable comments have been very useful for us.

Page 5 - I suggest removing this whole section: "A leading questions bias was controlled using validated questions, although this bias could not have been controlled at all, and some participants encouraged answers expected from the researcher." This is still very unclear - was the leading question biased controlled or not? How did participants encourage certain answers? If you used validated questions, you would need some more detail on this with references and reliability/validity statistics for the measure/s used. But in the discussion you say "Although the questionnaire was used in a previously study,[12] it has not been previously validated."

This is true. This section has been removed.

Discussion – this now gives a better overview of findings and comparison to other research, but some parts need to be reworded to better reflect the method of self-report of perceived drivers from health professionals:

Page 11 - “In this study, female patients and patients who suffer an unspecified or yet-undiagnosed pathology exerted greater pressure upon the professionals.” change to “In this study, health professionals reported greater pressure from female patients and patients who suffer an unspecified or yet-undiagnosed pathology.”

The suggested change has been made.

Page 12 – “This study’s findings reveal that as pressure from patients becomes more insistent... General practitioners are pressured more by their patients than pediatricians or nurses...” change to “This study’s findings reveal that as perceived pressure from patients becomes more insistent... General practitioners perceive more pressure from their patients than pediatricians or nurses...”

The suggested change has been made.

These parts need revised wording to clarify issues or explain what you mean:

Page 12 – clarify what you mean by the gatekeepers implication in “and suggests the gatekeeper’s implication could not prevent (or at least reduce) overuse.”

We have changed wording of this statement to avoid confusion.

Page 12 – need to revise wording to reflect that PSA tests do actually carry risks such as overdiagnosis and overtreatment of inconsequential cancers with no mortality benefit to treatment “It is unlikely that ordering a test such as the prostate-specific antigen test (PSA) in an asymptomatic male who insists so he can “rest easy” is the same as initiating a totally contraindicated treatment and one that poses risks for the patient.” – suggest changing “poses risks” to “poses immediate risks”, or using a different example.

The suggested change has been made.