## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Regional differences in the patient population of general practices in northern Germany - results of a mixed methods study |
| **AUTHORS** | Hansen, Heike; Schäfer, Ingmar; Pohontsch, Nadine; Kazek, Agata; Hardt, Hanna; Lühmann, Dagmar; Scherer, Martin |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Sean Lyons<br>Economic and Social Research Institute, Dublin, Ireland |
| **REVIEW RETURNED** | 06-Aug-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | This paper reports the results of mixed methods research into the types of patients presenting at GP practices in northern Germany. Variations are identified in the most prevalent types of patients at practices across a range of regional categories. The paper is well organised and clearly presented. The findings are plausible, and the authors include a useful comparison of the findings to previous research.<br><br>My only substantive concerns with the paper are to do with the regression analysis used to identify regional differences in the frequencies of patient types by region. First, the regression analysis needs to be more thoroughly explained. The description of these regressions is confined to three sentences at the end of p.7. It is usual when employing these methods to make it clear what equations were estimated, exactly which estimator was used and what variables were on the left- and right-hand-sides of the regressions. I can't work out what they have done from the description or from inspection of Tables 5 and 6. Often a sample equation is included to help illustrate these components. From the brief description it seems like the outcome variables were fractional; did the authors use a fractional logit estimator or (as the text says) a logistic?<br><br>Second, the authors use stepwise backward selection in their regression models to identify the types of patients that are more or less prevalent across areas. Again, little detail is provided on how this was done. However, in general stepwise regression is not a reliable method of variable reduction or selection (Thompson, 1995). The coefficients tend to be biased upwards in scale and the confidence intervals on them are biased in the direction of narrowness, i.e. the probability of false positives is increased (see e.g. Altman and Andersen, 1989). The results of the models tend to be highly sensitive to sampling error, which is a particular problem when the sample size is small (as in this case). There are better alternatives to stepwise regression for carrying out model selection. Some have a long pedigree such as 'all possible |

| | |
|---|---|
| | alternatives' methods, but these can give rise to other problems. A more recent method is to use a penalised regression approach such as Lasso. This is implemented in Stata (Ahrens et al., 2020), though I am not sure if it is available with a fractional outcome variable.<br><br>Apart from this, I have only minor suggestions on wording and typos:<br>- In the results para in the abstract, in place of "significantly lower associated" I suggest "significantly less prevalent in urban areas"<br>- In the final para of the abstract, "noncompliant" might be better than "incompliant".<br>- In bullet 2 on p3, "…we ensured to include both male and female GPs …" could be better phrased as "…we ensured that both male and female GPs were included…"<br>- In the 2nd line of the introduction, the references given on difficulties in rural recruitment both refer to Germany, so maybe it would be best to specify "…rural areas in Germany struggle…"<br>- At the start of p5, in place of "…into a characteristic property pattern…" I suggest something like "…into patterns of characteristic properties…"<br>- At the bottom of p10, I'm not sure what you mean by "had a special treatment effort in common". Is it that all these groups required additional treatment effort? The following sentence could also be clearer; again, do you mean to say that taking responsibility for older patients and organising their medical treatment led to a higher workload for these patients?<br>- On p11, line 44, where you say "concerning patient's behaviour" I wonder if you might mean "on the basis of common behaviours" or something like that.<br>- On p13, line 16, it might be clearer to replace "…associated with urban areas in comparison to rural areas." with "…more prevalent in urban areas than in rural areas."<br>- There are two places where you need to replace a ". Whereas" with ", whereas" to avoid sentence fragments: p.15 and p16.<br>- Finally, towards the end of p15, in place of "…but less NPs…" you might say "…but NPs less often…" and in the next sentence I wonder if where you say "cumulative workload" you mean "higher workload".<br>References<br>Ahrens, A., Hansen, C.B. and Schaffer, M.E., 2020, lassopack: Model selection and prediction with regularized regression in Stata. The Stata Journal 20(1): 176-235.<br>Altman, D. G. and Andersen, P. K., 1989, Bootstrap investigation of the stability of a Cox regression model. Statistics in Medicine 8: 771-783.<br>Thompson, B., 1995, Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. Educational and Psychological Measurement 55 (4): 525-534. |

| | |
|---|---|
| **REVIEWER** | E/Prof D'Arcy Holman<br>School of Population & Global Health<br>The University of Western Australia<br>Australia |
| **REVIEW RETURNED** | 16-Aug-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for this meticulously designed, executed and reported study on regional difference in the patient population of general practices in northern Germany. I had no concerns about the manuscript that were sufficiently serious to warrant a 'no' response |

| | |
|---|---|
| | on the review checklist. But nevertheless there are some issues I wish to raise with you and which I hope you will consider as a means to improve your paper:<br>1. Literal vs interpretative translation of statements by participants in focus groups: These statements are a valuable element of the manuscript. But would you please clarify your methodological approach to language translation, assuming that most of the original material was recorded in German and now reported in English. My impression is that you have used a parsimonious interpretive approach; ie, the interpretive component of the translation from the source language is the minimum required, in the judgment of the translator, to convey an adequate (but not necessarily complete) meaning in the target language, thus leaving some residual onus of interpretation to the reader. For example, on p13, line 35 you report “Oh, we’ve heard so much beautiful from you”, whereas a fully interpretive translation might be “Oh, I’ve heard so much praise about you”. If your intention was to apply a parsimonious interpretive approach then I support this decision. The point is that the method of translation itself is sufficiently important in trans-lingual qualitative research to warrant documenting it as part of your methods.<br><br>2. Appropriate use of logistic regression in a study of patient mix: Logistic regression models including potential confounders as covariates are appropriate to a research question such as, “Is the prevalence at presentation of patients of type (_____) associated with an urban (or rural) practice setting independent of other recorded patient characteristics?” This question is essentially one of aetiological significance and stems from a motivation to accrue a different type of knowledge that a research question such as, “Does a GP in an urban (or rural) practice setting see proportionally more patients of type (____)?” This second question is one about variations in the mix of actual patient workload and is the form of question most relevant to the assessment of GP training priorities. The control of confounding and the strength of unconfounded association between practice setting and patient type is irrelevant to the actual mix of patients that the GP sees. So, for example, at p14, line 12 you report that ‘minors accompanied by their parents’ accounted for 6.3% of rural patients vs 3.1% of urban patients. Thus, the prevalence ratio is 2.03 and the (superfluous) crude prevalence odds ratio is, as we would expect, a little higher at 2.10. The adjusted prevalence odds ratio derived from your logistic regression model is 1.40 (ie, 1/0.71). So, which of these measures provides the most valid reflection of the relative workload from this type of paediatric patient in rural vs urban GP practice settings? – I would say the prevalence ratio of 2.03 because a rural GP who sees 1,000 patients is twice as likely to see this type of patient than an urban GP who sees 1,000 patients (and not merely 40% more likely). I do not object to the inclusion of your logistic regression analysis, but do recommend you reconsider how to interpret the results and where your emphasis should be placed.<br><br>3. Priorities for medical training: At the end of your discussion, you conclude that GP training for urban vs rural practice settings warrants differences in syllabi to align with the different mixes of patients in these settings. I question if this could be an over-simplification and even, possibly, a conclusion based on unsubstantiated premise. For example, Table 4 indicates that 1 in 5 patients in the urban setting had an identified psychiatric |

| | |
|---|---|
| | disorder compared with 1 in 8 in the rural setting. But is this difference sufficient to warrant different training priorities? Patient with psychiatric disorders are commonplace in both settings and perhaps a more relevant question is whether mental health is given sufficient attention in GP training in general. A ground-breaking aspect of your research is your preference for classifying patients by complexity of morbidity, sociodemogrphics, specific needs and behavioural features more so than by traditional (eg, ICD) disease categories. Reading in your paper how commonly GPs encounter such issues as poor therapy adherence, hypochondriasis and prescription abuse raises questions as to whether GP in all settings are adequately prepared by their training to communicate about and deal effectively with these complex problems; but your discussion is silent on this important aspect of your work. What does the GP training literature have to say about how training priorities should be set? Is it simply a matter of matching GP training to predicted patient mix such that the most frequently encountered problems receive the greatest attention? Or is more appropriate that training is enhanced in areas that are most frequently at risk of poor management leading to poor outcomes due to complexity and unfamiliarity. These might not necessarily be the most common types of patient presentation (but nor should they be rare occurrences). The point is that your discussion and conclusions would benefit from a more critical consideration of the training implications of your important study. |

| | |
|---|---|
| **REVIEWER** | Mazumdar, Soumya<br>SWSLHD, Australia |
| **REVIEW RETURNED** | 31-Aug-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have done a detailed qualitative analysis on patients within a group of GPs in German. I believe this is an interesting study, and I had a few comments:<br><br>1. The term “environs” in line 58 is a bit strange. Is it something that the Federal Institute of Research on Building… have used, or is it something that the authors have decided to use. In the context the authors use this term it seems to me that they mean peri urban areas.<br>2. Line 60: The sentence may be better phrased as “with signs of urban agglomeration”<br>3. Do each of the categories in lines 51-60 have a population density band/threshold? This would be especially useful if researchers for other countries/jurisdictions/contexts were trying to do a comparison study.<br>4. Recruitment: Lines 19 to 47: More details of the survey parameters would be useful. What was the survey response rates? Did the authors gauge non response bias? Could the non-responding GPs have been different from the ones that responded?<br>5. Lines 51-60: Data Collection: Could there have been changes to the study populations between 2014 (quite a while back) and now? For instance, do any of the study areas have large concentrations of Syrian refugees, and that would have happened after 2014?<br>6. Was the study approved by an Ethics committee/Institutional Review Board?<br>7. Data Analysis: Lines 48-60: Is there a specific reason why the authors chose to use two logistic regressions instead of one multivariate logistic regression? |

| | |
|---|---|
| | 8. Results, line 14: The mean GP age is quite high... is this normal for Germany? A line or two on this may be useful to the international reader.<br>9. Results: Sample characteristics: This may be the place to discuss response rates and non-responder characteristics that I mentioned in 4.<br>10. Page 9: Line 42-43: Please check the grammar of this sentence, “the need to ward of desires…<br>11. Page 9: Line 56: Please change bad to poor and condition to conditions.<br>12. The qualitative coding is very interesting.<br>13. The discussion section is detailed and generally fine |

**VERSION 1 – AUTHOR RESPONSE**

Point-by-point response Reviewer 1:
Reviewer Name: Sean Lyons
Institution and Country: Economic and Social Research Institute, Dublin, Ireland
This paper reports the results of mixed methods research into the types of patients presenting at GP practices in northern Germany. Variations are identified in the most prevalent types of patients at practices across a range of regional categories. The paper is well organised and clearly presented. The findings are plausible, and the authors include a useful comparison of the findings to previous research.
My only substantive concerns with the paper are to do with the regression analysis used to identify regional differences in the frequencies of patient types by region. First, the regression analysis needs to be more thoroughly explained. The description of these regressions is confined to three sentences at the end of p.7. It is usual when employing these methods to make it clear what equations were estimated, exactly which estimator was used and what variables were on the left- and right-hand-sides of the regressions. I can’t work out what they have done from the description or from inspection of Tables 5 and 6. Often a sample equation is included to help illustrate these components. From the brief description it seems like the outcome variables were fractional; did the authors use a fractional logit estimator or (as the text says) a logistic?
Answer: As suggested, we now described which items were used as independent variables and how the dependent variables was defined. We also added the equation of the regression analysis. We now clarified that the regional categories were used as dependent variables and that they were not coded as fractional variables, but coded 0/1: “The full number (n) of identified patient types were introduced as independent variables (xi) into the backward selection and the regional category (coded 0/1) was used as dependent variable (y).”

Second, the authors use stepwise backward selection in their regression models to identify the types of patients that are more or less prevalent across areas. Again, little detail is provided on how this was done. However, in general stepwise regression is not a reliable method of variable reduction or selection (Thompson, 1995). The coefficients tend to be biased upwards in scale and the confidence intervals on them are biased in the direction of narrowness, i.e. the probability of false positives is increased (see e.g. Altman and Andersen, 1989). The results of the models tend to be highly sensitive to sampling error, which is a particular problem when the sample size is small (as in this case). There are better alternatives to stepwise regression for carrying out model selection. Some have a long pedigree such as ‘all possible alternatives’ methods, but these can give rise to other problems. A more recent method is to use a penalised regression approach such as Lasso. This is implemented in Stata (Ahrens et al., 2020), though I am not sure if it is available with a fractional outcome variable.
Answer: We assessed the alternative to stepwise regression you suggested. The ‘all possible alternatives’ methods are not feasible with our data. As described in the article of Ahrens et al., 2020,

with only 20 predictors, there are more than 1 million different models to consider and our data set includes 27 predictors. We also considered using Lasso, which is regularly implemented in Stata 16.0. However, Lasso was not convincing for us for a number of reasons. As it is based on penalized regression the sizes of the regression coefficients are individually down-weighted, ie, not proportional to coefficients from a normal logistic regression and therefore not clinically meaningful and extremely difficult to interpret in our view. Lasso also does not return standard errors, confidence intervals or any measure of statistical significance making it difficult to decide how well the Lasso results represent the population. It is possible to introduce Lasso coefficients in bootstrap regression analyses for obtaining p-values, but this even increase the problem that is will be very difficult for the average reader to understand and interpret analyses and results.
After doing some literature search, we gained the impression that at the present time not all statistical publications prefer penalized regression like Lasso over backward selection algorithms. We found that today, backward selection is still commonly used in medical publications and the proportion of Lasso analyses in publications with other subject areas than statistics seems to be tiny. We therefore decided to keep the backward selection analyses in our manuscript and to clearly communicate their weaknesses. For that reasons, we added a passage in the strengths and limitations section of the manuscript: “The stepwise variable selection used for identifying significant differences between the regions reacts sensitively to differences in the distribution of the variables and it is not considered a reliable method of variable selection [24]. The results from these analyses therefore describe only one possible, but not necessarily the best solution. Additionally, coefficients resulting from stepwise backward selection analyses tend to be biased upwards in scale and the probability of false positive results is increased [25]. For this reasons, these analyses should be interpreted with care and considered as purely explorative.”

Apart from this, I have only minor suggestions on wording and typos:
- In the results para in the abstract, in place of “significantly lower associated” I suggest “significantly less prevalent in urban areas”
- In the final para of the abstract, “noncompliant” might be better than “incompliant”.
- In bullet 2 on p3, “…we ensured to include both male and female GPs …” could be better phrased as “…we ensured that both male and female GPs were included…”
- In the 2nd line of the introduction, the references given on difficulties in rural recruitment both refer to Germany, so maybe it would be best to specify “…rural areas in Germany struggle…”
- At the start of p5, in place of “…into a characteristic property pattern…” I suggest something like “…into patterns of characteristic properties…”
Answer: All of these points were revised as suggested.

- At the bottom of p10, I’m not sure what you mean by “had a special treatment effort in common”. Is it that all these groups required additional treatment effort? The following sentence could also be clearer; again, do you mean to say that taking responsibility for older patients and organising their medical treatment led to a higher workload for these patients?
Answer: Thanks for this comment. We changed this paragraph to “Patients regularly needing home visits, patients living in a nursing home or senior citizens living on their own without caregivers had in common that they required an additional treatment effort. GPs took responsibility for their older patients and they have to organize their medical treatment which led to a higher workload.” This is hopefully clearer.

- On p11, line 44, where you say “concerning patient’s behaviour” I wonder if you might mean “on the basis of common behaviours” or something like that.
Answer: We changed “concerning patient’s behaviour” into “on the basis of common behaviours”. Thanks for this relevant note.

- On p13, line 16, it might be clearer to replace “…in comparison to rural areas.” with “…more

prevalent in urban areas than in rural areas."
Answer: We revised the paragraph as suggested.

- There are two places where you need to replace a ". Whereas" with ", whereas" to avoid sentence fragments: p.15 and p16.
Answer: We revised this.

- Finally, towards the end of p15, in place of "…but less NPs…" you might say "…but NPs less often…" and in the next sentence I wonder if where you say "cumulative workload" you mean "higher workload".
Answer: Thanks for this note. We revised this.

Point-by-point response Reviewer 2:
Reviewer: 2
Reviewer Name: D'Arcy Holman
Institution and Country: The University of Western Australia, Australia

Thank you for this meticulously designed, executed and reported study on regional difference in the patient population of general practices in northern Germany. I had no concerns about the manuscript that were sufficiently serious to warrant a 'no' response on the review checklist. But nevertheless there are some issues I wish to raise with you and which I hope you will consider as a means to improve your paper:
1. Literal vs interpretative translation of statements by participants in focus groups: These statements are a valuable element of the manuscript. But would you please clarify your methodological approach to language translation, assuming that most of the original material was recorded in German and now reported in English. My impression is that you have used a parsimonious interpretive approach; ie, the interpretive component of the translation from the source language is the minimum required, in the judgment of the translator, to convey an adequate (but not necessarily complete) meaning in the target language, thus leaving some residual onus of interpretation to the reader. For example, on p13, line 35 you report "Oh, we've heard so much beautiful from you", whereas a fully interpretive translation might be "Oh, I've heard so much praise about you". If your intention was to apply a parsimonious interpretive approach then I support this decision. The point is that the method of translation itself is sufficiently important in trans-lingual qualitative research to warrant documenting it as part of your methods.
Answer: Thank you for this comment. As you assumed we used a parsimonious interpretive approach to language translation and stayed as close as possible to a literal translation of the quotations. We added a short sentence in the methods section.

2. Appropriate use of logistic regression in a study of patient mix: Logistic regression models including potential confounders as covariates are appropriate to a research question such as, "Is the prevalence at presentation of patients of type (_____) associated with an urban (or rural) practice setting independent of other recorded patient characteristics?" This question is essentially one of aetiological significance and stems from a motivation to accrue a different type of knowledge that a research question such as, "Does a GP in an urban (or rural) practice setting see proportionally more patients of type (____)?" This second question is one about variations in the mix of actual patient workload and is the form of question most relevant to the assessment of GP training priorities. The control of confounding and the strength of unconfounded association between practice setting and patient type is irrelevant to the actual mix of patients that the GP sees. So, for example, at p14, line 12 you report that 'minors accompanied by their parents' accounted for 6.3% of rural patients vs 3.1% of urban patients. Thus, the prevalence ratio is 2.03 and the (superfluous) crude prevalence odds ratio is, as we would expect, a little higher at 2.10. The adjusted prevalence odds ratio derived from your logistic regression model is 1.40 (ie, 1/0.71). So, which of these measures provides the most

valid reflection of the relative workload from this type of paediatric patient in rural vs urban GP practice settings? – I would say the prevalence ratio of 2.03 because a rural GP who sees 1,000 patients is twice as likely to see this type of patient than an urban GP who sees 1,000 patients (and not merely 40% more likely). I do not object to the inclusion of your logistic regression analysis, but do recommend you reconsider how to interpret the results and where your emphasis should be placed.
Answer: Thank you for raising this very important point. We totally agree with you that the unadjusted prevalence ratios show the mix of patients GPs actually see and clarified the importance of the logistic regression analyses in the statistical analyses section: “As it might be that patient types are correlated, ie, patients systematically belong to more than one type, we also analysed in which patient types the biggest regional difference can be found. These variables were identified by logistic regression analyses…” We also revised the main findings section accordingly: “The biggest difference between urban and rural areas were found in five patient types being more prevalent in urban areas and in three patient types being more prevalent in rural areas.”

3. Priorities for medical training: At the end of your discussion, you conclude that GP training for urban vs rural practice settings warrants differences in syllabi to align with the different mixes of patients in these settings. I question if this could be an over-simplification and even, possibly, a conclusion based on unsubstantiated premise. For example, Table 4 indicates that 1 in 5 patients in the urban setting had an identified psychiatric disorder compared with 1 in 8 in the rural setting. But is this difference sufficient to warrant different training priorities? Patient with psychiatric disorders are commonplace in both settings and perhaps a more relevant question is whether mental health is given sufficient attention in GP training in general. A ground-breaking aspect of your research is your preference for classifying patients by complexity of morbidity, sociodemogrphics, specific needs and behavioural features more so than by traditional (eg, ICD) disease categories. Reading in your paper how commonly GPs encounter such issues as poor therapy adherence, hypochondriasis and prescription abuse raises questions as to whether GP in all settings are adequately prepared by their training to communicate about and deal effectively with these complex problems; but your discussion is silent on this important aspect of your work. What does the GP training literature have to say about how training priorities should be set? Is it simply a matter of matching GP training to predicted patient mix such that the most frequently encountered problems receive the greatest attention? Or is more appropriate that training is enhanced in areas that are most frequently at risk of poor management leading to poor outcomes due to complexity and unfamiliarity. These might not necessarily be the most common types of patient presentation (but nor should they be rare occurrences). The point is that your discussion and conclusions would benefit from a more critical consideration of the training implications of your important study.
Answer: Thank you for this again very important note. Due to your suggestions we revised the second paragraph of the implications for research and clinical practice. The new paragraph is the following: “The identified regional differences should also be included as learning content in the training of medical students and young GPs. In Germany the training of GPs is regulated by the respective regulations on continuing medical education of the federal states [45]. This results in a great variety and legal differences in the federal states. These trainings include the identified problems as psychosomatic primary care, addiction therapy or social medicine but to our knowledge they do not focus on regional differences [46]. The Baden-Württemberg General Practice Competence Center has developed Germany's first competence-based curriculum for general practice training assistants. GPs and the German College of General Practitioners and Family Physicians (DEGAM) were involved [47]. This curriculum does not include either the topic regional differences of patient types in general practice. Future revisions of these curricula should consider these regional differences. Future GPs could compensate the specific needs of their patient clientele with medical training aligned with the requirements of the region. For example, the training for GPs from urban areas should put an emphasis on the treatment of patients with psychiatric, social and cultural problems. Whereas rural GPs need advanced skills regarding the care for children or incompliant patients. Generally, GPs from all regions should be better prepared to address the problems with the worst outcomes, because the

differences in the frequencies of topics like psychiatric disorders, poor therapy adherence, hypochondria or drug abuse could also mean that these problems are less talked about or less identified in rural areas. Adjusting the training of GPs accordingly could facilitate a better response to these regional challenges in health care."

Point-by-point response Reviewer 3:
Reviewer Name: Soumya Mazumdar
Institution and Country: SWSLHD, Australia

The authors have done a detailed qualitative analysis on patients within a group of GPs in German. I believe this is an interesting study, and I had a few comments:
1. The term "environs" in line 58 is a bit strange. Is it something that the Federal Institute of Research on Building… have used, or is it something that the authors have decided to use. In the context the authors use this term it seems to me that they mean peri urban areas.
Answer: The term was defined and used in the former publications (see references 3, 8, 16 and 17). For that reason we would like to keep it. "Environs", by our definition, include urbanised districts and rural districts with signs of agglomeration.

2. Line 60: The sentence may be better phrased as "with signs of urban agglomeration"
Answer: We added your suggestion.

3. Do each of the categories in lines 51-60 have a population density band/threshold? This would be especially useful if researchers for other countries/jurisdictions/contexts were trying to do a comparison study.
Answer: Thank you for this comment. We added the corresponding population densities and revised the section as following: "The category "urban areas" included independent large cities constituting districts in their own right (over 100,000 inhabitants), the category "environs" urbanised districts (with a density of over 300 inhabitants/km²) and rural districts with signs of urban agglomeration (with a density of over 150 inhabitants/km²) and the category "rural areas" sparsely populated rural districts (with a density of less than 150 inhabitants/km²)."

4. Recruitment: Lines 19 to 47: More details of the survey parameters would be useful. What was the survey response rates? Did the authors gauge non response bias? Could the non-responding GPs have been different from the ones that responded?
Answer: A detailed description of the recruitment process and the stratification of groups can be found in Schäfer et al. 2020 (including additional files). We added a sentence about the response rates in the limitations as following: "We have to contact a high number of 4956 GPs which revealed a comparatively low participation rate of 4.3% interviewed GPs. In Quota sampling the participation rate is not important, however, it may still affect the representativeness of the GP population."

5. Lines 51-60: Data Collection: Could there have been changes to the study populations between 2014 (quite a while back) and now? For instance, do any of the study areas have large concentrations of Syrian refugees, and that would have happened after 2014?
Answer: We added a short sentence about this topic in the strenghts and limitations as following: "The identification of the patient types took place before the European refugee crisis in Germany arrived. The measurement of the frequencies of the patient types was carried out during this period (2015-2017). It can be assumed that the refugees have only slowly integrated into the general practices [22]. Therefore, the patient types "patients with migration background and culturally different disease concepts" and "patients with migration background and communication problems" could be nowadays found more frequently in general practices. It could also affect other patient types like "patients with psychiatric problems", which are frequently found in the refugee population [23]."

6. Was the study approved by an Ethics committee/Institutional Review Board?
Answer: Yes. The study was approved by the Ethics Commission of the Hamburg Medical Association on 12 August 2013 (file number PV 4535). All the authors commented on the draft and read and approved the final version of the manuscript. (s. declarations)

7. Data Analysis: Lines 48-60: Is there a specific reason why the authors chose to use two logistic regressions instead of one multivariate logistic regression?
Answer: We had to choose two logistic regression analyses, because the regional categories are the dependent variable in our analyses and we wanted to compare urban vs. rural areas and environs vs. rural areas. This is already described in the statistical analyses section.

8. Results, line 14: The mean GP age is quite high... is this normal for Germany? A line or two on this may be useful to the international reader.
Answer: We discussed this problem already in the publication by Schäfer et al. 2020. To gather some information on a possible selection bias, we performed a comparison of the data of study participants in the included regions with the statistics of the German national association of statutory health insurance physicians. This analysis showed, with one exception, only relatively minor deviations. Thus, GPs participating in our study had only been slightly older (urban areas: + 0.9 years; environs: + 0.4 years; rural areas: + 0.6 years) and slightly more often males than the basic study population of the selected districts (urban areas: + 3.6%; rural areas: + 3.6%). (s. KBV (Kassenärztliche Bundesvereinigung). Gesundheitsdaten: Regionale Verteilung der Ärzte in der vertragsärztlichen Versorgung. http://gesundheitsdaten.kbv.de/cms/html/16402.php. Accessed 15 Jan 2019.)
We added a sentence in the limitations as following: “Furthermore, we performed a comparison of the data of study participants in the included regions with the statistics of the German national association of statutory health insurance physicians [21]. GPs participating in our study had only been slightly older (urban areas: + 0.9 years; environs: + 0.4 years; rural areas: + 0.6 years) and slightly more often males than the basic study population of the selected dis-tricts (urban areas: + 3.6%; rural areas: + 3.6%).”

9. Results: Sample characteristics: This may be the place to discuss response rates and nonresponder characteristics that I mentioned in 4.
Answer: We discussed this in the limitations (s. our answer to your comment no. 4 and 8.)

10. Page 9: Line 42-43: Please check the grammar of this sentence, “the need to ward of desires…
Answer: We changed the sentence in “For some GPs caring for this patient group is stressful due to frequent and time consuming consultations. In addition, requests for prescriptions often have to be refused.”.

11. Page 9: Line 56: Please change bad to poor and condition to conditions.
Answer: We changed this as suggested.

12. The qualitative coding is very interesting.
13. The discussion section is detailed and generally fine

Thank you in advance for editorial consideration of our manuscript and we look forward to your decision regarding the suitability of this work for publication in BMJ Open.

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Sean Lyons<br>Economic and Social Research Institute, Ireland |
| **REVIEW RETURNED** | 20-Oct-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed all the queries I had on the paper. I am happy to recommend it for publication. |

| | |
|---|---|
| **REVIEWER** | Professor D'Arcy Holman<br>School of Population and Global Health<br>The University of Western Australia<br>Australia |
| **REVIEW RETURNED** | 15-Oct-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for the comprehensive responses to the issues raised by this reviewer and the other reviewers and the many improvements that you have made to your manuscript on the basis of these reviews. I have no further concerns and hope to see your work in print. |