PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of primary care availability on antibiotic consumption in Hungary: a population based panel study using unfilled general practices
AUTHORS	Biro, Aniko; Elek, Peter

VERSION 1 - REVIEW

REVIEWER	Yingfen Hsia Paediatric Infectious Diseases Research Group St. Georges University of London UK
REVIEW RETURNED	20-Jan-2019

GENERAL COMMENTS	1.This paper needs inputs from clinicians (i.e. infectious disease consultants) and policy makers in Hungary. Two researchers are excellent mathematicans and carreid out completed statistically modelings. However, the paper is lack of discussiion on how their resutls related to clinical practices and policy changing in Hungary. 2.I would suggest to shorten the statitical analysis in method section. The definition on narow-spectrum and broad-specrum antibioitcs should be in method not presented as footnote. Also, the incusion and exclusion on diagnoses (footnote 7) need to be in method section. 3.In 2017, the WHO Essential Mediine List working group released a new classification to group antibioitcs into: Access, Watch, and Reserve (https://www.who.int/medicines/news/2017/20th_essential_med-list/en/). This new AWaRe classification is intended to optimise antibiotic use (more Access antibiotics) and reserve the consumption of broad-spectrum antibioitcs. Authors should consider to use this new WHO classification to definte narrow-, and broad-specturm antibioitcs for their analyses if possible. 4.Authors should acknowledge the strengths and limitations of their works in the maniscript. 5.The conclusion should be strengthen. What's your suggestions and/or furutre works based on your resutls? 6.I would suggest authors to revised their maniscript and submit it to infectious diseases related journals as you can reach more appropriate readers.

REVIEWER	Huan ZHOU Sichuan University, China
REVIEW RETURNED	14-Feb-2019

GENERAL COMMENTS	This paper stated an interested topic, i.e. The effect of primary care availability on antibiotic consumption. However, some shortcomings cannot be accepted for publication. 1. It is not clear for “unfilled position”,”unfilled practices”and “unfilled general practices”, which are key words in this paper 2. In “Results” of Abstract, “public expenditures on antibiotics decrease by 4.1% (95% CI 1.6% to 6.4%)” didn’t consistent with the results section in the text (P16 Lin18 and Table 1). 3. In the background section, it is not clear to mention research purposes. 4. No discussion section Conclusion didn’t justify the main results. It should be brief and to the point.

REVIEWER	Roy Sabo Virginia Commonwealth University, USA
REVIEW RETURNED	23-Apr-2019

GENERAL COMMENTS	Specific Comments: 1. Methods Section, 2.1.1 Unfilled Practices: While in other countries using unfilled positions to represent Primary Care Access would have low construct validity, the manner in which the authors restrict their sample to Settlements with 1 GP makes this measure better. It certainly does not have high validity, but it has some. The remaining problems with this are (i) that this restriction to Settlements with 1 GP biases the sample (as the authors point out) to rural or village settings, which could affect the overall relationship the authors are investigating, and (ii) the authors do not account for the distance to or general availability of other GPs when a position is unfilled. 2. Methods Section, 2.1.1 Unfilled Practices: The statement "The majority of unfilled general practices are in the countryside..." is not cited or supported. 3. Methods Section, 2.1.1 (last paragraph): These are descriptions of the data and belong in the Results Section. 4. Methods Section, 2.1.2, last two paragraphs: These are descriptions of the data and belong in the Results Section. 5. Methods Section, 2.1.5, last sentence: goes in Results Section. 6. Methods Section, 2.2 Statistical Analysis: The models (1-3) are unclear and contain non-descriptive and unconventional terms. In particular, the "h" and "v" terms are not helpful to the reader and do not convey what they are supposed to represent. The author's description in text of what these terms are is likewise confusing and unclear. For instance, "various settlement-level indicators" is used for "h"; this needs to be a non-ambiguous list of concrete terms. It is not clear what an "age-gender distribution", how it is used as a proxy for health needs, or under what conditions it

	would be appropriate to do so. The authors use a fixed-effect for Settlements when the field of statistics is nearly unanimous in calling for a random effect in cases like this. The authors do not account for spatial dependence, which is remarkably important since geography is the underlying construct they are measuring here. 7. Results Section, 3.1.1: Inferential results are presented before summary statistics. It is not always clear where the authors get their "effects" from, as the Tables 1-3 are not presented in any standard manner, and are even missing descriptive column headings in some cases (Tables 2 and 4). It is not clear which table the results from the second paragraph on page 18 are coming from. 8. Conclusions: The authors do not contextualize their findings within the relevant literature base. The authors also do not discuss their strengths and limitations.

REVIEWER	Howard Cabral Boston University Boston, Massachusetts USA
REVIEW RETURNED	08-May-2019

GENERAL COMMENTS	This manuscript aims to examine the association of unfilled primary care positions in Hungary from 2010-2015 with antibiotic consumption. The study was approved by the Research Ethics Committee of the Hungarian Research Council. In general, the paper is well written though the framework supporting the link of unfilled positions to antibiotic consumption is not entirely well developed but is suggestive. There are some points of concern in the manuscript that need to be addressed. 1. The authors state that there is geographical and time variation in the data set. The statistical models as specified in the text do not note that potential clustering was accounted for in the analyses, though such is noted in the tables. How the clustering was taken into account in the robust standard errors should be explained in the methods text as well as in each table. 2. In the Abstract, the finding of a 3.2% decrease should perhaps be stated as a 3.2% decrease on average given that it was obtained from a fixed effect in a population-average model. 3. Antibiotic assumption is assumed to follow a Gaussian distribution in the statistical analyses. Analytic results supporting this assumption should be provided. 4. The authors state that the ratio of broad-spectrum antibiotics is “probably responsible” for emerging antibiotic resistance. References should be supplied to support this statement. 5. The main independent variable is not formulated as the number of providers per number of patients in the population. Why not? This is a common metric for judging coverage of care. The authors should explain why using the population sizes as weights would address this concern. The heteroscedasticity noted should be shown also. Moreover, in one of the models, there is an interaction that includes population size. How is this implemented given that the variable is also a weight?

	6. Given that multiple years of data were analyzed, it is good that the authors included a factor of time in their models. In general, though, more detail should be provided regarding how each variable is configured for the regression analysis as well as identifying the symbols used for the regression coefficients in each model. 7. The intercept coefficient in model (2) at the bottom of page 14 should have a subscript of 0. 8. There is no text in the Discussion on the limitations of the study. Why not? For example, it seems that there could be uncontrolled confounding given the limitations of the data set. 9. The sample size in each analysis should be stated as a footnote in each table.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Yingfen Hsia
Institution and Country: Paediatric Infectious Diseases Research Group, St. Georges University of London UK

1.This paper needs inputs from clinicians (i.e. infectious disease consultants) and policy makers in Hungary. Two researchers are excellent mathematicians and carried out completed statistically modelings. However, the paper is lack of discussion on how their results related to clinical practices and policy changing in Hungary.

Thank you for pointing out this limitation. During the revision process, we put an effort into strengthening the discussion section, e.g. by discussing the institutional frameworks (types of primary care systems) where our results are likely to be generalisable.

2. I would suggest to shorten the statistical analysis in method section. The definition on narrow-spectrum and broad-spectrum antibiotics should be in method not presented as footnote. Also, the inclusion and exclusion on diagnoses (footnote 7) need to be in method section.

The definition of narrow- and broad-spectrum antibiotics is now provided in the main text of section 2.1.3. The previous footnote 7 is no longer included in the text because we omitted the results related to urinary tract infections and respiratory tract infections from the revised version (see our response to the next comment). To address the comments raised by reviewers 2 and 3, we could not shorten the statistical analysis section substantially.

3. In 2017, the WHO Essential Medicine List working group released a new classification to group antibiotics into: Access, Watch, and Reserve (https://www.who.int/medicines/news/2017/20th_essential_med-list/en/). This new AWaRe classification is intended to optimise antibiotic use (more Access antibiotics) and reserve the consumption of broad-spectrum antibiotics. Authors should consider to use this new WHO classification to define narrow-, and broad-spectrum antibiotics for their analyses if possible.

Thank you for this recommendation, we extended the analysis with the AWaRE classification. These results have been added to Table 3, and are discussed in section 3.4. At the same time, to keep the size of Table 3 within reasonable limits, we omitted the results related to urinary tract infections and respiratory tract infections from the revised version.

4. Authors should acknowledge the strengths and limitations of their works in the manuscript.

The strengths and limitations are now presented in the revised Discussion section.

5. The conclusion should be strengthened. What's your suggestions and/or future works based on your results?

While we decided not to include a separate conclusions section in the manuscript, these questions are now addressed in the last two paragraphs of the revised Discussion section.

6. I would suggest authors to revise their manuscript and submit it to infectious diseases related journals as you can reach more appropriate readers.

We think that the implications of primary care availability can potentially be of interest for a broad range of readers, and publishing in BMJ Open ensures that our manuscript reaches a wide audience.

Reviewer: 2
Reviewer Name: Huan ZHOU
Institution and Country: Sichuan University, China

1. It is not clear for "unfilled position", "unfilled practices" and "unfilled general practices", which are key words in this paper.

We clarified the terminology, and use "unfilled general practices" as key words throughout the paper.

2. In "Results" of Abstract, "public expenditures on antibiotics decrease by 4.1% (95% CI 1.6% to 6.4%)" didn't consistent with the results section in the text (P16 Lin18 and Table 1).

Table 1 has been made more transparent, now the first line of column (3) in Table 1 reflects the estimated effect of unfilled general practices on antibiotic expenditures.

3. In the background section, it is not clear to mention research purposes.

The title of section 1 has been changed to Introduction, to better reflect its contents. Institutional background and information on the consumption of antibiotics in Hungary are now presented in two separate subsections, 1.1 and 1.2.

4. No discussion section

A detailed discussion section is now included in the manuscript.

5. Conclusion didn't justify the main results. It should be brief and to the point.

To address this concern, we implemented two modifications. First, the Conclusions part of the Abstract has been substantially shortened. Second, the title of the last section of the manuscript has been changed to Discussion, to be in line with its content. In this section, the last paragraph briefly concludes the manuscript, based on the main results.

Reviewer: 3
Reviewer Name: Roy Sabo
Institution and Country: Virginia Commonwealth University, USA

Specific Comments:
1. Methods Section, 2.1.1 Unfilled Practices: While in other countries using unfilled positions to represent Primary Care Access would have low construct validity, the manner in which the authors restrict their sample to Settlements with 1 GP makes this measure better. It certainly does not have high validity, but it has some. The remaining problems with this are (i) that this restriction to Settlements with 1 GP biases the sample (as the authors point out) to rural or village settings, which could affect the overall relationship the authors are investigating, and (ii) the authors do not account for the distance to or general availability of other GPs when a position is unfilled.

The first two paragraphs of section 2.1.1 have been modified to make the sample restrictions clearer. We restrict the sample to villages, but keep villages that have more than 1 GPs. (As noted in the text, 77% of the villages are covered by only one general practice.) The results are robust to restricting the sample to villages that are covered by only one general practice.

In the last paragraph of section 1.1, we explain that if a general practice becomes unfilled then deputy GPs provide some care there. Unfortunately, we do not observe the opening hours of the general practices. If a general practice is unfilled then patients might seek care at other healthcare providers (e.g. specialist care or GPs of other districts). As an indicator for the availability of other healthcare services, we include the local specialist outpatient capacities (annual specialist outpatient hours per 1,000 inhabitants) as a control variable in our models.

Also, in the 3rd paragraph of the Discussion section we discuss the findings if the number of providers per number of inhabitants is used as the main independent variable (showing positive effect of better availability on antibiotic use). In the same paragraph, we also explain why we use our preferred measure (ratio of unfilled practices) as the main independent variable.

2. Methods Section, 2.1.1 Unfilled Practices: The statement "The majority of unfilled general practices are in the countryside..." is not cited or supported.

We now provide supporting statistics as follows: “the lack of GP care availability is typically an issue in villages: although only 30% of the population live there, 81% of unfilled general practices are located there”.

3. Methods Section, 2.1.1 (last paragraph): These are descriptions of the data and belong in the Results Section.

These results have been moved to section 3.1.1 (2nd paragraph).

4. Methods Section, 2.1.2, last two paragraphs: These are descriptions of the data and belong in the Results Section.

These results have been moved to section 3.1.2.

5. Methods Section, 2.1.5, last sentence: goes in Results Section.

This sentence has been moved to the end of section 3.1.1.

6. Methods Section, 2.2 Statistical Analysis: The models (1-3) are unclear and contain non-descriptive and unconventional terms. In particular, the "h" and "v" terms are not helpful to the reader and do not convey what they are supposed to represent. The author's description in text of what these terms are is likewise confusing and unclear. For instance, "various settlement-level indicators" is used for "h";

this needs to be a non-ambiguous list of concrete terms. It is not clear what an "age-gender distribution", how it is used as a proxy for health needs, or under what conditions it would be appropriate to do so. The authors use a fixed-effect for Settlements when the field of statistics is nearly unanimous in calling for a random effect in cases like this. The authors do not account for spatial dependence, which is remarkably important since geography is the underlying construct they are measuring here.

Thank you for these remarks, we address them in the following way.

The paragraphs below equation (1) have been substantially revised to explain the factors that are included in terms “h” and “v”, and why those terms are included in the model. Specifically, the variables of age-gender distribution are now defined more precisely and their role in affecting health needs is discussed.

In the 3rd paragraph of section 2.2.1 we argue why the fixed-effects specification is our preferred specification in this specific setting, in line with the econometric literature (see e.g. Wooldridge 2010). However, we note that the coefficient of the ratio of unfilled general practices is robust to the choice between fixed-effects and random-effects specifications. For instance, if the outcome variable is the logarithmic per capita antibiotic DOT, the random-effects specification gives a coefficient of -0.031, as opposed to the fixed-effects coefficient of -0.033 (SE 0.012), so the estimates are practically identical. We point out this robustness property of our results in the 2nd paragraph of the Discussion section.

As we now state in the last sentence of section 2.2.1, we cluster standard errors on the micro-regional level to take into account potential interdependencies (spatial dependence) in the local health care systems.

7. Results Section, 3.1.1: Inferential results are presented before summary statistics. It is not always clear where the authors get their "effects" from, as the Tables 1-3 are not presented in any standard manner, and are even missing descriptive column headings in some cases (Tables 2 and 4). It is not clear which table the results from the second paragraph on page 18 are coming from.

The tables have been restructured to a standard format. Also, to improve readability, their contents have been reduced by presenting only the results for the logarithmic outcomes (apart from the DOT per 1,000 inhabitant days in Table 1).

8. Conclusions: The authors do not contextualize their findings within the relevant literature base. The authors also do not discuss their strengths and limitations.

We have changed the title of the Conclusions section to Discussion. In addition to a brief literature review in the Introduction, the new Discussion section links our findings to the relevant literature (4th paragraph), and discusses the strengths (2nd and 3rd paragraphs) and limitations (5th paragraph).

Reviewer: 4
Reviewer Name: Howard Cabral
Institution and Country: Boston University, Boston, Massachusetts, USA

1. The authors state that there is geographical and time variation in the data set. The statistical models as specified in the text do not note that potential clustering was accounted for in the analyses, though such is noted in the tables. How the clustering was taken into account in the robust standard errors should be explained in the methods text as well as in each table.

Thank you for this remark. The last sentence of section 2.2.1 now explains how clustering was done. Also, this information is now provided in the note of each table.

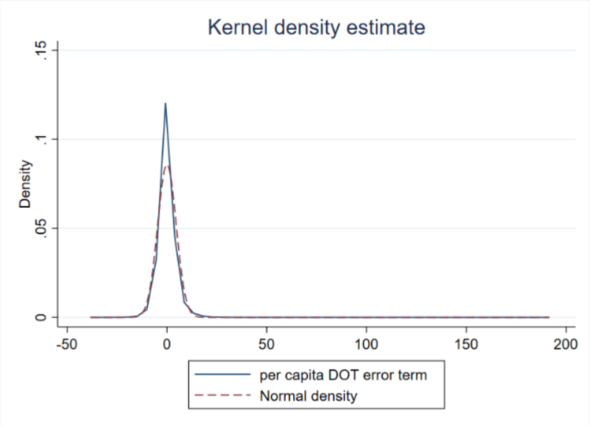
2. In the Abstract, the finding of a 3.2% decrease should perhaps be stated as a 3.2% decrease on average given that it was obtained from a fixed effect in a population-average model.

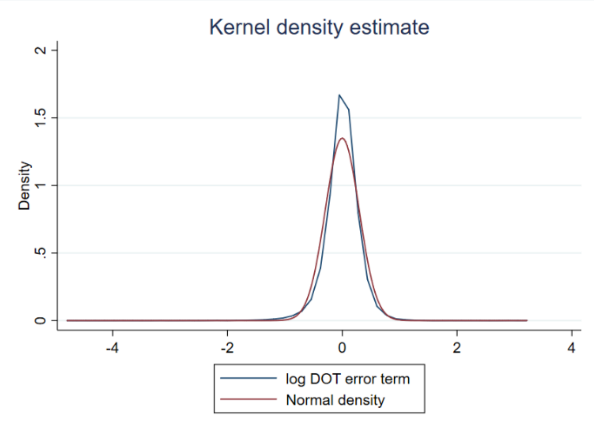
We now clearly indicate that these estimates are average effects.

3. Antibiotic assumption is assumed to follow a Gaussian distribution in the statistical analyses. Analytic results supporting this assumption should be provided.

Since our sample is reasonably large, the parameter estimates are approximately normally distributed according to the central limit theorem, thus the normal distribution of the error term is not needed for the validity of our results.

Here we provide density plots of the error term distributions in the models of antibiotic DOT per 1,000 inhabitant-days and of logarithmic DOT. The Shapiro-Wilk test rejects the null hypothesis of normal distribution in both cases, but the error terms are not far from normality in the logarithmic case.





4. The authors state that the ratio of broad-spectrum antibiotics is “probably responsible” for emerging antibiotic resistance. References should be supplied to support this statement.

References have been added. The statement has been rewritten: “Broad-spectrum antibiotics are known to have a high potential to cause antimicrobial resistance (WHO 2018, Melander et al. 2018).”

5. The main independent variable is not formulated as the number of providers per number of patients in the population. Why not? This is a common metric for judging coverage of care.
The authors should explain why using the population sizes as weights would address this concern. The heteroscedasticity noted should be shown also. Moreover, in one of the models, there is an interaction that includes population size. How is this implemented given that the variable is also a weight?

Thank you for these recommendations. Since our focus in this paper is on the implications of unfilled general practices, we prefer using an independent variable directly related to the occurrence of unfilled practices. However, in the 3rd paragraph of the Discussion section we discuss the findings if the number of providers per number of inhabitants is used as the main independent variable (showing positive effect of better availability on antibiotic use). In the same paragraph, we also explain why we use our preferred measure as the main independent variable.

To illustrate heteroscedasticity and the use of the weighted regression, we added Figures A4 and A5 to the Appendix (which we refer to in section 2.2.1).

In the interaction model, we analyse if the impact of primary care availability on antibiotic use varies by population size, which does not contradict using population size also as a weight.

6. Given that multiple years of data were analyzed, it is good that the authors included a factor of time in their models. In general, though, more detail should be provided regarding how each variable is configured for the regression analysis as well as identifying the symbols used for the regression coefficients in each model.

We made two steps to address this comment. First, the paragraphs below equation (1) have been substantially revised to explain what factors are included in terms “h” and “v”, and why those terms are included in the model. Second, we restructured the result tables (Tables 1-4), the full sets of coefficient estimates are now reported (apart from the monthly date effects and settlement fixed effects). Thus, the connection between equations 1-3 and Tables 1-4 is now more transparent.

7. The intercept coefficient in model (2) at the bottom of page 14 should have a subscript of 0.

Thank you for this remark. Subscript 0 was included also in the original version, maybe it did not appear correctly in the complied version. It might be an issue of different versions of Word – our document was written in Word 2016. This is how model (2) is specified:

$$\mathrm{antibiotics}_{s,2010} = \gamma_0^{2010} + \gamma_u^{2010}\mathrm{unfilled}_{s,2010} + \gamma_h^{2010}\mathrm{h}_{s,2010} + \varepsilon_{s,2010}$$

8. There is no text in the Discussion on the limitations of the study. Why not? For example, it seems that there could be uncontrolled confounding given the limitations of the data set.

Limitations are now discussed in the 5th paragraph of the Discussion section. Regarding the question of confounding, we note that settlement-level fixed effects control for all time-invariant settlement-level confounding variables.

9. The sample size in each analysis should be stated as a footnote in each table.

The sample sizes are now presented in table footnotes, except for Table 3, where the sample size varies by columns, therefore it is presented in the last row of that table.

VERSION 2 – REVIEW

REVIEWER	Roy T. Sabo Virginia Commonwealth University, USA
REVIEW RETURNED	26-Jun-2019

GENERAL COMMENTS	The authors have addressed some of my previous concerns, but others remain (to me) unaddressed, while new concerns have arisen. Major Recurring Concern: I had previously been critical of the authors' use modeling of settlements with fixed effects as opposed to random effects, stating the nearly unanimous standard that incidental factors (such as hospital, school, or in this case, geographic settlement area) are better modeled randomly, especially in cases with large numbers of levels, which is certainly the case here. The authors did not opt to change their model, but instead argue that the fixed-effect model can "capture all time-

	invariant characteristics of the settlements..." and "yields unbiased estimates for the parameters even if we allow that settlements with unfilled and filled practices may systematically differ from each other across dimensions that we cannot measures." The first claim is NOT germane to this research, as the settlement-level associations in question are in fact NOT time invariant. The second claim is dubious at best, as the fixed effect model in question, and the manner in which the authors have implemented it WITHOUT an interaction effect, means that the stated regression parameter estimates assume no geospatial interaction in the associations between filled/unfilled GPs and the outcomes. This is an enormously limiting assumption, which I brought up previously. Other concerns: 1. Strengths and Limitations of the Study: The authors state as a strength that they "identify causal effects of access to primary care...". Causal effects cannot be inferred from a retrospective study. 2. Section 2.1.3, first paragraph, last sentence: The authors claim that measuring overall prescriptions from both primary and specialist care is akin to "taking into account a possible substitution from PC to SC" when GP positions are unfilled. This overall measure DOES NOT take such a substitution into account; rather it confounds it entirely, making it impossible to see if it has occurred. 3. Section 2.2.1, last paragraph: The authors attempt to account for potential inter-dependencies in local health care systems by using "micro-regional level" standard errors. However, the process in which they have done so is not explained, nor are any citations provided. 4. Section 2.2.2: There is a reference that did not format correctly. 5. Section 2.2.2, last paragraph: "...and settlement-level controls (h_s) are teh same as in equation (2)." Should this refer to equation (1)? "Then, we re-estimate equation (2), with adding the interaction term..." Should this also refer to equation (1)? It makes no sense to refer to equation (2) in these cases as that is only the baseline 2010 model. 6. Section 3: There are nearly 2 pages of results relegated to Appendices. This seems extreme; if the material is this important, why isn't it in the main text? 7. Section 3.2: There is another unformatted citation. 8. Sections 3.2, 3.3 and 3.4: Nearly all of the outcomes modeled by the authors is on the log-scale, yet results are interpreted by absolute % change on the original scale. The authors need to re-calibrate these effect sizes or specifically mention Log-outcome for each instance. 9. Section 3.4, second paragraph, last sentence: the authors compare the effect of unfilled GPs on narrow-spectrum antibiotics and broad-spectrum antibiotics, but these estimates are from

	separate models and should NOT be compared, anecdotally or otherwise.

REVIEWER	Howard Cabral Boston University United States of America
REVIEW RETURNED	18-Jun-2019

GENERAL COMMENTS	This manuscript aims to examine the association of unfilled primary care positions in Hungary from 2010-2015 with antibiotic consumption. There are some points of concern in the manuscript that remain to be addressed after the response by the authors, which in general was complete. 1. Previous critique: “Antibiotic assumption is assumed to follow a Gaussian distribution in the statistical analyses. Analytic results supporting this assumption should be provided.” The authors note correctly that the approximate Gaussian distribution applies to the coefficients of their models. This is the case for models of non-Gaussian outcomes also. The previous comment related to the appropriateness of the use of a linear model. The authors’ response did not completely address this concern. 2. Previous critique: “The main independent variable is not formulated as the number of providers per number of patients in the population. Why not? This is a common metric for judging coverage of care. The authors should explain why using the population sizes as weights would address this concern. The heteroscedasticity noted should be shown also. Moreover, in one of the models, there is an interaction that includes population size. How is this implemented given that the variable is also a weight?” The authors responded that they felt it appropriate to include the population size as a fixed effect in their interaction model as well as a weight. This response is not all that clear. Why, then for example, would there be an offset in Poisson regression models that model such an effect but without a coefficient assigned to it? Providing a reference or other technical justification for using a variable as a predictor and as a weight in the same model would be of value here. 3. Previous critique: “There is no text in the Discussion on the limitations of the study. Why not? For example, it seems that there could be uncontrolled confounding given the limitations of the data set.” The authors noted that their models included all of the confounders that they could measure. This is exactly the point. That is, there could be other confounders relevant to their study question that were not available in their data set. Some mention of this and the likely variables that would have been informative would be useful here.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 4
Reviewer Name: Howard Cabral

1. Previous critique: “Antibiotic assumption is assumed to follow a Gaussian distribution in the statistical analyses. Analytic results supporting this assumption should be provided.”
The authors note correctly that the approximate Gaussian distribution applies to the coefficients of their models. This is the case for models of non-Gaussian outcomes also. The previous comment related to the appropriateness of the use of a linear model. The authors’ response did not completely address this concern.

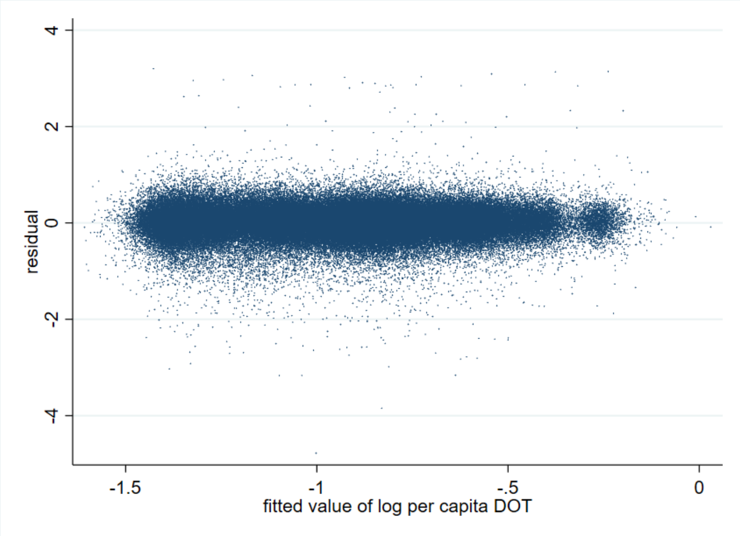
Thank you for this remark. It was not clear to us that the previous comment related to the use of a linear model. To address this concern, we took the following two steps.

First, we now present that our results are robust to a nonlinear specification. The estimation results are virtually unaffected if the continuous indicator of the ratio of unfilled practices is replaced with a categorical variable that takes three distinct values: it equals 0 if the ratio of unfilled practices is zero; equals 1 if the ratio of unfilled practices is between zero and one; and equals 2 if the ratio of unfilled practices is one. In other words, here we distinguish the settlements where no practices are unfilled; where all practices are unfilled; and where some practices are filled and some are unfilled. Additionally, to check for nonlinearities, we include the squared terms of all control variables in the model.

According to the parameter estimates in this specification, if all GP practices (typically the single available GP practice) become unfilled, that implies a 3.7% lower per capita antibiotic DOT (95% CI: 1.5% to 5.8%) compared to the “none unfilled” category, which is very close to the estimate of the linear specification presented in Table 2 of the paper (3.2%). Meanwhile, if there are more than one GP practices and some but not all of them become unfilled (middle category), that has no significant effect on antibiotic consumption (effect size is -0.5%, 95% CI: -2.5% to +1.4%).

Second (not presented in the paper), as a diagnostic tool for the linearity assumption, we plot the residuals versus the fitted values, based on the fixed effects model of logarithmic per capita antibiotic DOT. The plot does not indicate important nonlinearities in the model.

All in all, nonlinearities do not affect substantially the estimated effect of unfilled general practices on antibiotic consumption. For simplicity and easy interpretability, we present the numerical results of the linear model in the paper and refer to the nonlinear results as a robustness check.



2. Previous critique: “The main independent variable is not formulated as the number of providers per number of patients in the population. Why not? This is a common metric for judging coverage of care. The authors should explain why using the population sizes as weights would address this concern. The heteroscedasticity noted should be shown also. Moreover, in one of the models, there is an interaction that includes population size. How is this implemented given that the variable is also a weight?”

The authors responded that they felt it appropriate to include the population size as a fixed effect in their interaction model as well as a weight. This response is not all that clear.
Why, then for example, would there be an offset in Poisson regression models that model such an effect but without a coefficient assigned to it? Providing a reference or other technical justification for using a variable as a predictor and as a weight in the same model would be of value here.

We appreciate the concerns of the Reviewer, thus we decided to omit the interaction term between the indicator of unfilled practices and population size. We still present the interactions with the indicators of healthcare supply and socio-economic indicators. Section 3.3 has been rewritten accordingly.

3. Previous critique: “There is no text in the Discussion on the limitations of the study. Why not? For example, it seems that there could be uncontrolled confounding given the limitations of the data set.”
The authors noted that their models included all of the confounders that they could measure. This is exactly the point. That is, there could be other confounders relevant to their study question that were not available in their data set. Some mention of this and the likely variables that would have been informative would be useful here.

Thank you for this remark, we inserted the following sentence to the discussion of limitations: “There could still remain unobserved time-varying confounders that are related both to antibiotic use and the availability of GP care (such as changing population health status and health behaviour).”

Reviewer: 3
Reviewer Name: Roy T. Sabo

Major Recurring Concern: I had previously been critical of the authors' use modeling of settlements with fixed effects as opposed to random effects, stating the nearly unanimous standard that incidental factors (such as hospital, school, or in this case, geographic settlement area) are better modeled randomly, especially in cases with large numbers of levels, which is certainly the case here. The authors did not opt to change their model, but instead argue that the fixed-effect model can "capture all time-invariant characteristics of the settlements..." and "yields unbiased estimates for the parameters even if we allow that settlements with unfilled and filled practices may systematically differ from each other across dimensions that we cannot measures." The first claim is NOT germane to this research, as the settlement-level associations in question are in fact NOT time invariant. The second claim is dubious at best, as the fixed effect model in question, and the manner in which the authors have implemented it WITHOUT an interaction effect, means that the stated regression parameter estimates assume no geospatial interaction in the associations between filled/unfilled GPs and the outcomes. This is an enormously limiting assumption, which I brought up previously.

Thank you for these detailed comments. We acknowledge that different fields view the decision between the fixed and the random effects specification slightly differently. Here we follow the econometric practice (e.g. Wooldrige, 2010, sections 10.4-10.5). In the next paragraphs we outline why we present the fixed effects model as our baseline specification. However, in light of the Reviewer’s comments, we think that the best approach is to show that the estimates are robust in this respect, therefore we now include the estimates from both specifications in Table 2 and have also made some other changes, see below. Also, the other three Reviewers did not raise objections against the fixed effects formulation, so we did not think it appropriate to change the baseline specification at this stage.

Since some unobserved time-invariant characteristics of the settlements may be related to the unfilled status of the general practice and to the other regressors, the random effects specification may be biased, and the econometric literature takes this possible bias seriously. In our case, a formal Hausman-type test (implemented with the xtoverid function of the Stata statistical package) prefers the fixed effects model over the random-effects one at even the 0.1% significance level.

At the same time, we agree with the Reviewer that it is advisable to present the random effects specification as well. Therefore, we changed the text as follows. Most importantly, we extended the table of main results (Table 2) with the random effects estimates. The table indicates that the main parameter of interest (coefficient of the indicator of unfilled practices) is robust to the choice between the fixed and random effects specifications, while the parameter estimates of the control variables may differ more.

Also, we slightly modified the paragraph describing the fixed effects model in section 2.2.1, pointing out that unbiasedness holds only if the unobserved confounders are time-invariant. In the second to last paragraph of section 4 (Discussion), we acknowledge among the limitations that some unobserved time-varying confounders might still remain (with examples in the text), despite the richness of the data we use.

The Reviewer also raised that there are no geospatial interactions in our models. Due to the lack of available data on e.g. where the deputy GP comes from, we cannot include these interactions in the models. Some deputy GPs come from neighbouring settlements, others from the micro-regional centre (where a few GPs operate). Hence, on one hand, simple geographical distance may not be a good measure of the interaction, and on the other hand, the micro-regional level is definitely too large to examine the interaction. We briefly mention this as a limitation in the Discussion section.

Other concerns:

1. Strengths and Limitations of the Study: The authors state as a strength that they "identify causal effects of access to primary care...". Causal effects cannot be inferred from a retrospective study.

Thank you for this remark, we rewrote that statement as “[t]o estimate the effects of access to primary care on the consumption of antibiotics, we use geographical and time variation in unfilled general practices in Hungary.”

2. Section 2.1.3, first paragraph, last sentence: The authors claim that measuring overall prescriptions from both primary and specialist care is akin to "taking into account a possible substitution from PC to SC" when GP positions are unfilled. This overall measure DOES NOT take such a substitution into account; rather it confounds it entirely, making it impossible to see if it has occurred.

Thank you for this remark. We intended to express with the original sentence that the combined effect was estimated. We removed the second part of the original sentence.

3. Section 2.2.1, last paragraph: The authors attempt to account for potential inter-dependencies in local health care systems by using "micro-regional level" standard errors. However, the process in which they have done so is not explained, nor are any citations provided.

Thank you for this remark, we inserted references to this part of the text.

4. Section 2.2.2: There is a reference that did not format correctly.

We apologise for this problem, we now typed the equation number in the text (instead of using a cross-reference).

5. Section 2.2.2, last paragraph: "...and settlement-level controls (h_s) are the same as in equation (2)." Should this refer to equation (1)? "Then, we re-estimate equation (2), with adding the interaction term..." Should this also refer to equation (1)? It makes no sense to refer to equation (2) in these cases as that is only the baseline 2010 model.

We apologise for this confusion, unfortunately we had problems with cross-references to the equation numbers. We now typed the equation numbers in the text to avoid such problems.

6. Section 3: There are nearly 2 pages of results relegated to Appendices. This seems extreme; if the material is this important, why isn't it in the main text?

We moved the previous Table A1 to the main text (now Table 1). Also, we removed the previous Appendix A3 from the appendix because its content is sufficiently explained in the text. We think the remaining content of the appendix is important (partly following Reviewers’ recommendations), but we are reluctant to include those in the main text as the submission guidelines state that the article should have up to 5 tables and figures (which is now exceeded by one). If the editorial policy allows, we are willing to include the remaining content of the appendix in the main text.

7. Section 3.2: There is another unformatted citation.

Again, we apologise for this problem, we now typed the equation number in the text (instead of using a cross-reference).

8. Sections 3.2, 3.3 and 3.4: Nearly all of the outcomes modeled by the authors is on the log-scale, yet results are interpreted by absolute % change on the original scale. The authors need to re-calibrate these effect sizes or specifically mention Log-outcome for each instance.

Thank you for this remark, we have recalculated all results to % change but also mention the results on the log scale since the Tables contain the latter estimates. The differences between the % effects and the log effects are not large because the effects are small. We have rewritten footnote 7 accordingly.

9. Section 3.4, second paragraph, last sentence: the authors compare the effect of unfilled GPs on narrow-spectrum antibiotics and broad-spectrum antibiotics, but these estimates are from separate models and should NOT be compared, anecdotally or otherwise.

Thank you for this remark, this sentence has been rewritten to avoid such comparison.

VERSION 3 - REVIEW

REVIEWER	Roy T. Sabo Virginia Commonwealth University, USA
REVIEW RETURNED	17-Jul-2019

GENERAL COMMENTS	The authors have addressed my concerns.

REVIEWER	Howard Cabral Boston University United States of America
REVIEW RETURNED	16-Jul-2019

GENERAL COMMENTS	This manuscript aims to examine the association of unfilled primary care positions in Hungary from 2010-2015 with antibiotic consumption. There is only one remaining question regarding the re-formulated outcome variable, all other critiques having been addressed. The outcome variable has be re-configured as an ordered categorical variable (ordinal). The typical statistical approach to analyzing such a variable is to employ, if appropriate, a proportional odds logit model. Although such a ordered, categorical variable could yield linear regression coefficients if employed in such a form, the question can be raised as to this being properly used in contrast to the logit model that could yield odds ratios. Why the authors did not choose to construct the logit model is not clear.

VERSION 3 – AUTHOR RESPONSE

Reviewer: 4

Reviewer Name: Howard Cabral

This manuscript aims to examine the association of unfilled primary care positions in Hungary from 2010-2015 with antibiotic consumption. There is only one remaining question regarding the re-formulated outcome variable, all other critiques having been addressed.

The outcome variable has be re-configured as an ordered categorical variable (ordinal). The typical statistical approach to analyzing such a variable is to employ, if appropriate, a proportional odds logit model. Although such a ordered, categorical variable could yield linear regression coefficients if employed in such a form, the question can be raised as to this being properly used in contrast to the logit model that could yield odds ratios. Why the authors did not choose to construct the logit model is not clear.

We thank the Reviewer for the thorough review of the latest version of the manuscript. Regarding the only remaining question, there may have been some misunderstanding. Not the outcome variable, but the main explanatory variable has been reformulated as a categorical variable in one of the specifications. The outcome variables are continuous, the same as in the previous versions. We have made this clearer in the current version with a minor modification.

VERSION 4 - REVIEW

REVIEWER	Howard Cabral Boston University United States of America
REVIEW RETURNED	08-Aug-2019

GENERAL COMMENTS	This manuscript aims to examine the association of unfilled primary care positions in Hungary from 2010-2015 with antibiotic consumption. The authors have now responded to all points of critique.