PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Factors Influencing Subspecialty Choice Among Medical Students: A Systematic Review and Meta-analysis
AUTHORS	Yang, Yahan; Li, Jiawei; Wu, Xiaohang; Wang, Jinghui; Li, Wangting; Zhu, Yi; Chen, Chuan; Lin, Haotian

VERSION 1 – REVIEW

REVIEWER	E. Benjamin Puertas Pan Amerinca Health Organization PAHO/WHO, El Salvador
REVIEW RETURNED	25-Feb-2018

GENERAL COMMENTS	The study is complete and covers the most important papers on the subject. It refers to sub specialties, and includes Family Medicine as such, although I would consider it a specialty.

REVIEWER	Emily Fletcher University of Exeter Medical School
REVIEW RETURNED	22-Mar-2018

GENERAL COMMENTS	Introduction Rephrase the first sentence as this does not make sense. Sentence 2 - it is not obvious to me how social economy and improved living standards leads to increased demand for physicians. Surely, this should be more to do with people living longer, with increased workload on doctors through increased number of consultations, and having to manage patients with multi-morbidity, complex long-term health conditions etc. Methods The fact that all studies reviewed were cross sectional means there is a lack of information about what happens in reality. Hypothetical choices during medical school will be subject to change until the postgraduate doctor actually applies for specialty training and it is research into the factors influencing the decision at this point (in addition to, or rather than, during medical school) which may shed light on how policy/intervention can address problems with workforce. Overall, this was an interesting paper and I enjoyed reading it, many thanks.

REVIEWER	Charlotte BEAUDART University of Liège, Belgium
REVIEW RETURNED	22-May-2018

GENERAL COMMENTS	BMJ Open
	Methods :
	- Authors should state at the beginning of the methods section that they are following the PRISMA guidelines.
	- Item n°5 of PRISMA guidelines refers to the presence of a registered or published a-priori-protocol. Authors should indicate the presence/absence of such a protocol.
	- The last search of the SR is October 2016. The reviewer is wondering about the relevance of an update of this search.
	- Authors are not clear regarding their methodology for study selection. Was the study selection performed by two reviewers independently? If yes, how did they resolve conflicts between both researchers?
	- An additional search is often recommended (ex. Grey literature, hand searching in bibliography of relevant studies, etc.)
	- Regarding search strategy:
	o Please indicate for which database this search strategy is adapted?
	o Even if it is too late, some points of the search strategy are not optimal. The reviewer is sceptical about the last term "factor" which appears to be very restrictive for this search. Moreover, authors did not use in an optimal way some "*" for example to replace a letter (ex: use "cross sectional stud*" to include "cross sectional study" but also "cross sectional studies)
	- Please provide a description of the NOS scale.
	- Q-test should be mentioned in the methods section.
	- For publication bias, authors are recommended to use also other methods such as "fail safe number" or "trim and fill method".
	- Table 1. Please provide additional information to let the reader understand that "score" means "quality score of the NOS scale".
	- Table 2. Please provide CI 95% for estimates and Q-test values in this table.
	- Table 3. Please provide the point estimate of the regression and the 95% CI
	- Table S2. Please provide CI 95% for the estimates and I².

REVIEWER	Clare Davenport
	University of Birmingham, UK
REVIEW RETURNED	27-May-2018

GENERAL COMMENTS	This research addresses an important and topical question and a thoughtful investigation using a systematic review are one way of investigating this. However the write up detracts from providing an accurate review of the quality of question formulation and review methods. I consider that the authors' conclusions are not supported by their findings and the research limitations, including the use of meta-regression to explore heterogeneity are not considered fully.
	More detailed comments are provided below:
	Methods:
	It is important to define your PICO or equivalent in your Methods section – for example medical students at what stage in their career; at least an example of some of the factors you were investigating (your exposure) and a distinction between individual and organisational factors (which you identify as a source of heterogeneity for subgroup analysis. In addition to demonstrating rigour in your systematic review methods an example of some of the factors you were investigating demonstrate a priori hypotheses and if included in your abstract will help dissemination of your findings for those looking for research on this topic.

You do not state why your commenced your search from 1977. Why did you only include peer reviewed publications as this strategy has the potential to introduce publication bias.

There is no mention of whether you applied a Language restriction.

You stated you searched the Cochrane library but you were looking for cross sectional studies. What did you expect to find in the Cochrane library?

The Newcastle Otowa Scale that you used for quality assessment is designed for case control and cohort studies, not cross sectional studies?

Analysis:

Generally this section is confusing with regard to the presence of a priori sources of heterogeneity and potential influencing factors and the difference between these. For example you state that you used meta-regression to identify possible sources of heterogeneity. Do you mean you used meta-regression to explore the effect of a priori defined potential sources of heterogeneity? If so these should be stated. For example what is the rationale for subgrouping studies conducted before and after 2010? Meta-regression should not be used to identify sources of heterogeneity not least because the method has limited power which is influenced by the number of studies available for each variable being considered. The potential result of using meta-regression to identify sources of heterogeneity is that results will be driven by the data available – for example the number of studies including any particular potential variable.

You state that you 'identified each factor when at least 5 papers were available'. A generally accepted rule is a minimum of 10 studies per factor being investigated in meta-regression.

Results: study characteristics

You describe potential factors influencing sub-speciality choice in your results section. If these are your pre-defined factors they should be introduced in the methods section.

Some information on the study designs employed would be helpful here for example how students were recruited; how influencing factors were identified / measured and techniques employed by included studies to investigate the association between influencing factor and sub-speciality choice.

Results: analysis:

When presenting subgroup analysis you present numerical results by world region and survey year and state other sources of heterogeneity investigated were non-significant. The results for world region are also non-significant.

Discussion / conclusions:

In general I consider that the firmness in which you present your conclusions is not supported by your results. There is considerable heterogeneity between studies that may mask associations. In addition the method of measurement used in studies and the potential error and bias inherent in transforming this into your outcome measure is a major limitation.

Some of the information in the discussion would have been more helpful in the background and results sections – for example.

I consider that the stage in a medical student's career is crucial to interpretation of many of the influencing factors measured, for example competencies. Indeed I would go so far as to say that this aspect of study design should have been investigated as a source of heterogeneity.

	In terms of further research would you consider that studies should be restricted by region? In terms of influencing policy then the most informative studies might be restricted by organisational and medical training factors; this will reduce heterogeneity and allow more meaningful investigation of other factors, for example there are large disparities in income between students in high income countries. Did you consider the value of qualitative research to further investigate your question?

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:
Comment: The study is complete and covers the most important papers on the subject. It refers to sub specialties, and includes Family Medicine as such, although I would consider it a specialty.
Response: We thank the reviewer's appreciation about our study. The relevant sections have been modified to make it more precise (Page 4, Line 75-77; Page 10, Line 239-240; Page 11, Line 257-263). We also added statements as follows: Studies were included if they reported data on medical students, were published in peer-reviewed journals, and used a validated method to assess the extent of a factor's influence on the choice of subspecialty, such as pediatric gastroenterology and vascular surgery, or its corresponding specialty, such as pediatrics and surgery. A guide to medical specialty, available at https://www.abms.org/member-boards/specialty-subspecialty-certificates/, was used to identify the medical specialty and subspecialty of our research. (Page 5, Line 117-119; Page 6, Line 120-128).

Reviewer #2:
Comment (1): Introduction
Rephrase the first sentence as this does not make sense.
Response: Thanks so much for your comments. The purpose of this sentence is to define the medical subspecialties in the research. We have rephrased this sentence and repositioned it in Page 5, Line 117-119 and Page 6, Line 120-121.

Comment (2): Sentence 2 - it is not obvious to me how social economy and improved living standards leads to increased demand for physicians. Surely, this should be more to do with people living longer, with increased workload on doctors through increased number of consultations, and having to manage patients with multi-morbidity, complex long-term health conditions etc.
Response: We have cited the reference in Page 17, Pine 383-390 and rephrased the sentence as follows: because of the population aging, increased workload on doctors through increased number of consultations and in managing patients with multi-morbidity, the demand for physicians continues to increase; however, an imbalance in the supply of physicians in different subspecialties has become a growing concern in both developed and developing countries. (Page 4, Line 71-72).

Comment (3): The fact that all studies reviewed were cross sectional means there is a lack of information about what happens in reality. Hypothetical choices during medical school will be subject to change until the postgraduate doctor actually applies for specialty training and it is research into the factors influencing the decision at this point (in addition to, or rather than, during medical school) which may shed light on how policy/intervention can address problems with workforce.
Response: Many thanks for your comments. We have included the cross-sectional studies because cohort studies on this topic are limited. We have rephrased the statements in the limitation section as

follows: first, the students involved in our study included medical students at different stages of their medical education. Students' perception about different subspecialties may change during medical training until the students applies for specialty training. For example, compared to an intern, a freshman student may place greater emphasis on income and prestige when considering a career choice. A subgroup analysis stratified by the stages of medical education and a secondary meta-analysis of longitudinal studies may better reflect changes in influencing factors and the extent of their influence over time (Page 14, Line 332-335).

Reviewer #3:

Comment (1): Authors should state at the beginning of the methods section that they are following the PRISMA guidelines.

Response: Many thanks for your comments. We have added the statement at the beginning of the methods section as follows: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was used to ensure the reporting quality of this review (Fig. S1) (Page 5, Line 108-110).

Comment (2): Item n°5 of PRISMA guidelines refers to the presence of a registered or published a-priori-protocol. Authors should indicate the presence/absence of such a protocol.

Response: We have added a statement in the methods section as follows: we developed a review protocol (registration number: PROSPERO CRD42017053781) prior to commencing the study. (Page 5, Line 107-108).

Comment (3): The last search of the SR is October 2016. The reviewer is wondering about the relevance of an update of this search.

Response: Thanks for your constructive suggestion. We have updated the search strategy and included three new studies in our analysis (Vo's study, reference 111; Grasreiner's study, reference 112; and Alkhannen's study, reference 113). The relevant sections have been modified (Page 2, Line 33-50; Page 3, Line 51-52; Page 5, Line 112; Page 7, Line 162-169; Page 8, Line 191; Page 9, Line 217-218; Page 10, Line 220-237; Page 11, Line 249-253; Page 12, Line 271-272; Page 22, Line 685-694; Details in Table 1, Table 2, Table 3, Table S1, Table S2, Figure 1, Figure 2, Figure S1, Figure S2, Figure S4, Figure S5, Figure S8, Figure S9, Figure S11, Figure S14).

Comment (4): Authors are not clear regarding their methodology for study selection. Was the study selection performed by two reviewers independently? If yes, how did they resolve conflicts between both researchers?

Response: We have added a statement as follows: a third investigator was consulted if disagreements occurred (Page 6, Line 136).

Comment (5): An additional search is often recommended (ex. Grey literature, hand searching in bibliography of relevant studies, etc.)

Response: We agree with the reviewer that an additional search is helpful for improving the quality of reporting of meta-analyses In this study, we have conducted the search using five databases manually. To ensure all relevant literatures were covered, we also conducted an additional search using OpenGrey. However, no additional articles were further included. We have included the information about OpenGrey in method section (Page 6. Line 129-130)

Comment (6): Please indicate for which database this search strategy is adapted?

Response: Thank you for your kind reminding. We have rephrased the statement as follows: a detailed example of search strategy for Medline/Pubmed is shown in Methods S1 (Page 5, Line 116-117).

Comment (7): Even if it is too late, some points of the search strategy are not optimal. The reviewer is sceptical about the last term "factor" which appears to be very restrictive for this search. Moreover, authors did not use in an optimal way some "*" for example to replace a letter (ex: use "cross sectional stud*" to include "cross sectional study" but also "cross sectional studies)

Response: In order to verify the reliability of the search strategy, we have searched the Medline/Pubmed using our original research strategy and using "cross sectional stud*" to include "cross sectional study" but also "cross sectional studies. The previous search result includes 306 items while the latter one includes 284 items.

Comment (8): Please provide a description of the NOS scale.

Response: Thanks for your suggestion. We made an inappropriate statement of our quality assessment tool in the previous manuscript. As is shown in the supporting information, an 11-item checklist rather than the NOS scale was actually used to assess the quality of the studies. We have reworded our statement and added a detailed description in the method section (Page 6, Line 137-140).

Comment (9): Q-test should be mentioned in the methods section.

Response: We have mentioned Q-test in the methods (Page 7, Line 153-155).

Comment (10): For publication bias, authors are recommended to use also other methods such as "fail safe number" or "trim and fill method".

Response: Many thanks for your comments. We have used two generally accepted tools, funnel plot test and Egger's test, to provide qualitative and quantitative measurement of publication bias, respectively. (Page 7, Line 157)

Comment (11): Table 1. Please provide additional information to let the reader understand that "score" means "quality score of the NOS scale".

Response: We have added a footnote in Table 1.

Comment (12): Table 2. Please provide CI 95% for estimates and Q-test values in this table.

Response: Thanks for your recommendation. We have added CI 95% for estimates in Table 2. Additionally, Q-test values of the meta-analyses analysis stratified by region and survey year are shown in Table S2.

Comment (13): Table 3. Please provide the point estimate of the regression and the 95% CI

Response: We have added CI 95% for estimates in Table 3.

Comment (14): Table S2. Please provide CI 95% for the estimates and I^2.

Response: We have added CI 95% for estimates and I^2 in Table S2.

Reviewer #4:

Comment (1): This research addresses an important and topical question and a thoughtful investigation using a systematic review are one way of investigating this. However the write up detracts from providing an accurate review of the quality of question formulation and review methods. I consider that the authors' conclusions are not supported by their findings and the research limitations, including the use of meta-regression to explore heterogeneity are not considered fully. More detailed comments are provided below:

Methods:

It is important to define your PICO or equivalent in your Methods section – for example medical students at what stage in their career; at least an example of some of the factors you were

investigating (your exposure) and a distinction between individual and organisational factors (which you identify as a source of heterogeneity for subgroup analysis.

Response: Thanks so much for your agreement on our manuscript and we really appreciate your constructive comments and suggestions. We agree that it is critical to define PICO. First, we have added statements to provide details of the medical students we recruited as follows: because of the differences between medical education systems in the world, the medical students we recruited includes the student in medical school, internship, residency training and fellowship, containing the students who about to make a specialty choice and students who has just made a specialty choice. (Page 6, Line 122-125). Second, there is actually no pre-defined factors in our research and we have provided brief descriptions of the factors which may cause confusion in the result section as follows: The influencing factors for subspecialty choice were then classified according to 17 aspects, including academic interests, controllable lifestyle or flexible work schedule (defined as flexibility that allows physicians to control the number of hours devoted to practicing the specialty), competencies, patient service orientation, medical teachers or mentors, career opportunities, workload or working hours (characterized by the physician's time spent on professional responsibilities), income, prestige, length of training, advice from others (advice from family, friends, and other students), student debt, experience with the subject, working environment, personality, gender and job security. (Page 7, Line 169; Page 8, Line 170-178). Third, "individual and organizational factors" haven't been studied in meta-regression, since country, survey years, specialty and sample size were identified as potential sources of heterogeneity (Page 9, Line 212-214).

Comment (2): In addition to demonstrating rigour in your systematic review methods an example of some of the factors you were investigating demonstrate a priori hypotheses and if included in your abstract will help dissemination of your findings for those looking for research on this topic.

Response: Names of all the factors we investigated in this study have been stated in the abstract (Page 2, Line 41-46). However, due to the word limit, we did not provide additional description of each factor.

Comment (3): You do not state why your commenced your search from 1977.

Response: Our search is not limited by publication date. Among the included studies, the first article was published in 1977. We have modified the method section to make the statements about the search strategy clearer (Page 5, Line 113; Page 7, Line 162-163).

Comment (4): Why did you only include peer reviewed publications as this strategy has the potential to introduce publication bias.

Response: We only included peer reviewed publications because the scientific quality of research published in peer reviewed journals is superior to that in non-peer reviewed publications. Besides, Egger's tests revealed that most of the publication bias in our research was small ($P>0.05$).

Comment (5): There is no mention of whether you applied a Language restriction.

Response: Thank you for your kind reminding. We have added a statement in method section as follows: we performed a literature search in June 2018 using the Cochrane Library, Medline, Web of Science, CNKI and ERIC databases without language restrictions (Page 5, Line112-113).

Comment (6): You stated you searched the Cochrane library but you were looking for cross sectional studies. What did you expect to find in the Cochrane library?

Response: As we shown in the supporting information, we searched systematic review and cross-sectional studies, and the Cochrane library was used to searched systematic review. However, cross-sectional studies were also included in our research. We have reworded the statements in manuscript to avoid the potential controversy (Page 5, Line 112-113; Page 7, Line 162-163).

Comment (7): The Newcastle Otowa Scale that you used for quality assessment is designed for case control and cohort studies, not cross sectional studies?

Response: Thanks for your constructive suggestion. We made inappropriate statements of our quality assessment tool. As was shown in the Table S1, an 11-item checklist for cross-sectional studies, rather than NOS scale, were used to assess the quality of the studies. Each item and the scores of each study are specified in Table S1. We have reworded our statement and added a detailed description of the ARHQ methodology checklist in the method section (Page 6, Line 137-140).

Comment (8): Generally this section is confusing with regard to the presence of a priori sources of heterogeneity and potential influencing factors and the difference between these. For example you state that you used meta-regression to identify possible sources of heterogeneity. Do you mean you used meta-regression to explore the effect of a priori defined potential sources of heterogeneity? If so these should be stated. For example what is the rationale for subgrouping studies conducted before and after 2010? Meta-regression should not be used to identify sources of heterogeneity not least because the method has limited power which is influenced by the number of studies available for each variable being considered. The potential result of using meta-regression to identify sources of heterogeneity is that results will be driven by the data available – for example the number of studies including any particular potential variable.

Response: We agree with the reviewer that meta-regression has limited power which is influenced by the number of studies available for each variety. However, meta-regression is generally used to understand heterogeneity (W. L. Baker, Int J Clin Pract. 2009), and we performed meta-regression to discern whether a linear relationship exists between EOI and the four covariates, exploring heterogeneity. Besides, we grouped studies conducted before or after 2010 to ensure the maximum number of studies available for research year.

Comment (9): You state that you 'identified each factor when at least 5 papers were available'. A generally accepted rule is a minimum of 10 studies per factor being investigated in meta-regression.

Response: Thanks for your suggestion. We agree with you and actually most of the factors being investigated contains over 10 studies. However, five papers are also acceptable. (Michael Borenstein, Introduction to meta-analysis, 2009; D. A. Mata, JAMA. 2015).

Comment (10): You describe potential factors influencing sub-speciality choice in your results section. If these are your pre-defined factors they should be introduced in the methods section.

Response: There is no pre-defined factors in our research. As we shown in the result section, among the included studies, the influencing factors were ranked according to the frequency of occurrence and each factor was identified when at least 5 papers were available. The top 17 influencing factors were identified, and personality, gender, experience with the subject, the working environment and job variety were then excluded for explained reasons. Thus, the analysis of the remaining 12 influencing factors were shown in this manuscript. We have reworded the statements to make it clear (Page 7, Line 169; Page 8, Line 170-183).

Comment (11): Some information on the study designs employed would be helpful here for example how students were recruited; how influencing factors were identified / measured and techniques employed by included studies to investigate the association between influencing factor and sub-speciality choice.

Response: Thank you for your kind reminding. First, we have added statements to provide details of the medical students we recruited (Page 6, Line 122-125). Second, we have reworded the result section to indicate how influencing factors were identified (Page 7, Line 169; Page 8, Line 170-183). Third, we have added a statement to describe the method employed by included studies to investigate the influencing factors (Page 8, Line 183-185).

Comment (12): When presenting subgroup analysis you present numerical results by world region and survey year and state other sources of heterogeneity investigated were non-significant. The results for world region are also non-significant.

Response: Our wording might be confusing. As claimed in the result, the meta-regression was used to understand the potential source of heterogeneity, and Table 3 showed that some of the heterogeneities observed among the 12 factors can be partially explained by all four varieties (country, survey years, specialty and sample size). The subgroup analyses were performed to compare the studies stratified by region and survey years, and Table S2 showed that significant differences were noted between developed countries and developing countries in academic interests and prestige.

Comment (13): In general I consider that the firmness in which you present your conclusions is not supported by your results. There is considerable heterogeneity between studies that may mask associations. In addition the method of measurement used in studies and the potential error and bias inherent in transforming this into your outcome measure is a major limitation.

Response: Thank you for your critical comments. First, the estimates were pooled using a random-effects meta-analysis model due to the between-study heterogeneity (Page 7, Line 145-148). Second, ARHQ methodology checklist widely used for cross-sectional studies were applied to assess the quality of the studies (Page 6, Line 137-140). Therefore, we believe that the overall conclusions claimed in our study are well supported by the results.

Comment (14): Some of the information in the discussion would have been more helpful in the background and results sections – for example. I consider that the stage in a medical student's career is crucial to interpretation of many of the influencing factors measured, for example competencies. Indeed I would go so far as to say that this aspect of study design should have been investigated as a source of heterogeneity.

Response: We have added a statement of the stage in a medical student's career in method section (Page 6, Line 122-125).

Comment (15): In terms of further research would you consider that studies should be restricted by region? In terms of influencing policy then the most informative studies might be restricted by organizational and medical training factors; this will reduce heterogeneity and allow more meaningful investigation of other factors, for example there are large disparities in income between students in high income countries.

Response: Thanks for your suggestion. We agree with the reviewer that in terms of further research we would consider restricting studies by region. However, in the scope of the current study, the number of articles that can be retrieved after considering the organizational and medical training factors is limited for the analyses. We appreciate your suggestion and added a statement in our discussion section accordingly as follows: subgroup analysis stratified by organizational and medical training factors would provide more information of the factors influencing subspecialty choice among medical students (Page 14, Line 340-342).

Comment (16): Did you consider the value of qualitative research to further investigate your question?

Response: Thanks for your great suggestion. We did consider the value of qualitative research in investigating our question as you suggested. In the current manuscript, our study used the methodology of systematic review and meta-analysis to investigate our questions in a more defined, quantitatively manner. Qualitative research would be of great value in further studies but is beyond the scope of the current manuscript for clarifying the objective of this study. Your suggestion for further investigation based on qualitative research is very much appreciated.

REVIEWER	E. Benjamín Puertas Pan American Health Organization PAHO/WHO
REVIEW RETURNED	22-Aug-2018

GENERAL COMMENTS	Interesting article, specially since it found that income potential not necessarily is the main influeincing factor.

REVIEWER	Charlotte BEAUDART University of Liège
REVIEW RETURNED	10-Jul-2018

GENERAL COMMENTS	The authors have replied properly to the majority of my comments. However, I still have minor modifications request: Comment 4: the information about study "selection" has not been added in the manuscript. Please provide information about the number of reviewers involved in this process. Comment 8: please provide references for this new quality tool used – is it a valid tool? Comment 9: Q-test is not only used to assess difference between subgroups. Q-test is a test for heterogeneity. So yes, it assesses heterogeneity between groups but it also assesses heterogeneity between studies included in a meta-analysis. This heterogeneity is quantified with the I². Therefore, the I² result is complementary to the Q test. Please indicate, as requested in my first review, the use of Q-test in the methodology section to assess global heterogeneity of the model (line 145-146) and add Q-test values in table 2 (not only in table S2). Comment 10: I agree with the authors that they used generally accepted method to assess publication bias. However, since they noticed the presence of potential publication bias (Egger test is significant), they are encouraged to implement the Trim and Fill method in their results to assess the potential impact of this publication bias on their results.

REVIEWER	Clare Davenport University of Birmingham, UK
REVIEW RETURNED	30-Aug-2018

GENERAL COMMENTS	My comments are minor at this stage of revision. I think the review raises important questions. The authors appropriately discuss the limitations (largely as a result of the evidence available rather than their review methods). Introduction: Line 77- 79: is it the long training or competition for places that is the problem. I think the latter prolongs the journey to specialist. Eligibility criteria: Study design is not included here although it appears you looked for cross sectional studies. Methods: statistical analysis: Pre-defined sources of heterogeneity informing subgroup analysis and meta-regression should be specified here (12 factors that were split into economic and non economic factors). Pease reiterate what the EOI value is here and define what variables you will be investigating. The statement: 'Potential sources of heterogeneity were identified using

	meta-regression' suggest data driven analysis. Do the authors mean that pre-defined variables (see above) were investigated using meta-regression? I think these are the 12 variables presented in the results section?
	Results:
	The authors make statements about the results of subgroup analysis that do not take into account statistical significance. For example 'The EOI value of academic interests in developed 217 countries was higher than that in developing countries (79.66% [95% CI, 70.73%; 86.39% vs. 60.41% [95% CI, 43.44%; 75.19%]; does not appear significant whereas 'lower EOI value of prestige was found in studies conducted in developed countries than in developing countries (23.96% [95% CI, 19.20%; 29.47%] vs. 47.65% [95% CI, 34.41%; 61.24%]; Q=4.71 P=0.01)' appears statistically significant. Although the statistical tests are limited in power because of the number of studies and heterogeneity this distinction is important.
	Discussion:
	Your discussion about economic factors does not feature in your results although they are represented in a figure. I think that if you are to highlight these in the discussion they should be represented in the text of the results.

VERSION 2 – AUTHOR RESPONSE

Our point-by-point responses are as follows:

Editorial Requirements:

Comment (1): Since these are surveys among medical students the choices are hypothetical (what they plan) – can they explain that in the Abstract? The stage of their training will be crucial to that choice.

Response: Thanks for your constructive suggestion. We have added a statement in the abstract as follows: The recruited medical students includes students in medical school, internship, residency training and fellowship, who are about to or have just made a specialty choice (Page 2, Line 36-39).

Reviewer #1:

Comment: Interesting article, specially since it found that income potential not necessarily is the main influencing factor.

Response: Thank you so much for your appreciation about our study.

Reviewer #3:

Comment (1): The authors have replied properly to the majority of my comments.

However, I still have minor modifications request:

Comment 4: the information about study "selection" has not been added in the manuscript. Please provide information about the number of reviewers involved in this process.

Response: We have reworded the statement as follows: Each article was reviewed by two trained investigators (Y.Y. and J.L.) and the following information was independently extracted from each selected article using a standardized form: study design, geographic location, years of survey, journal, sample size, average age of the participants, the number and percentage of male participants, and the influencing factors and the extent of their influence (Page 6, Line 134-138).

Comment (2): Comment 8: please provide references for this new quality tool used – is it a valid tool?

Response: Thanks for your suggestion. We have cited the reference in Page 6, Pine 141.

Comment (3): Comment 9: Q-test is not only used to assess difference between subgroups. Q-test is a test for heterogeneity. So yes, it assesses heterogeneity between groups but it also assesses heterogeneity between studies included in a meta-analysis. This heterogeneity is quantified with the I². Therefore, the I² result is complementary to the Q test. Please indicate, as requested in my first review, the use of Q-test in the methodology section to assess global heterogeneity of the model (line 145-146) and add Q-test values in table 2 (not only in table S2).

Response: We have reworded the statements in the methods: Between-study heterogeneity was assessed using the Cochran' s Q-test, and was quantified with the I2 statistic, which was calculated to describe the percentage of total variation caused by heterogeneity across studies, with ≥50% indicating considerable heterogeneity (Page 7, Line 148). Also, we added Q-test values in **Table 2**.

Comment (4): Comment 10: I agree with the authors that they used generally accepted method to assess publication bias. However, since they noticed the presence of potential publication bias (Egger test is significant), they are encouraged to implement the Trim and Fill method in their results to assess the potential impact of this publication bias on their results.
Response: Many thanks for your comments. We have added statements in the manuscript: Fill and trim approach, which imputes estimates from hypothetical negative unpublished reports, was also used to investigate the publication bias if the Egger's test was significant (Page 7, Line 165-167). There was evidence of small study effect in the meta-analysis of "patient service orientation" (Egger's test P=0.02). However, the trim-and-fill method showed the publication-bias corrected estimate remained statistically significant (63.79%, 95% CI, 58.20%; 69.04%) (Page 10, Line 242-244).

Reviewer #4:

Comment (1): My comments are minor at this stage of revision. I think the review raises important questions. The authors appropriately discuss the limitations (largely as a result of the evidence available rather than their review methods).

Introduction:

Line 77- 79: is it the long training or competition for places that is the problem. I think the latter prolongs the journey to specialist.

Response: Thanks for your suggestion. We have reworded the statement: Some specialties and subspecialties, such as family medicine and palliative medicine, are experiencing a desperate shortage of physicians, whereas other specialties and subspecialties, such as cardiology, ophthalmology and ear, nose and throat (ENT) surgery, are highly competitive specialties with low success rate for candidates (Page 4, Line 80-81).

Comment (2): Eligibility criteria:

Study design is not included here although it appears you looked for cross sectional studies.

Response: Thank you for your kind reminding. We have reworded the statement in study eligibility section: Studies were included if they were systematic review or cross-sectional studies, reported data on medical students, were published in peer-reviewed journals, and used a validated method to assess the extent of a factor's influence on the choice of subspecialty, such as pediatric gastroenterology and vascular surgery, or its corresponding specialty, such as pediatrics and surgery (Page 5, Line 120).

Comment (3): Methods: statistical analysis:
Pre-defined sources of heterogeneity informing subgroup analysis and meta-regression should be specified here (12 factors that were split into economic and non-economic factors). Pease reiterate what the EOI value is here and define what variables you will be investigating.
The statement: 'Potential sources of heterogeneity were identified using meta-regression' suggest data driven analysis. Do the authors mean that pre-defined variables (see above) were investigated using meta-regression? I think these are the 12 variables presented in the results section?
Response: First, EOI value here is the same as we stated at the beginning of the manuscript (extent of their influence), thus we think there is no need to restate. Second, we have reworded the statements in statistical analysis section to make the variables of meta-regression clear in the method section: Potential sources of heterogeneity were identified using meta-regression. Four categorical covariates were identified as potential sources of heterogeneity by examining the studies conducted in the United States (US) vs. the studies conducted in other countries, the studies conducted before 2010 vs. those

conducted after 2010, the studies concerning subspecialty only vs. those that were not specific to a subspecialty, and the studies with a sample size <200 vs. the studies with a sample size ≥200. (Page 7, Line 151-156). Third, the 12 influencing factors are not the variables of the meta-regression.

Comment (5): *Results:*

The authors make statements about the results of subgroup analysis that do not take into account statistical significance. For example 'The EOI value of academic interests in developed countries was higher than that in developing countries (79.66% [95% CI, 70.73%; 86.39% vs. 60.41% [95% CI, 43.44%; 75.19%]; does not appear significant whereas 'lower EOI value of prestige was found in studies conducted in developed countries than in developing countries (23.96% [95% CI, 19.20%; 29.47%]vs. 47.65% [95%CI, 34.41%; 61.24%]; Q=4.71 P=0.01)' appears statistically significant. Although the statistical tests are limited in power because of the number of studies and heterogeneity this distinction is important.

Response: Thanks for your suggestion. As we stated in the result section: the EOI value of academic interests in developed countries was higher than that in developing countries (79.66% [95% CI, 70.73%; 86.39% vs. 60.41% [95% CI, 43.44%; 75.19%]; *Q*=3.51 *P*=0.02, Page 10, Line 227-229). Note that the *P*-value in this subgroup analysis is 0.02 <0.05, which indicates the statistical significance, rather than "not appear significant".

Comment (6): *Discussion:*

Your discussion about economic factors does not feature in your results although they are represented in a figure. I think that if you are to highlight these in the discussion they should be represented in the text of the results.

Response: Our wording might be confusing. As claimed in the discussion section, we classified all the 12 influencing factors into two categories, economic factors and non-economic factor (Page 11, Line 248-249). All the 12 factors are mentioned throughout the manuscript (abstract, method, result, discussion, and the according tables and figures).

VERSION 3 – REVIEW

REVIEWER	Charlotte Beaudart University of Liège
REVIEW RETURNED	08-Jan-2019

GENERAL COMMENTS	No further comment