Dear Dr. Merino,

Please find enclosed our revised manuscript entitled "Quality of Care Delivered by General International Medical Graduates in U.S. Hospitals (Manuscript ID BMJ.2016.035412)." We appreciate the issues raised by the reviewers and the editors, and are grateful for the opportunity to respond. Below are the comments we received from the editor and reviewers (in italics) and our responses (in bold).

We paid particular attention to the statistical editor's queries and suggestions as well as tried to fully address each of the reviewer's concerns and comments.

We believe that the reviewer comments and those of the statistical editor were very helpful and addressing them has strengthened the manuscript, and we hope you agree. If there are any further issues or concerns that we can address, please do not hesitate to contact us.

Sincerely,
Ashish Jha, MD, MPH

Editors' Comments:

First, please revise your paper to respond to all of the comments by the reviewers. Their reports are available at the end of this letter, below.

Thank you for your thoughtful comments and giving us an opportunity to revise our manuscript. We think addressing the points raised by editors and reviewers has substantially improved the paper.

We found the topic interesting and important but have some concerns, particularly regarding some of the statistical analyses. these are summarized in the review done by Dr. Julie Morris, out statistical expert. Please address all her points. Her review is included below.

We appreciate the guidance and have carefully addressed all of Dr. Morris's suggestions. We re-ran our models using multivariable logistic regression (instead of linear probability models) for all analyses in our revised manuscript. We also tested 3 different methods for attributing patient outcomes to physicians (including the main approach using the largest Part B spending), and found that our results were not sensitive to how we attributed patients to physicians. We hope these have sufficiently addressed the concerns. Please see below (response to reviewer 6) for additional details. In addition, because the latest 2014 Medicare data have now been released (that were not available when we first submitted our manuscript), we included the 2014 data for our analyses in the revised manuscript.

Please provide greater international context. What is the situation in other countries? Where do IMGs in the US come from?

This is an important issue and we appreciate the opportunity to provide a greater context. IMGs make up a substantial proportion of physician workforce in many high-income countries, which is why we believe that our paper likely has relevance far beyond the U.S. We have added the following sentences in our introduction to provide greater international context (page 9, paragraph 1). We are happy to expand more about the statements about international context if the editors prefer that.

"Licensure process for IMGs varies modestly between these countries. IMGs have to pass English language and Professional and Linguistic Assessments Board (PLAB) tests in order to practice in the UK.1 In the U.S., IMGs must pass 2 examinations testing medical knowledge and 1 examination assessing clinical skills, and must acquire accredited residency training in the U.S. (see supplementary appendix 1 for more details).2"

Reviewer 1 (Prof. Balakrishnan R Nair):

International Medical Graduates (IMGs) make up more than 20% of the workforce in USA, UK, Canada and Australia. They are the backbone of some specialties and geographical areas where locally trained doctors are reluctant to work. However the IMGs are often targeted when things go wrong , often from system issues, in the media. So this article looking at the performance of IMGs and measuring the usual bench marks like mortality, readmission rates and cost is very timely. The authors have measured this in common clinical conditions where the internists take care of the patients. When this happens in lower socio economic groups in smaller set ups , the results are more significant. The methodology and sample size are adequate and the paper is well written.

Thank you for your comments.

It is interesting to note more IMGs work in internal medicine compared to the local graduates.

As we show in Table 2, IMGs make up nearly 45% of all Internists (slightly fewer than local U.S. graduates). However, given that IMGs were on average treating a larger number of patients, the number of patients treated by IMGs (813,014) was larger than that of USMGs (679,977).

May be better outcomes are from more "holistic" care given by the internists than subspecialists?

As we described in methods section (page 12, paragraph 2), we compared IMGs and USMGs who were all general internists to avoid comparing physicians of different specialties. Although not directly examined in our paper, it is indeed possible that internists provide more "holistic" care than specialists; however, this is beyond the scope of our study as our paper focused on outcomes for USMG internists versus IMG internists. We clarify this point in the discussion of our revised manuscript (page 12, paragraph 2).

Overall good paper with good take home messages for IMGs and people involved in assessing them and licencing bodies

Thank you for your thoughtful review.

Reviewer 2 (Prof. Stephen Leeder):

The performance of doctors trained overseas compared to that of doctors trained in the country in which they practice - the US or the UK in particular - is important to patients and providers alike. By comparing the outcomes of care for selected Medicare patients treated by international medical graduates (IMGs) and those trained in the US, a picture emerges in which the IMGs essentially performed as well and maybe even a little better than their home grown counterparts.

The study, which uses routinely collected data, is thoughtfully constructed. There is no reason to suppose that the data are influenced at the stage of collection and coding by an awareness of the origin of the Hospitalist responsible for the care of the patient. The statistical methods employed are appropriate and not overly elaborate, allowing the reader to understand the logic and limitation of their application.

Thank you for your comments.

Given the quality of the data and the analytical methods that make allowance for many sources of variance in the outcome data such as hospital type, level of service specialisation etc., the findings appear sound. Whether differences in outcome might be present in other forms of illness management (psychiatry, intensive care, geriatric care) where cultural affinity may play a more prominent role is a residual question of interest - but that would require a different study.

We agree with the reviewer that whether we find similar findings among other specialties such as psychiatry and intensive care is a very important question and of substantial interest. Unfortunately, it is, as the reviewer notes, outside the scope of the current paper. However, we think such work should be done to better understand those issues. We emphasize this point in the limitation section of our revised manuscript (page 25, paragraph 1):

"Future research is warranted to understand whether patient outcomes differ between IMGs and domestically-trained doctors for other specialties (e.g., psychiatry, intensive care, geriatric care), in the outpatient setting, in the non-Medicare populations, and in other high-income countries."

I found only one typo (!) and the paper is really well written. The diagrams are clear, well labelled and easy to understand. I would recommend its acceptance without change.

Thank you again for a very thoughtful review. We double-checked for any typos and attempted to correct them in our revised manuscript.

Reviewer 3 (Dr. Sachin H. Jain):
The authors undertake a careful analysis on a subject of great importance for US and UK workforce policy--namely the quality of care provided by physicians of foreign origin in the United States. The analysis focuses on hospitalists and the outcomes--mortality and readmission rates--are well-respected quality measures for the hospital medicine field. The authors take great care to risk adjust their findings and evaluate the level of acuity. Their finding that quality is stable across both foreign and US trained physicians is an important one, worthy of the pages of the British Medical Journal.

We appreciate the reviewer's comments.

The paper's discussion could be strengthened in the following ways:
1) More explicitly discuss the xenophobia and bias against physicians who have trained abroad and take on the implicit assumption in the paper about the quality of medical training in the US versus abroad

This is an excellent point. As suggested, we added a discussion about implicit bias against physicians trained abroad in the introduction:

"Furthermore, studies have suggested that there are implicit biases and discrimination against physicians who have trained abroad with an assumption that quality of care delivered by IMGs may be inferior to that of USMGs.3-5"

2) More explicitly discuss a significant limitation of the paper--which is that it looks at quality of care provided by international medical graduates, but it does not acknowledge the incredible heterogeneity that exists within the broader cohort "international medical graduates." There are international medical graduates who grow up abroad and migrate to the US and there are US residents who go abroad for medical training and return to the US. These groups are typically viewed quite differently and the authors should acknowledge that there may be different results if they further stratified their analyses.

Thank you for this thoughtful point with which we agree wholly. We could not consistently identify U.S. citizens who go abroad for a medical training and return to the U.S. to practice. Most of the U.S. citizens IMGs (USIMGs) typically graduate from medical schools in Central America and the Caribbean (73% of USIMGs graduate from medical schools in these countries). When we excluded all physicians from these countries, our findings were essentially unchanged. However, we agree with the reviewer's main point here (about heterogeneity) and acknowledge so in the revised version of the manuscript.

As suggested, we added the following explanation in the discussion of our revised manuscript (page 22, paragraph 1):

"First, we could not distinguish between foreign-born IMGs and U.S. citizens who travel abroad for medical training. Approximately 73% of U.S. citizens who are IMGs attend medical schools in Central America and the Caribbean.6-8 Excluding physicians from these countries in our sensitivity analysis did not affect our findings. However, the results of our analyses by country suggest that there may be some heterogeneity among IMGs based on both where they trained and their reasons for coming to the U.S."

3) The paper should more explicitly discuss the caveat that while 30-day mortality and readmission rates are accepted measures of quality for hospital medicine, there are other aspects of clinical care that may not be captured by those measures--patient satisfaction; effective use of evidence-based practice guidelines; etc.

This is an excellent point. Based on the suggestion, we added the following sentences to the limitation section of the discussion (page 22, paragraph 1):

"Finally, although 30-day mortality and readmission rates are widely-accepted measures of quality for hospital care, they are not fully comprehensive. There are other important aspects of inpatient care, such as patient experience, which we did not measure."

Overally, this is a great topic--and a well executed paper. With some minor revisions to the discussion, I would recommend that the editors accept it for publication.

Thank you again for your helpful comments.

Reviewer 4 (Laurent Gerbaud):

This paper is interesting, well done, based on validated methods, and assesses a very few explored questions about the links between quality of care place of physicians' initial training.
The assessment criteria (30 day mortality and readmission rate) is not unusual and due to the use of large database, authors were able to take into account a lot of confusing variables in the main analyses as well as in sensitivity analyses and/or for specific clinical conditions and/or for some specific medical exercises.
The results are consistent with the data and statistical methods and lead to discussion and conclusion congruent with these results.
However, on the discussion section, the authors should take care to two factors:
- The IMG are younger than USMG, that is closer to their initial training and to best practices (that are more recently teach)

Thank you for the careful review and the very helpful comments. Based on the reviewer's suggestion, we have added the following sentences in the discussion of the revised manuscript (page 23, paragraph 1):

"although IMGs were younger and therefore closer to their initial training, which might have allowed them to be more aligned with best practices, adjustment for physicians' age did not affect our findings, suggesting that physicians' age does not explain the difference in mortality.

- The IMG may have a different strategy about continuous training: as they are foreigner they may be more cautious not to fail and, this way, more interested in continuous training and willing to apply any new skill in their current practice.

This is an interesting point that we had not considered but agree is plausible. We have added the following sentence to the discussion (page 24, paragraph 1):

"It is also possible that IMGs may be more concerned about professional failure, and therefore, are more engaged in continuous training and updating their skills and knowledgebase, although we lack evidence to either support or refute this hypothesis.

Reviewer 5 (Patrick Dowling):

This observation study is of significance as it compares the of quality care rendered by US Medical graduates (USMGs) and IMGs , for hospitalized patients in the US, age 65 or older. . The authors are to be commended for putting together the data sets with several sources involving over 900,000 patients and 40,500 allopathic physicians. Moreover, although the US currently has over 220,000 IMGs out of a total of almost 900,000 physicians, it faces a physician shortage as demand is increasing secondary to the 20 million Americans who gained health insurance as a result of President Obama's Affordable Care Act.
Unfortunately, the paper has several limitations in my opinion. The first involves the order of references as they did not seem to match the text as discussed below.

Thank you for such a careful review, and our apologies for the errors in the order of references. For some reasons, the EndNote program did not function appropriately and we should have caught this error. We are grateful to the reviewer for identifying and bringing it to our attention. We have double-checked the references and fixed these problems in the revised manuscript.

Second, the number of statistical tests conducted and the possibility of randomly finding statistically significant results. The large numbers of tests are not grouped in an intuitive way in the text and therefore it is somewhat confusing to track them all.

We appreciate this point and apologize for not having been clear. Although we did carry out several statistical tests, all tests were chosen a priori, which should reduce concerns about increase in the risk of type 1 errors. However, we agree with the reviewer that given that we examined 3 outcomes, we should be using stricter threshold for p-value (known as the Bonferroni correction). Therefore, in our revised manuscript, we used the p-value less than 0.0167 to be considered as statistically significant, and we also added the following sentence to clarify this point (page 15, paragraph 1):

"Because we evaluated 3 outcomes, $p<0.017$ was considered as statistically significant."

Finally, important physician variables are missing from analysis or seem unclear.

Physician variables included in initial analyses were physician age and sex. Given that we restricted our analyses to patients treated by general internists, all the physicians' specialties were, by definition, the same. Based on the reviewer's comments, we considered what else we might account for and included another important physician variable that was available – patient volume – in our regression models in our revised manuscript. Unfortunately, there were no other physician-level variables available in our data.

Introduction:
Page 9 of 54 (PDF), lines 44-54 represents an example of the references not matching the text . This occurs on the very first page of the narrative. Lines 44-54 involve two sentences which mention concerns about the quality of IMGs and the limited studies which actually compare outcomes between IMGs and USMGs. Five references are cited--# 7-11. Reference, #7, is by Jena AB who is the 2nd author on this paper . It addresses salary differences by gender among faculty members, by specialty, of public medical schools in the US. IMGs are not mentioned nor is quality discussed; Ref. #8 is the US News & World Report, a non-medical news magazine in the US as opposed to a peer reviewed medical journal ; Ref #9 by McWilliams, JM in N Engl. J Med does not address IMGs ; Ref # 10 is the web page for FAIMER, the Foundation for Advancement of International Medical

Education and Research) The resource listed, the "World Directory of Medical Schools" is a work in progress by which the organization aspires to list all the medical schools in the world with the goal of developing a method for accreditation in the future. Once the process is finished , it may address quality issues of IMGs by different schools Finally, Ref # 11 is the web site for the Association of American Medical Colleges (AAMC) which is very extensive . No specific subsite was listed.
This problem continues on page ten, lines 4-20, which begins with "The studies that have examined patient…" is not properly cited. Cited references # 2-4 have nothing to do with patient outcome studies. Given that , It is not possible to know what studies the authors are referring to that support their assertions. It is very disconcerting to find such inaccuracies in the introduction of a study.

Again, we are sorry about these errors in our references and very much appreciate the reviewer raising them. We have fixed these in our revised manuscript.

Methods:
(Page 12, first paragraph, lines 1-10) The physician sample was limited to internist. How was this determined? And how accurate is the source? Was it self-reported?

Thank you for raising this point. Physicians' specialties were identified using data from external sources such as the American Board of Medical Specialties and other specialty societies. We clarified this in our revised manuscript (page 12, paragraph 1).

(Page 12, second paragraph, lines 17-36) For physicians residency training is critical as it sets them on the pathway to board certification that is used as a proxy for the quality of the care rendered by a physician. Given that, the important question here with respect to quality is residency completion in the US AND board certification. What was the percentage of board certification in the two comparison groups?

This is an important question. We found that 83.0% of USMGs and 77.8% of IMGs were board-certified in our data. Additional adjustment for the board-certification status did not affect our findings (supplementary table E-G).

(Page 12, second paragraph, lines 30-36) states that the dated on medical school of graduation was found on 83.6% of the physicians. Were the 16.4% of physicians with missing medical school information in Doximity different than the rest of the sample?

This is an important issue and to address the reviewer's concern, we ran new analyses examining this question. Below is the table that examines differences among two groups of physicians: physicians with and without missing data on medical schools. The data in Tables 1 and 2 (below) suggest that physicians with missing data on medical schools were somewhat younger and more likely to be female, but patient characteristics were very similar between physicians with and without data on medical schools.

Table 1. Physician and patient characteristics, physicians with missing medical school data

	Physicians with data on medical schools (n=46,823)	Physicians without data on medical schools (n=8,374)
Physician characteristics		
Physician's age, y, mean (SD)	46.9 (11.1)	42.2 (10.7)
Female, %	30.9%	37.8%
No. of hospitalizations per year, n	196	174
Patient characteristics		
Number of patients, n	1,543,859	244,955
Patient's age, y, mean (SD)	76.0 (13.6)	75.9 (13.6)
Female, %	59.4%	59.2%
Race, % White	77.5%	77.0%
Black	13.6%	13.4%
Hispanic	5.9%	6.4%
Other races	3.1%	3.3%
Household income, $, mean (SD)	55,561 (22,193)	56,319 (22,339)
Medicaid status, %	31.2%	31.3%
Coexisting condition, % CHF	19.0%	18.9%
COPD	26.0%	25.6%
Diabetes	33.1%	33.2%
Renal failure	21.7%	22.2%
Neurological disorders	16.7%	16.3%
Cancer	6.6%	6.6%
Mental illness	18.0%	17.6%
Discharge location, % Home	61.1%	62.2%
Skilled nursing facility	23.9%	23.8%
Rehabilitation facility	2.3%	2.1%
Hospice	4.0%	3.9%
Others	8.7%	8.0%

Table 2. Hospital characteristics, physicians with missing medical school data

	Physicians with data on medical schools (n=46,823)	Physicians with missing data on medical schools (n=8,374)
Hospital characteristics		
Hospital size, % Large (400+ beds)	39.1%	40.2%
Medium (100-399 beds)	53.4%	53.0%
Small (1-99 beds)	7.5%	6.9%
Teaching status, % Major teaching	25.4%	27.0%
Minor teaching	34.1%	33.9%
Non-teaching	40.5%	39.1%
Ownership, % For-profit	12.9%	12.5%

Non-profit 75.9% 76.9%
Public 11.1% 10.6%
Region, % Northeast 22.8% 27.4%
Midwest 23.6% 21.0%
South 34.1% 31.4%
West 18.8% 18.5%
Rural-urban status, % Urban 86.9% 87.1%
Suburban 2.0% 1.9%
Large rural 8.9% 8.9%
Small rural 2.2% 2.1%
ICU, % Hospitals with ICU 86.9% 86.4%
Hospitals without ICU 13.1% 13.6%

To address the concern of the reviewer about the potential impact of the missing physicians, we re-ran our primary models to account for the missing data. We used a widely-accepted approach of building a weighted model where weights were generated based on the inverse probability of medical school data being observed, as recommended by Parezen et al..9 We found that our findings were not affected by using this approach (see Table 3 below).

We appreciate the reviewer's concern and feel reassured that the small amount of missing data were unlikely to have affected our results.

Table 3. Adjusted 30-day mortality rates, IMGs vs USMGs, after accounting for missing data on medical schools
Adjusted patient outcomes
(95%CI) Adjusted
risk difference
(95%CI)
IMG − USMG p-value
IMGs USMGs
Main model (excluding physicians with missing data on medical schools) 11.2%
(11.1% to 11.3%) 11.6%
(11.5% to 11.7%) -0.5%
(-0.6% to -0.3%) <0.001
Weighted regression to account for missing data 11.1%
(11.0% to 11.2%) 11.6%
(11.5% to 11.7%) -0.5%
(-0.6% to -0.3%) <0.001
Note: The probability of missing data was modeled using patient and physician characteristics, and the outcome regression model was weighted by the inverse of the predicted probability of missingness.

(Page 13, first paragraph, lines 20-26) Only physician age and gender are included in the analysis but what about sub-specialties such as pulmonology, geriatrics etc... etc…? What about board certifications status? These are critically important characteristics that impact care rendered to patients.

Thank you for raising these points. In order to avoid comparing physicians with different specialties, we excluded subspecialists and focused just on general internists. To address the reviewer's concern, we adjusted for patient volume for all analyses we conducted. In terms of board certification, as stated above, we found that 83.0% of USMGs and 77.8% of IMGs were board-certified. When we included board certification in our models, the findings did not change (see supplementary tables E-G).

(Page 13, first paragraph, lines 26-32) Where teaching hospitals excluded? In teaching hospitals, the care is provided by residents and services billed by the attending physician. Where hospitals categorized based on categories such as community or academic teaching hospital. Are results reflective of hospital types because IMGs are employed in certain types of hospitals versus others?

These are important issues, and we appreciate the opportunity to clarify them.

First, teaching hospitals were NOT excluded from our analyses (see Table 1 of our manuscript). We included hospital fixed effects in our regression models, and therefore, we are only examining outcomes of IMGs to USMGs within the same hospital. To the extent that IMGs might cluster in different hospitals should not affect our findings.

We agree with the reviewer that it would be interesting to examine differences in patient outcomes between IMGs and USMGs within certain subgroups of hospitals such as major teaching hospitals. Therefore, in response to the reviewer, we built new multivariable models that were stratified by teaching status. We found that IMGs have lower mortality regardless of the teaching status, i.e., even in non-teaching hospitals. We included these results in Table H of the supplementary material.

Table 4. Stratified analysis by teaching status of hospitals
No. of hospitalizations
(No. of physicians) Adjusted patient outcomes
(95%CI) Adjusted
OR/difference
(95%CI)
IMG vs USMG p-value
IMGs USMGs
30-day mortality rate Major teaching hospitals 183,383
(13,242) 10.2%
(9.9% to 10.4%) 10.8%
(10.5% to 11.0%) 0.93
(0.88 to 0.97) 0.001
Minor teaching hospitals 408,108
(18,734) 11.3%
(11.1% to 11.4%) 11.7%

(11.5% to 11.8%) 0.95
(0.93 to 0.98) 0.002
Non-teaching hospitals 580,542
(21,911) 11.5%
(11.3% to 11.6%) 11.9%
(11.3% to 11.6%) 0.95
(0.93 to 0.97) <0.001
30-day readmission rate Major teaching hospitals 184,352
(13,418) 16.7%
(16.5% to 17.0%) 16.7%
(16.4% to 16.9%) 1.01
(0.98 to 1.04) 0.71
Minor teaching hospitals 400,402
(18,390) 15.1%
(15.0% to 15.3%) 15.2%
(15.1% to 15.4%) 0.99
(0.97 to 1.01) 0.51
Non-teaching hospitals 556,810
(20,760) 15.2%
(15.0% to 15.3%) 15.2%
(15.1% to 15.4%) 0.99
(0.97 to 1.01) 0.54
Total Part B costs per hospitalization Major teaching hospitals 199,626
(14,313) $1224
($1212 to $1237) $1172
($1159 to $1185) +$52
(+$33 to +$72) <0.001
Minor teaching hospitals 431,304
(19,191) $1167
($1159 to $1175) $1118
($1110 to $1126) +$49
(+$37 to +$61) <0.001
Non-teaching hospitals 601,845
(22,188) $1103
($1096 to $1109) $1061
($1054 to $1069) +$41
(+$30 to +$52) <0.001
Note: Adjusted for patient and physician characteristics and hospital fixed effects.

(Page 14, third paragraph, lines 19-26) How were internists identified as being "general internists"? It is important to note this in the methods section.

Thank you for raising this point. Physicians' specialties were identified using multiple sources including the data collected by American Board of Medical Specialties and other specialty societies. We clarified this in our revised manuscript (page 12, paragraph 1).

(Page 14, third paragraph, lines 23-33) Looking at the top medical conditions seems too simplistic. Some include in the list are ambulatory sensitive conditions and other are not. A more robust and validated framework to categorize conditions is needed if the goal is to attribute 30 day mortality outcomes to physicians caring for patients in hospitals only. This seems somewhat arbitrary to this reviewer.

We very much appreciate the reviewer's point that there are important clinical differences across the 6 conditions. We chose this approach because, we wanted to examine whether the effects we were seeing overall were also present within individual conditions. We think the fact that the patterns of care seen overall also show up in individual conditions is important validation of the overall findings.

We chose the most common conditions for three reasons – first we wanted to ensure we were not being arbitrary. Any specific conditions we chose, if we had picked them based on our clinical intuition, could have been subject to criticism that other clinicians might choose different ones. Picking the most common means that we were not able to influence which conditions were included versus not.

The second reason to take this approach is a public health one – these are the conditions that occur most often for hospitalized patients and therefore, outcomes for these conditions have the biggest health impact for the hospitalized patients. Finally, there was a statistical benefit of choosing these conditions – because they were the most common, they afforded the most amount of power to find any difference that might be present, reducing the risk that we would make a type 2 statistical error (failing to find a difference when there was one).

That all said, we agree with the reviewer that one could imagine other approaches to picking individual conditions. Our examination of the literature did not turn up any standard frameworks that we could apply to choosing some individual conditions but not others. However, if the editors wish for us to take a different approach to choosing individual conditions, we are happy to do so.

(Page 14, fourth paragraph, lines 49-54) It is not clear which physicians were included in the analysis. Early in the manuscript, the authors state that it is internists, then general internists, and now in this paragraph state that it is "hospitalists". These categorizations all have very different meaning in workforce research and it is critical that this is absolutely clear.

We apologize for not being clear about this. We analyzed the patients treated by "general internists" for our main analyses, and we evaluated patients cared for by "hospitalists" in our sensitivity analysis (because unmeasured confounding is less of a potential concern among patients of hospitalists). We double-checked our manuscript and made sure that we are using either "general internists" or "hospitalists" throughout our manuscript.

(Page 15, first paragraph, lines 10-17) Many graduates from a particular university in Guadalajara Mexico are also US citizens. They are the so-designated as USIMGs Although the authors excluded USIMGs for the for-profit Caribbean medical schools, were they aware that this school in Mexico had a longstanding program—the 5th pathway, for US students ? How was this handled in the sensitivity analysis? Mexico is listed as sending the 6th most IMGs to the US but the vast majority of these may be from the 5th pathway program and US citizens making it different that the other countries listed.

Thank you for raising this point. Based on the reviewer's comment, we created a new sensitivity analysis where, instead of just excluding IMGs who graduated from Caribbean medical schools (as we did originally), we excluded IMGs who graduated from both Central America and the Caribbean countries. Previous studies have found that these countries make-up approximately 73% of IMGs who are U.S. citizens.6-8Making this additional exclusion did not meaningfully change our results (see tables E-G of the supplementary material).

Results:
Table 1, which is mentioned on page 16, and found on page 24, lists Physician and hospital characteristics.
As mentioned the authors state these were internist, they do not comment on whether or not they were all board certified or if they had done additional sub-specialty training but were practicing as general hospitalists. Further , there were very significant differences between the hospital characteristics of the treating IMG physicians and USMG physicians , some of which I have listed from the table.
IMGs USMGs
Non-teaching hosp. 46.9 % 36.7%
Major teaching 18.8 29.0
Public Hosp. 8.6 12.4
Sans ICU 15.4 11.6
>400 beds 33.8 42.2
For-profit 15.7 11.2

Thank you for your point, and we agree that this is important. We agree with the reviewer that there are differences between IMGs and USMGs in terms of their practice location. This is part of the reason we ran models using hospital fixed effects so we can compare IMGs and USMGs within the same hospital. Therefore, the fact that IMGs practice in different hospitals than USMGs should not impact our findings. We apologize if this was not clear initially and emphasized this in our revised manuscript.

Several major cities in the US have large (>400 beds) not for profit, public (usually county or city hospitals) that are major teaching hospitals. They care for a large number of indigents and low income patients. They have several hundred and as many as 900 physicians is residency and fellowship training programs. These characteristics are more likely to be found in the column under USMGs above. Yet it is not uncommon that over 50% of the residents and fellows are IMGs , especially in the non-surgical, less lucrative specialties such as general internal medicine. In these hospital wards IMG trainees would be providing most of the care under the supervision of an attending physician, which would most likely be a USMG as per the above. Hence even though the bulk of the care is being provided by the IMGs , the billing physician , under Part B of Medicare , would be the attending physician which could be the USMG. Thus rather than the comparison be between IMGs and USMGs, the comparison might actually be between IMG trainees and IMGs. Were the authors aware of this nuance in the billings in teaching hospitals? If this was not controlled for , the data could be misleading.

This is an important point, and indeed, we were aware that there are differences in trainee make-up between hospitals. However, we believe that the attending physician is important for patient outcomes, even in teaching hospitals, and therefore, we think the outcomes grouped by the physician of record is still important.

That said, we think the reviewer raises an important point. In order to address this (and the reviewer's point earlier), we conducted new analyses where we built models stratified by teaching status to examine whether the patterns would look the same across hospitals based on teaching status. If the reviewer's very reasonable hypothesis is right, we might not see differences in patient mortality between IMGs and USMGs in non-teaching hospitals, if the observed difference in patient mortality is solely due to the difference in residents composition.

We found that the results looked essentially identical across the three strata of hospitals (major teaching, minor teaching, and non-teaching) suggesting that any misclassification we might have at major teaching hospitals is likely to be very small and does not appear to affect our findings (see Table 3 above).

(Page 17, first paragraph, lines 7-9) What was the unit of analysis? Hospitalizations or patients? This is not clear in the text.

The unit of analysis was hospitalizations and not patients. We are sorry for the confusion. We clarified this in our revised manuscript (page 11, paragraph 2).

Other comments:
Licensing and credentialing vary from state to state in the US. A sensitivity analysis looking at his would help clarify the claims in the conclusions about systems of credentialing and the quality of IMGs.

This is an interesting point. Based on the reviewer's suggestion, we conducted a stratified analysis by state. In order to avoid unstable estimates, we restricted to states with at least 10,000 observations in our national dataset.

Out of 30 states that satisfied the criterion of 10,000 or more hospitalizations in our data, we found lower mortality rates for IMGs in 27 states (though the differences were not always statistically significant). The fact that in 27 of the 30 states (where the analyses were not powered to find differences at the individual state level), the results were consistent with the national results suggest that state level variation in credentialing and quality of IMGs does not threaten the validity of the national findings.

Table 4. Adjusted 30-day mortality rate between IMGs and USMGs, by state
State No. of hospitalizations Adjusted 30-day mortality rate
(95%CI) Adjusted risk difference
(95%CI)
IMG − USMG p-value

IMGs USMGs
AL 21,220 11.2% (10.3% to 12.1%) 11.9% (11.3% to 12.5%) -0.7% (-2.0% to +0.5%) 0.26
AZ 19,089 9.9% (9.4% to 10.4%) 10.7% (9.8% to 11.7%) -0.8% (-2.0% to +0.3%) 0.17
AR 10,790 12.9% (11.6% to 14.3%) 13.0% (12.2% to 13.8%) -0.1% (-1.9% to +1.8%) 0.95
CA 81,670 12.0% (11.7% to 12.4%) 12.8% (12.4% to 13.2%) -0.8% (-1.3% to -0.2%) 0.01
CO 10,944 9.8% (8.3% to 11.2%) 11.6% (11.0% to 12.2%) -1.8% (-3.5% to -0.2%) 0.03
CT 18,398 11.2% (10.5% to 11.8%) 12.3% (11.3% to 13.3%) -1.2% (-2.5% to +0.1%) 0.08
FL 89,422 10.5% (10.2% to 10.7%) 10.8% (10.4% to 11.3%) -0.4% (-0.9% to +0.2%) 0.17
GA 32,295 11.9% (11.4% to 12.5%) 12.2% (11.7% to 12.8%) -0.3% (-1.1% to +0.5%) 0.44
IL 62,756 11.4% (11.0% to 11.7%) 11.4% (10.9% to 11.9%) -0.05% (-0.7% to +0.6%) 0.89
IN 27,943 11.4% (10.9% to 12.0%) 11.6% (10.9% to 12.2%) -0.1% (-1.1% to +0.8%) 0.77
KY 19,770 10.3% (9.6% to 11.1%) 11.3% (10.6% to 12.0%) -0.9% (-2.1% to +0.2%) 0.12
LA 16,624 10.5% (9.5% to 11.4%) 12.0% (11.3% to 12.7%) -1.5% (-2.8% to -0.3%) 0.02
MD 33,751 9.8% (9.3% to 10.2%) 11.5% (10.7% to 12.3%) -1.7% (-2.7% to -0.8%) 0.001
MA 40,513 9.2% (8.8% to 9.7%) 9.8% (9.3% to 10.3%) -0.6% (-1.3% to +0.2%) 0.15
MI 54,746 10.6% (10.2% to 10.9%) 10.9% (10.3% to 11.5%) -0.3% (-1.1% to +0.4%) 0.41
MN 10,865 11.2% (10.1% to 12.3%) 10.4% (9.7% to 11.1%) +0.8% (-0.6% to +2.2%) 0.27
MS 12,965 11.3% (9.7% to 12.8%) 12.2% (11.3% to 13.1%) -0.9% (-3.0% to +1.2%) 0.40
MO 26,835 12.3% (11.6% to 13.0%) 10.9% (10.2% to 11.6%) +1.4% (+0.3% to +2.5%) 0.02
NJ 38,378 10.9% (10.5% to 11.3%) 11.7% (10.9% to 12.4%) -0.8% (-1.7% to +0.1%) 0.09
NY 72,899 11.1% (10.8% to 11.5%) 11.8% (11.3% to 12.3%) -0.7% (-1.3% to -0.02%) 0.04
NC 45,966 11.7% (11.2% to 12.3%) 11.7% (11.2% to 12.1%) +0.04% (-0.7% to +0.8%) 0.92
OH 49,252 10.2% (9.9% to 10.6%) 10.6% (10.1% to 11.1%) -0.4% (-1.0% to +0.3%) 0.28
OK 11,945 12.0% (11.1% to 12.9%) 12.5% (11.6% to 13.4%) -0.5% (-2.0% to +1.0%) 0.50
PA 49,571 10.7% (10.3% to 11.1%) 10.8% (10.3% to 11.3%) -0.1% (-0.8% to +0.6%) 0.76
SC 16,679 12.7% (11.9% to 13.5%) 12.9% (12.0% to 13.7%) -0.2% (-1.5% to +1.1%) 0.73
TN 31,446 12.2% (11.5% to 12.9%) 12.3% (11.7% to 12.8%) -0.1% (-1.1% to +0.9%) 0.89
TX 84,707 11.1% (10.8% to 11.4%) 11.4% (11.1% to 11.8%) -0.3% (-0.8% to +0.2%) 0.25
VA 37,881 12.1% (11.5% to 12.6%) 12.9% (12.3% to 13.5%) -0.8% (-1.7% to +0.1%) 0.07
WA 23,484 12.8% (12.0% to 13.6%) 13.3% (12.7% to 13.9%) -0.5% (-1.6% to +0.6%) 0.41
WI 20,960 10.9% (10.2% to 11.5%) 11.5% (10.9% to 12.1%) -0.6% (-1.7% to +0.5%) 0.26
Note: Adjusted for patient and physician characteristics and hospital fixed effects. We used multivariable linear probability models to examine the difference in patient mortality between IMGs and USMGs.

The title should include "by internists" to denote the physicians of focus in the study.

Based on the reviewer's suggestion, we changed the title to "Quality of Care Delivered by Internists who Graduated from Foreign versus U.S. Medical Schools in U.S. Hospitals: Observational Study."

The abstract should include mention of the physician sample being internists etc. It is somewhat deceptive otherwise and will lead readers to think the analysis includes all physicians.

We apologize for this omission. Based on the reviewer's suggestion, we have clarified that we focused on patients treated by general internists in our abstract.

The conclusions are somewhat overstated and should be revised slightly given the study design limitations and assumptions. I would reject in present form as I believe it needs significant revisions .

Thank for your comments. We toned down our conclusions to be as follows (page 20, paragraph 2):

"Using a nationally representative sample of hospitalized elderly patients in the U.S., we found no evidence that patient outcomes were worse for IMGs than for USMGs. In fact, patients treated by IMGs had modestly lower 30-day mortality than those cared for by USMGs."

We appreciate for the reviewer's thoughtful comments. We tried to address the reviewer's concerns by building new statistical models, running new analyses, and rewriting sections of the manuscript to be more in line with the reviewer's interpretation. We hope we have sufficiently addressed the reviewer's concerns. If there are remaining issues, please let us know, and we would be happy to address them further.

Reviewer 6 (Julie Morris):

This paper compares patient outcomes between international medical graduates and US medical graduates for hospitalized patients 65 years and above who were in US Medicare.

There are some issues relating to the statistical analysis which require clarification. In addition, the interpretation of some of the results needs to be re-considered.

1. Linear regression analysis for binary outcomes (ie. 30-day mortality) rather than logistic/probit regression is not commonly used. More details should be given as to why this method was used (Page 13). Was an identity link used or a logit link used to ensure predicted values were in the correct range? Were robust standard errors employed? It is unclear why logistic regression (the alternative, more appropriate analysis?) was then used as a sensitivity analysis (Page 15). This latter analysis gave almost identical results to the primary analysis (see Table H).

The statistical editor raises and important point, and we apologize that we did not clarify the reason for our approach earlier. In our initial manuscript, we used simple OLS, i.e., an identity link, instead of a logit link to ensure predicted values were in the correct range. Therefore, the editor was right that it was possible that predicted values took negative values or values larger than 1. And the editor is correct that we clustered robust standard errors at the physician level.

While we wholly agree that logistic regression is more commonly used for binary outcomes, we initially chose to use a linear probability model for computational efficiency, and checked the OLS model with a logistic regression of the overall model to ensure that our results were qualitatively similar.

Due to large sample sizes and the complexity of the models (with over 3500 variables in each model since we are using hospital fixed effects), each logistic regression model takes 7-10 days to converge. The computational time is on average <1 hour for linear probability models. Because in the overall model, they gave nearly identical answers, we felt comfortable proceeding with a linear probability model.

However, we very much appreciate the statistical editor's concern. In order to address it, we have now re-run all of our models using a logistic approach and present those in the main revised manuscript. We again clustered robust standard errors at the physician level which we clarified in our manuscript.

2. The attribution of patients to specific physicians is of some concern in terms of assessing the effects of a single physician on the treatment and care of a patient rather than the actual influence of a team of physicians. However, the results obtained using an attribution based on higher spending and the results from a sensitivity analysis which had attribution based on the largest number of claims are said to be very similar. This finding does go some way to allay these concerns.

Thank you for raising this point, and we agree that there is no perfect attribution method. We agree that ultimately, care is delivered by a group of providers and attributing care to a single physician is imperfect. That said, we do believe that for common, major medical conditions like the ones we examined, the attending physician largely influences the course of care, and to a large extent, patient outcomes.

We agree that there are multiple reasonable approaches to attributing patients to the physician, and we too were reassured that the approach we used did not meaningfully change the results. To further address this issue, we added one more approach for attributing patient outcomes to physicians. In the revised manuscript, we attributed patient outcomes to (1) physicians with largest Part B spending (main analyses), (2) physicians who accounted for the largest number of evaluation-and-management (E&M) claims, and (3) physicians who billed the first E&M claims. We found that our findings were consistent across all 3 methods employed.

3. Statistically significant differences between mortality rates were found (in favour of IMGs). However the actual difference is very small (11.0% vs 11.7%) and there is no discussion of the apparent lack of any real clinical significance of this difference.

This is an important point, and the degree to which adjusted OR of 0.95, or a 0.4 percentage point difference, is clinically meaningful is debatable – but we wholly agree that the difference is not large. We believe it is clinically meaningful because when we compare it to treatments we use commonly in inpatient settings, many that are considered quite effective have effect sizes on mortality that are considerably smaller than 0.4 percentage points.

We wholly agree with the editor's point that this issue needs far clearer discussion than we had previously included. In response, in our revised manuscript, we added new paragraph in the discussion section. The new text argues that a 0.4 percentage point difference in patient mortality is at, the very most, of modest clinical significance but it may be seen as not being clinically meaningful at all. It is reassuring that IMGs certainly don't seem to be worse than USMGs (page 23, paragraph 2).

4. Page 18. "The lack of differences in mortality for patients with COPD and UTI may be because of lack of statistical power to detect differences in the relatively low overall 30-day mortality rates for this population".
But the mortality rates for IMGs and USMGs are identical for UTIs (6.1% vs 6.1%) and only 0.1% apart for COPD (4.6% vs 4.7%) – see Figure 1 (A). Hence it seems inappropriate to comment on 'lack of statistical power' in the absence of any real difference.

Thank you for bringing this to our attention, and we fully agree. Based on the editor's comment, we removed the corresponding sentences from our revised manuscript.
References
1. General Medical Council. Applying for registration as an international medical graduate [Available from: http://www.gmc-uk.org/doctors/before_you_apply/imgs.asp accessed November 18 2016.
2. American Medical Association. Practice medicine. Four key steps to begin practicing medicine in the U.S. Chicago, IL [Available from: http://www.ama-assn.org/ama/pub/about-ama/our-people/member-groups-sections/international-medical-graduates/practicing-medicine.page accessed June 17 2016.
3. Delamothe T. The plight of the international medical graduate. BMJ 2014;348:g2894.
4. Desbiens NA, Vidaillet HJ, Jr. Discrimination against international medical graduates in the United States residency program selection process. BMC Med Educ 2010;10:5. doi: 10.1186/1472-6920-10-5
5. Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. BMJ 2013;347:f5662. doi: 10.1136/bmj.f5662
6. McAvinue MB, Boulet JR, Kelly WC, et al. U.S. citizens who graduated from medical schools outside the United States and Canada and received certification from the Educational Commission for Foreign Medical Graduates, 1983-2002. Acad Med 2005;80(5):473-8.
7. Eckhert NL, van Zanten M. U.S.-citizen international medical graduates--a boon for the workforce? N Engl J Med 2015;372(18):1686-7. doi: 10.1056/NEJMp1415239
8. Norcini J, Anderson MB, McKinley DW. The medical education of United States citizens who train abroad. Surgery 2006;140(3):338-46. doi: 10.1016/j.surg.2006.06.001
9. Parzen M, Lipsitz SR, Ibrahim JG, et al. A weighted estimating equation for linear regression with missing covariate data. Stat Med 2002;21(16):2421-36. doi: 10.1002/sim.1195

1. Cover letter IMG BMJ FINAL.docx PDF HTM