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)	Antibiotic use and clinical outcomes in the acute setting under management by an Infectious Diseases Acute Physician versus other clinical teams: a cohort study
AUTHORS	Fawcett, Nicola; Jones, Nicola; Quan, T; Mistry, Vikash; Crook, Derrick; Peto, Tim; Walker, Sarah

VERSION 1 - REVIEW

REVIEWER	Kuang-Hung Hsu, PhD Chang Gung University, Taiwan
REVIEW RETURNED	22-Jan-2016

GENERAL COMMENTS	The manuscript titled 'Antibiotic use and clinical outcomes in the acute setting under management by an infectious diseases acute physician versus other clinical teams: a cohort study' demonstrated IDPs can reduce antibiotic uses while keeping the medical quality. In addition, the study found that the patients may be holding and be admitted longer for observation and more conservative prescription of antibiotics. The study samples were collected from a single medical setting and patients were treated by an experienced IDP. While this study provides evidence of the effectiveness of IDP, several weakness of this study were found. Comments by section are as follows. General idea of the manuscript A vast body of evidence has demonstrated that the employment of IDPs can reduce antibiotic prescriptions and guards the patient's prognosis. As exemplified by a review article (Pulcini et al, 2014; Clinical Microbiology and Infection 2014; 20: 963–972), a nationwide population based analysis (Shih et al, 2014; Journal of Microbiology, Immunology and Infection 2014; 47: 297-303), and others, this study confirms the findings by analyzing data from their medical setting. This study adds information on how the IDP managed patients to reduce antibiotics. However, due to the limited number of IDPs observed in this study, the generalizability of the findings is uncertain. Introduction Since the core significance of this study is to examine the effect of IDP on reducing antibiotics prescriptions and patients' clinical outcomes. The literature review of the related papers should be more comprehensive and in a balance way (the second paragraph of this part). There are a few literatures addressing positive effects of IDPs on reducing antibiotics and clinical outcomes. In addition, the management strategy or practice behavioral changes between IDPs and non-IDPs should be addressed in literature review in which may lead to why to examine the prescribing holding behavior or length of stay in study. Therefore, the hypothesis testing and study design followed the rationale. It seems to be insufficient information and a

great gap of knowledge in this section.

Methods

The study actually collected data on one-week basis and confirmed the clinical outcomes by a large database. However, in the methods section, the readers could not understand this arrangement and why. The questions raised from your methods are 1) how is the one-week data collection representable to the rest of unobserved periods? 2) Is the IDP/team similar to other IDPs in UK or a particularly trained professional? 3) You have so many supplement information (4 tables and 2 figures). You need to re-organize and decide which one to be included. The manuscript should be presented in a concise and logically sound way.

Results

1. Patient's characteristics were not presented. The readers cannot judge whether the patient profile was fairly comparable between the two study groups. Have you performed individual matching or propensity score matching to assure the comparability and validity of this study?

2. In table 1, one cannot tell which regression model was used in each patient-level metric. In addition, the way of presenting descriptive statistics was rather busy and confusing. What were those numbers in parenthesis with 'sum' meaningful to the readers? The cell, DOT and Non-IDP, may be a typing error. Please check.

3. You tried to conclude nonsignificant results in this section without providing power calculation based on this sample size of which is not acceptable. For example, the last sentence of page 12 was deemed as speculation.

4. Page 13, the description of the two paragraphs will lead the readers to ask why you don't use the three-year dataset only to address the questions in this study. The statistical test on delays in treating sepsis between non-IDP (11/20) and IDP (4/5) was not surprisingly to be nonsignificant (p=0.62). However, the 55% vs 80% difference may be statistically significant if sample size is sufficiently enough, say in your three-year dataset. It is hard for us to swing back and forth between one-week trial data and three-year dataset, especially the differences of 30-day mortality (15%/13% vs. 6.3%/5.7%), and admitted overnight (70%/88% vs. 79%/83%) were not consistency between two datasets in which patient profile may be different.

5. Figure 1 presented with different color of dots may be not so explicit instead of using numbers in table. The readers want to catch your points at a short glance but not spending time on counting dots in different colors and figuring out the meaning of their sequences.

6. I think that the better way of expressing this data is to use large dataset and reorganize tables and figures into three parts, such as descriptive statistics/univariate tests, multiple regression models on practice behaviors, and multiple regression models on clinical outcomes and/or medical consumption.

Discussion

1. I think that the most important issue of this study is to address the significance and knowledge gap provided beyond current literatures.

2. The authors should do more comprehensive literature review. Page 17, the last sentence of the first paragraph may not be true. There were literatures to address the association of patients' clinical outcome and antibiotics reduction by IDPs. The two papers listed in the previous section are only part of the samples.

3. I think that the cost-effectiveness issues may be interesting and important for the professional society. The authors may elaborate the issues based on their dataset.

REVIEWER	Jaap ten Oever
	Radboudumc; Netherlands
REVIEW RETURNED	15-Mar-2016

GENERAL COMMENTS	Fawcett et al. show that teams led by Infectious Diseases Acute Physician prescribe less antibiotics than teams led by non-IDPs without an increase in adverse events. The authors explain this difference by showing that IDP teams withhold antibiotic treatment in case of uncertain infective diagnosis (at the expense of an increased length of stay). Numbers are small, but the methodology is good and the findings are consistent.
	I have a few comments/questions:
	-If I understand correctly, only patients admitted to the medical service were included. Does this mean that patients that were sent home immediately after an emergency department visit were excluded? If so, why were those patients not included? Could a difference in admission policy and/or antibiotic prescription to patients sent home explain the findings?
	-In line 145 the authors assume that 20% of patients were under IDP during the audit period. In line 161 (the power calculation for the larger data set) the authors assume that 1% of the admission were seen by the IDP team. What is the explanation for this difference?
	-In line 290 the authors state that they included "take-home"antibiotics. It would be easier for the reader if the authors explain that antibiotic treatment after discharge was taken into account when determining DOT (first alinea page 8).
	-My suggestions would be to more explicitly mention that the principal diagnosis was copied from the attending physician and not retrospectively made. The definition of likelihood of infection (Figure 1) should be explained in the methods. Was likelihood of infection included in the analysis of predictors of antibiotic prescription?
	-I miss the abbreviations in the tables/figures.

REVIEWER	Hui Nian
	Department of Biostatistics
	Vanderbilt University Medical Center
	U.S.A.
REVIEW RETURNED	02-May-2016

GENERAL COMMENTS	The study assessed the antibiotic use and clinical outcomes between IDP patients and non-IDP patients in the acute setting, and found that the IDP-led group used significantly less antibiotic therapy with no adverse clinical outcome. The manuscript is clearly written. There are some statistical issues that I think can be addressed in a revision.
	1. In the sample size justification (the first paragraph of Statistical Analysis), please clarify the type of statistical test/model used and other information needed in order to derive the number of 367 patients.
	2. "DOT and length-of-stay were truncated at 18 and 17 (95th centiles) respectively to reduce the influence of outliers. "
	If the data were accurately measured, they shouldn't be excluded from the analysis. Otherwise, the estimates could be biased.

3. I wonder if the authors had considered using negative binomial model, which also allows for overdispersion and is a simpler model to estimate and interpret. AIC or BIC statistics can be used to evaluate whether negative binomial model is preferred over zero inflated poisson model.

4. Is there a particular reason why the authors used a different model (logistic regression model instead of zero inflated poisson model) to do variable selection (backwards elimination)? Also, how many predictors (degrees of freedom) to start with? Is it true that the sample size of this study is not big enough to include all the clinically relevant predictors in the model and such variable selection must be performed? If so, have the authors considered using propensity score method?

5. Minor edit. Line 148. "ranksum tests" → "Wilcoxon rank-sum test"

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name
Kuang-Hung Hsu, PhD

Institution and Country
Chang Gung University, Taiwan

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

Please leave your comments for the authors below
The manuscript titled 'Antibiotic use and clinical outcomes in the acute setting under management by an infectious diseases acute physician versus other clinical teams: a cohort study' demonstrated IDPs can reduce antibiotic uses while keeping the medical quality. In addition, the study found that the patients may be holding and be admitted longer for observation and more conservative prescription of antibiotics. The study samples were collected from a single medical setting and patients were treated by an experienced IDP. While this study provides evidence of the effectiveness of IDP, several weakness of this study were found. Comments by section are as follows.

General idea of the manuscript
A vast body of evidence has demonstrated that the employment of IDPs can reduce antibiotic prescriptions and guards the patient's prognosis. As exemplified by a review article (Pulcini et al, 2014; Clinical Microbiology and Infection 2014; 20: 963–972), a nationwide population based analysis (Shih et al, 2014; Journal of Microbiology, Immunology and Infection 2014; 47: 297-303), and others, this study confirms the findings by analyzing data from their medical setting. This study adds information on how the IDP managed patients to reduce antibiotics. However, due to the limited number of IDPs observed in this study, the generalizability of the findings is uncertain.

Introduction
Since the core significance of this study is to examine the effect of IDP on reducing antibiotics prescriptions and patients' clinical outcomes. The literature review of the related papers should be more comprehensive and in a balance way (the second paragraph of this part). There are a few literatures addressing positive effects of IDPs on reducing antibiotics and clinical outcomes. In

addition, the management strategy or practice behavioral changes between IDPs and non-IDPs should be addressed in literature review in which may lead to why to examine the prescribing holding behavior or length of stay in study. Therefore, the hypothesis testing and study design followed the rationale. It seems to be insufficient information and a great gap of knowledge in this section.

Response: We acknowledge the suggestion that a broader review of the IDP literature would improve the manuscript, and have expanded our introduction accordingly. We are aware of the large literature related to IDP input to ward-based management, providing reviews of hospital inpatients already admitted and managed under a separate clinician. However our study was looking at a different aspect of IDP management, namely the evidence for difference in prescribing and outcomes for patients managed from point of referral by an IDP-led acute team, taking into account decision-making during early diagnostic uncertainty, in a highly heterogenous group of patients and including patients discharged rapidly, often before a chance to perform a ward-based review.
We have clarified this greatly in our introduction.
We have also attempted to draw the distinction between the literature on the effect of IDP ward based reviews, and the studies examining (retrospectively) the effect of IDP-led management of acute patients.

Methods
The study actually collected data on one-week basis and confirmed the clinical outcomes by a large database. However, in the methods section, the readers could not understand this arrangement and why. The questions raised from your methods are
1) how is the one-week data collection representable to the rest of unobserved periods?

Response: We apologise for a lack of clarity in our description of the 1 week and 3 year datasets. We have extended this in both a short sentence at the end of our introduction where we say how we plan to address our objectives, and in more detail in the methods section. In brief, the 3 year dataset was from an electronic health records database which contained only administrative data (mortality, admissions, length of stay etc), but not patient-level clinical data or antibiotic prescribing. We have also included a new Supplementary Table 3 comparing the 1 week and 3 year cohorts showing that there were no statistically significant differences in available patient characteristics namely age, gender, Charlson score and admission at weekend, and refer to this in the Results section. We have also added a discussion of this point to the study limitations.

2) Is the IDP/team similar to other IDPs in UK or a particularly trained professional?

Response: We acknowledge that our studied IDP may not represent IDPs in the UK as a whole, and have added this in the discussion of study limitations. Their combined training in both an infectious diseases/microbiology specialty combined with general medicine is common in the UK, as is IDP participation in the acute medical services.

You have so many supplement information (4 tables and 2 figures). You need to re-organize and decide which one to be included. The manuscript should be presented in a concise and logically sound way.

Response: as per subsequent suggestions by the reviewer below, we have changed the order of our tables and results section, moving one supplementary table (patient characteristics) to the main text. However, we consider that the other supplementary material contains information about the patients managed under IDP vs non-IDP that may be of interest to specific readers; as far as we are aware there is no limit on the amount of supplementary material and so we would prefer to keep it as supplementary (which is only 3 pages), unless the Editors strongly preferred we reduce it.

Results
1. Patient's characteristics were not presented. The readers cannot judge whether the patient profile was fairly comparable between the two study groups. Have you performed individual matching or propensity score matching to assure the comparability and validity of this study?

Response: We apologise that this was not clear in the original submission. The patient characteristics of the IDP and non-IDP groups were presented in the original supplementary information table 3, together with a formal statistical comparison, and referred to at the start of the Results. We have moved this table into the main body of the text as suggested.

2. In table 1, one cannot tell which regression model was used in each patient-level metric. In addition, the way of presenting descriptive statistics was rather busy and confusing. What were those numbers in parenthesis with 'sum' meaningful to the readers? The cell, DOT and Non-IDP, may be a typing error. Please check.

Response: Each patient-level metric has a separate regression model, adjusting for patient-level characteristics as described in the Methods. We present the full multivariable regression model for DOT in the Table 3 (Table 1 in the original submission), and described in the Table 2 footnote the fact that predictors were similar for other outcomes. In the light of the comment of the reviewer above, we do not feel that adding another 6 supplementary tables containing all these multivariable models would be particularly helpful. The specific effect of different case-mix variables is not the main point of interest for this paper, rather we have focussed on the fact that, after adjusting for them, the differences between IDP vs non-IDP groups remain. We have instead clarified the existing table 2 footnotes and ensured that the information provided is consistent across the different models. We have also reorganised the table order following the reviewer's further suggestions below to improve clarity. The sum numbers were included to provide raw data on DOT and DDD in each group and they have been removed as suggested. We have corrected the DOT and non-IDP cell as noted by the reviewer.

3. You tried to conclude nonsignificant results in this section without providing power calculation based on this sample size of which is not acceptable. For example, the last sentence of page 12 was deemed as speculation.

Response: we included our sample size calculation in the original manuscript which was based on the primary outcome as stated. Having observed a significant difference on our primary outcome, we go on to explore potential drivers of this observed difference. We have amended this section to state clearly that we observed no evidence of difference on the various quality metrics. We had already noted that numbers were small: we have amended the text to state more clearly that, as numbers were small, we are not concluding that there is no difference in quality between IDP and non-IDP, but that the lack of evidence of difference (and similar point estimates) mean that quality is not likely to have driven the overall significant difference in antibiotics prescribed between the groups.

4. Page 13, the description of the two paragraphs will lead the readers to ask why you don't use the three-year dataset only to address the questions in this study. The statistical test on delays in treating sepsis between non-IDP (11/20) and IDP (4/5) was not surprisingly to be nonsignificant (p=0.62). However, the 55% vs 80% difference may be statistically significant if sample size is sufficiently enough, say in your three-year dataset. It is hard for us to swing back and forth between one-week trial data and three-year dataset, especially the differences of 30-day mortality (15%/13% vs. 6.3%/5.7%), and admitted overnight (70%/88% vs. 79%/83%) were not consistency between two datasets in which patient profile may be different.

Response: We agree that it would have been preferable to examine data on the 3 year dataset to

improve power; unfortunately the data on prescribing was not available in the 3 year dataset, which was purely of administrative information (admissions, mortality etc). We have now clarified our description of the two datasets, how they were obtained and what data they contained, and highlighted the limit of data from the 3 year dataset, though it offers greater dataset size. We have also restructured our clinical outcomes section in the Results to examine first the 1 week dataset in detail, then make it clear we are replicating the findings in a much larger, separate dataset.

5. Figure 1 presented with different color of dots may be not so explicit instead of using numbers in table. The readers want to catch your points at a short glance but not spending time on counting dots in different colors and figuring out the meaning of their sequences.

Response: We have changed Figure 1 to a table as requested.

6. I think that the better way of expressing this data is to use large dataset and reorganize tables and figures into three parts, such as descriptive statistics/univariate tests, multiple regression models on practice behaviors, and multiple regression models on clinical outcomes and/or medical consumption.

Response: We have reorganised the text and tables as suggested, so that first, patient characteristics are presented together with prescribing metrics, followed by regression modelling and discussion and results of differing practice, followed by mortality/clinical outcomes. As above, we are not able to use the large dataset for anything except a limited number of outcomes.

Discussion
1. I think that the most important issue of this study is to address the significance and knowledge gap provided beyond current literatures.
2. The authors should do more comprehensive literature review. Page 17, the last sentence of the first paragraph may not be true. There were literatures to address the association of patients' clinical outcome and antibiotics reduction by IDPs. The two papers listed in the previous section are only part of the samples.

Response: We have added key references to the literature of IDP effects in the introduction, including the review highlighted by the reviewer above, and also noted the difference between the literature on IDP consultations and IDP-led management of acute patients which is the context in which our study was conducted. We have also extended our discussion to include this. However, cognisant of the points made by the reviewer at the start of their comments, we have also added an explicit comment on the limitation of including only one IDP in the Discussion.

3. I think that the cost-effectiveness issues may be interesting and important for the professional society. The authors may elaborate the issues based on their dataset.

Response: We have added a sentence to our summary paragraph to highlight this issue. However, we have not added more information on this in order to remain within the 4000 word count limit; rather we have used these words to more fully describe the literature as we felt that this was a more important point to address.

Reviewer: 2

Reviewer Name
Jaap ten Oever

Institution and Country
Radboudumc; Netherlands

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

Please leave your comments for the authors below
Fawcett et al. show that teams led by Infectious Diseases Acute Physician prescribe less antibiotics than teams led by non-IDPs without an increase in adverse events. The authors explain this difference by showing that IDP teams withhold antibiotic treatment in case of uncertain infective diagnosis (at the expense of an increased length of stay). Numbers are small, but the methodology is good and the findings are consistent.

I have a few comments/questions:
-If I understand correctly, only patients admitted to the medical service were included. Does this mean that patients that were sent home immediately after an emergency department visit were excluded? If so, why were those patients not included? Could a difference in admission policy and/or antibiotic prescription to patients sent home explain the findings?

Response: We have extended our description of the medical services in the Methods section to clarify that all patients who were referred to the medical team by the Emergency Department (ED) and who received a documented review from the medical team were included. Patients who attended the ED who were not referred to the medical team were not included, since the data used for the study did not exist. On reflection of ways that this could potentially affect our findings, it is theoretically possible that the ED could change referral behaviour based on knowledge of the acute physician on-call. Local experience suggests that usual knowledge in ED of the on call team is limited to the bleep number to phone. We have thus not elaborated further in our discussion of this point (rather prioritising other additions suggested above), but could add more text on this instead if the Editors preferred.

-In line 145 the authors assume that 20% of patients were under IDP during the audit period. In line 161 (the power calculation for the larger data set) the authors assume that 1% of the admission were seen by the IDP team. What is the explanation for this difference?

Response: The IDP was on full clinical duties during the audit period, but was not continuously on duty during the whole of the 3 year period, which is why the percentages differ. We have now included a more in-depth description of our larger dataset, including how the IDP undertook other clinical duties and had periods of leave over the 3 year period, so overall they admitted a lower percentage of patients.

-In line 290 the authors state that they included "take-home"antibiotics. It would be easier for the reader if the authors explain that antibiotic treatment after discharge was taken into account when determining DOT (first alinea page 8).

Response: We have amended this as suggested.

-My suggestions would be to more explicitly mention that the principal diagnosis was copied from the attending physician and not retrospectively made. The definition of likelihood of infection (Figure 1) should be explained in the methods. Was likelihood of infection included in the analysis of predictors of antibiotic prescription?

Response: We have clarified the description of the provenance of the principal diagnosis and added how likelihood of infection was determined to the methods as suggested. Likelihood of infection was not included in the analysis of predictors, as we wished to include purely objective criteria.

-I miss the abbreviations in the tables/figures.

Response: We apologise and have added these to the tables.

Reviewer: 3

Reviewer Name
Hui Nian

Institution and Country
Department of Biostatistics
Vanderbilt University Medical Center
U.S.A.

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

Please leave your comments for the authors below
The study assessed the antibiotic use and clinical outcomes between IDP patients and non-IDP patients in the acute setting, and found that the IDP-led group used significantly less antibiotic therapy with no adverse clinical outcome. The manuscript is clearly written. There are some statistical issues that I think can be addressed in a revision.

1. In the sample size justification (the first paragraph of Statistical Analysis), please clarify the type of statistical test/model used and other information needed in order to derive the number of 367 patients.

Response: We have added that the sample size calculation was done using simulation from a Poisson model with the rates and proportions in the IDP vs non-IDP group as indicated. We apologise that the reduction the study was powered to detect (30%) had mistakenly been summarised as one-third.

2. "DOT and length-of-stay were truncated at 18 and 17 (95th centiles) respectively to reduce the influence of outliers. "
If the data were accurately measured, they shouldn't be excluded from the analysis. Otherwise, the estimates could be biased.

Response: No data were excluded: as stated, very large outliers (due to extended lengths of stay) were truncated at the 95th percentiles in regression models to avoid problems with undue influence and over-fitting to extreme values. We have clarified this in the Methods.

3. I wonder if the authors had considered using negative binomial model, which also allows for overdispersion and is a simpler model to estimate and interpret. AIC or BIC statistics can be used to evaluate whether negative binomial model is preferred over zero inflated poisson model.

Response: We initially explored both standard Poisson and standard negative binomial models as well as the zero-inflated Poisson model – in fact the standard negative binomial model was a far worse fit. We attribute this fact to the fact that only 33% of patients received antibiotics, but in those that did the median DOT was 6 days, with IQR from 5 to 9 days (Table 2) – thus any rate model which does not include a zero-inflated component struggles to fit the data. Further we found that several factors were important predictors of this zero component (Table 3). As a specific example, fitting a standand negative binomial model to our Table 3 dataset including all the factors in the Poisson model has an AIC of 968.9; in contrast the Table 3 zero-inflated Poisson model has an AIC of 799.2.

4. Is there a particular reason why the authors used a different model (logistic regression model instead of zero inflated poisson model) to do variable selection (backwards elimination)? Also, how many predictors (degrees of freedom) to start with?

Response: the zero-inflated Poisson model has two components, the zero-component (receipt of antibiotics at all) and the rate component (DOT). Variable selection therefore needs to consider both parts of the model. However, we were aware of the potential for non-linearity in predictors, particularly age, CRP etc, and so we wanted to use a variable selection method that allowed for this, such as fractional polynomials. However, the mfp common in Stata which does this does not incorporate models such as the zero-inflated poisson model which has the two distinct linear predictors. As above, there is rather clear separation between the DOT in those receiving antibiotics and the "zero" DOT in those not receiving antibiotics. We therefore considered it would be reasonable to identify the predictors of receiving vs not receiving antibiotics using a standard multivariable fractional polynomial logistic regression; and then use these identified covariates in the zero-inflation component of the Poisson model, using multivariable fractional polynomial regression to then identify predictors of the volume of antibiotics received in those receiving them. We have clarified this in the revised text, and also added the number of predictors we started with.

Is it true that the sample size of this study is not big enough to include all the clinically relevant predictors in the model and such variable selection must be performed? If so, have the authors considered using propensity score method?

Response: The goal of the regression models was to adjust the observed difference between IDP vs non-IDP for case-mix, rather than to try to make inference about specific factors affecting prescription and volume of antibiotics received. We are therefore not primarily concerned with the estimated effect of these other factors, and the reviewer is correct that we did not consider that the study sample size was large enough to include all 20 factors in the model without variable selection (and in fact never considered this approach in our analysis plan).

We agree that propensity scores would have been a valid way to adjust for case-mix for the zero-inflation part of the model (receiving antibiotics or not), but cannot immediately see how we could incorporate different predictors in this zero-inflation component vs the volume component of the Poisson model using a standard propensity score analysis. Perhaps not unsurprisingly, we did find that the predictors of these two aspects were somewhat different (albeit accepting the reviewer's point about power). In particular, we identified some evidence for opposite effects, whereby patients with frailty syndrome were more likely to be prescribed antibiotics, but received lower volumes when prescribed, which makes sense clinically, but we think would be very difficult to adequately cover using one propensity score.

As this is a statistical point, we have not incorporated any information on this into the current draft, but could do so if the Editors wished.

5. Minor edit. Line 148. "ranksum tests" → "Wilcoxon rank-sum test"

Response: we have changed this.

VERSION 2 – REVIEW

REVIEWER	Jaap ten Oever Radboud university medical centre
REVIEW RETURNED	25-Jul-2016

GENERAL COMMENTS	The authors have addressed all the raised issues adequetely

REVIEWER	Hui Nian Vanderbilt University Department of Biostatistics USA
REVIEW RETURNED	22-Jul-2016

GENERAL COMMENTS	All the comments have been fully addressed.