PEER REVIEW HISTORY

BMJ Medicine publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evidence at Time of Regulatory Approval and Cost of New Antibiotics, 2016-2019
AUTHORS	Kesselheim, Aaron; Mitra-Majumdar, Mayookha; Powers, John; Brown, Beatrice

VERSION 1 - REVIEW

REVIEWER 1	Outterson, Kevin; Boston University School of Law. Competing Interest: I lead CARB-X, which makes grants to antibacterial product developers. I have no financial conflict of interest with drug R&D. CARB-X is funded only by governments and independent charitable foundations.
REVIEW RETURNED	17-Jan-2022

GENERAL COMMENTS	2/29 “come with high price tags” – true for the first in class TB drugs, less true for the antibiotics for Gram-positive and Gram-negative bacteria, especially when compared to other drug classes. New (innovative) drugs will always be more expensive than the generic comparators, otherwise there will be no funding for R&D. Non-inferiority is indeed an issue, though. 4/7 The fatality figures are in the US only. Shouldn’t “Gram” be capitalized as a proper noun? 4/10 “reportedly slowed” – see the recent CID article by Dheman and FDA colleagues. Undoubtably, it has slowed; no need for the “reportedly” modifier that raises doubt. 4/14 Slow uptake is directly related to both the financial struggles (including bankruptcies) but also the issue you raise on non-inferiority. Connected issues. 5/3 Excellent work on these methods, continuing to build that PORTAL database 5/40 Is there a literature on the relationship b/w Red Book and actual prices? 6/13 This list of “antibiotics” differs from some others in the published literature, for example by including the TB drugs and inhaled amikacin. You have also included a combination with a non-NME component and an mAb. Can you provide a more precise definition of your inclusion and exclusion criteria? (such as – all NME antibacterials J01 & J06…etc., systemic, as you think best). Otherwise, it is difficult to replicate the study for longer periods of time. This request may require you to re-run the results, or to give estimates of the impact of inclusion and exclusion criteria on the results and conclusions.

	6/52 If most of the NI margins were 10-15%, please report this in addition to the largest margin of 20%. What was the reason the regulatory allowed the wider margin? 7/40 There is a literature on NI trials in antibiotics, some of which could be cited in this discussion paragraph. The reader should know that others have written on this topic. 8/6 HPV is another example of indirect endpoints being useful and necessary. 8/11 “unvalidated indirect endpoints” – this implies that the particular endpoints in these antibiotic pivotal trials were “unvalidated”. If that is true, please explain. 8/20 “limited evidence of questionable rigor” – what has been shown about pretomanid after approval? What has been the experience at MSF, the Global Fund, TB Allliance, and others post-approval? You cast negative aspersions on the drug, which may or may not have been confirmed by subsequent evidence. See, e.g., https://pubmed.ncbi.nlm.nih.gov/34400340/. 8/42 Are you saying that plazomicin is more dangerous than other ahminoglycosides that are effective? 9/7 “new antibiotics meant to fill unmet medical needs lack confirmatory evidence that they do so” – you only present data from the pivotal trial, not the evidence present today. The FDA saw enough evidence to approve the drugs, so what do you mean by “confirmatory evidence”? Are you saying the FDA standards are too low? 9/12 Agreed that any new incentives must have stronger indicia of quality (satisfying unmet medical needs)

REVIEWER 2	Ramachandran, Reshma; Yale School of Medicine. Competing Interest: I serve on the Doctors for America Drug Affordability Action Team, which is funded by the Laura and John Arnold Foundation. I also serve on the boards of Universities Allied for Essential Medicines and the American Medical Student Association. I have worked with one of the study authors, Dr. John Powers in a volunteer capacity through the National Physicians Alliance and more recently, in his role with the Laura and John Arnold Foundation.
REVIEW RETURNED	17-Jan-2022

GENERAL COMMENTS	Review of “Characteristics of the Pivotal Trials and Cost of New Antibiotics, 2016-2019” In this follow-on study to their prior analysis published in 2016, the authors characterized the pivotal clinical trials of antibiotics approved by the FDA between October 2016 and December 2019. In addition to extracting specific characteristics of these trials, the authors examined the launch prices for these antibiotics as well as the status of meeting post-marketing requirements or commitments requested by the FDA upon approval. The latter piece is a new addition to their previous work.

	Like the previous article, the authors find that the pivotal trials for newly-approved antimicrobials have largely demonstrated uncertain efficacy and a majority are non-inferiority trials. They also found that no drug within their study cohort has completed or submitted its post-marketing requirements or commitments. Finally, they also found wide price variation for these drugs with some drugs being priced quite high. While the study is of a timely and interesting topic, there are areas for methodological improvement as detailed in the comments below. Additionally, the lack of summary tables of the study's key findings makes it difficult to follow the results and understand how the findings relate to each other. The study would be significantly strengthened if the authors make clear the connections between evidentiary standards for approval and the cost analyses for the drug. For instance, are there clear differences in the costs and costs ratios between the drugs that were approved based on non-inferiority margins versus those tested in a superiority trial? Additionally, did those drugs with direct versus indirect endpoints also have differences in terms of costs and cost ratios? Comparative conclusions are also drawn of these study findings, but it is unclear to what cohort the study findings are made to. Were these drugs approved based on fewer pivotal trials than all FDA approved drugs or the previous cohort studied? Further specific comments on the study can be found below - Major comments: (Abstract, page 2, lines 19-20): The authors write that all drugs were approved based on changes to surrogate measures, but this does not seem to be reflected in the tables in the manuscript. In the actual manuscript (Results), it seems that this has been clarified further but should be edited in the abstract to reflect this finding. (What this Study Adds, page 3, lines 17-18): Can we say for sure that policy incentives led to a steady increase of novel antibiotics without examining the prior period or is this a possible suggestion? Given the duration needed for antibiotic drug development, would the initiation of development for drugs within this cohort predated the incentives? (Methods, page 5, Lines 26-27): In reviewing the FDA Orange Book and the data descriptions (https://www.fda.gov/drugs/drug-approvals-and-databases/orange-book-data-files), it is unclear to me where there is information regarding Post-Marketing Requirements and Commitments that were used to confirm these for each antibacterial product. Could the authors point to more specifically where they extracted this information? (Methods, page 5, lines 28-29): Further details on how the authors categorized Post-Marketing Requirements and Commitments as "open" and "closed" as well as the other subcategories should be included as well as what database was used to do so. Did the authors verify this through ClinicalTrials.gov and through publication searches as well? (Methods, page 5, lines 55-56): Could the authors clarify why the cost of treatment calculated based on dosage and administration used in the pivotal trials rather than the labels as the latter would more likely inform clinical practice? Were there significant differences between the comparator regimens used within the clinical trials and those in the label and/or as recommended in

	clinical practice? And to further clarify, was the regimen used within the pivotal trials or the label used to calculate the cost of approved antibacterial? (Results, page 6, lines 44-45): The authors should consider determining the median number of patients across trial arms for those pivotal trials that did have active or placebo comparators? (Results, page 6, lines 52-54): Further clarity on how the authors determined whether the trial design was that of a non-inferiority or superiority trial would be helpful – was this based on FDA's determination or that of the manufacturer? For example, with Recarbrio, the FDA stated that the trial for the cUTI indication were "not powered to demonstrate non-inferiority in the appropriate patient population" (Multi-Discipline Review, page 22). (Results, page 6, lines 51-55): Did the authors ascertain whether the appropriate active comparator (as recommended in clinical practice) was used for these pivotal trials? It may be worthwhile to also include this if there is a discordance between the comparator chosen by the sponsor and what is recommended for use for these indications. (Results, page 7, lines 3-5): The coding for the primary endpoints to these outcomes should be included within Table 3. (Results, page 7, lines 10-13 and lines 15-17): Is this supposed to sum up to 52? It appears there are 53 here across the 4 categories (25+21+2+5) and across their statuses (27+7+3+1+8+4+3). Also, it is unclear from this section which are PMRs and PMCs – aggregate figures for each category should be listed here to make clear which are mandatory (PMRs) and which are voluntary (PMCs). (Discussion, page 7, lines 41-45): Without a clear comparison to the number of antibiotics prior to policy incentives, it seems these sentences should be modified. For this cohort, did development of these drugs begin or move more quickly from phase to another after certain policy incentives were introduced? Were special regulatory pathways sought after policy incentives were introduced? It's unclear in looking at this list if there is a clear causal pathway between the policy incentives and the findings in this piece. Similarly, it is unclear what the authors are comparing these findings to in saying "fewer pivotal trials,…,greater allowable losses." Was this cohort compared to the previously studied cohort from 2010-2015 and found to have fewer pivotal trials? (Discussion, page 8, lines 7-8): Further explanation is needed here from the authors, perhaps using specific examples from the cohort studied, of why validated indirect endpoints including clinician-reported or observer-reported outcomes is questionable in acute diseases. Are there cases where this may be appropriate? If not, why? (Discussion, page 8, lines 11-13): Table 3 currently does not include the categorization of direct or indirect outcomes or the further subcategories so it's unclear how costs relate to those drugs approved based on indirect outcomes. The authors might consider a stratified analysis looking at median costs or cost ratios for those drugs approved based on indirect outcomes versus those based on direct outcomes.

	(Discussion, page 9, lines 9-12): Theoretically, in reading through the statute and the authors' description of the PASTEUR Act, wouldn't the committee determining Critical Need Antimicrobials serve as an entity to deny this designation and therefore, such contracts if the drug does not have public health value or is the authors' recommendation that additional criteria regarding evidentiary standards be used in making this additional designation? Minor comments: (Introduction, page 4, lines 16-18): The authors should consider including mention of the other incentives made possible through the GAIN Act including one that relates to one of the outcome measures studied – eligibility to receive Fast Track designation if requested as well as Priority Review for a new qualified infectious disease product. (Methods, page 3, line 30): The word "approval" seems to be missing after FDA. (Methods, page 5, lines 20-21): This sentence should be modified to clarify that Post-Marketing Commitments are not "imposed" by the FDA as these are voluntary as explained later in the paragraph. (Methods, page 5, lines 44-50): To further clarify, was it only for pretomanid and secnidazole that an active comparator was not used within the pivotal trial? Noting that this was in the table, the authors should make clear that the total cost of the pretomanid, bedaquiline, and linezolid regimen as tested in the clinical trial was calculated. (Methods, page 6, lines 8-10): Were costs adjusted for inflation to one year? The authors should consider including mention of this within the main article. (Results, page 6, lines 20-32 and 33-41): Might the authors consider constructing tables to demonstrate these aggregated results, particularly for the second paragraph? Related to the results presented in the first paragraph, it may also be worthwhile to point out the number of antibacterial drugs that received simultaneous approval for multiple indications (e.g. cUTI and cIAI for instance for a few). (Results, page 6 and 7, lines 44-56, 3-5): The authors should consider summary tables to depict these results. (Results, page 7, line 8): The title should be changed to "Post-Marketing Requirements and Commitments." (Results, page 7, lines 21-23): Might the authors consider calculating the median cost and/or cost ratios for these antibacterial drugs (possibly excluding some outliers) or perhaps using a stratified approach in doing so with interquartile ranges? Additionally, including that these are all 2021 prices would be useful. (Discussion, page 7, lines 49-51): The authors should consider including the non-inferiority range determined in the prior study.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Comment	Author Response
2/29 “come with high price tags” – true for the first in class TB drugs, less true for the antibiotics for Gram-positive and Gram-negative bacteria, especially when compared to other drug classes. New (innovative) drugs will always be more expensive than the generic comparators, otherwise there will be no funding for R&D. Non-inferiority is indeed an issue, though.	Agreed. However, our aim was to compare new antibiotics to the older effective antimicrobial agent rather than other classes of drugs, approximating the choice clinicians may have to make. Other studies have pointed out that newer agents should improve patient outcomes to be of clinical value and justify higher prices.
4/7 The fatality figures are in the US only. Shouldn’t “Gram” be capitalized as a proper noun?	Left as is based on Cochrane guidance.
4/10 “reportedly slowed” – see the recent CID article by Dheman and FDA colleagues. Undoubtably, it has slowed; no need for the “reportedly” modifier that raises doubt.	Removed “reportedly” in line 66 on page 5 of the tracked draft and added citation 4 (Powers 2004).
4/14 Slow uptake is directly related to both the financial struggles (including bankruptcies) but also the issue you raise on non-inferiority. Connected issues.	Agreed.
5/3 Excellent work on these methods, continuing to build that PORTAL database	N/A
5/40 Is there a literature on the relationship b/w Red Book and actual prices?	Red Book provides WAC and AWP prices, but actual costs inclusive of rebates will vary. Since rebate agreements are private, it is difficult to ascertain the relationship between unit price and the ultimate cost to a payor. We have addressed this in lines 363 – 365 on page 18 of the tracked draft.
6/13 This list of “antibiotics” differs from some others in the published literature, for example by including the TB drugs and inhaled amikacin. You have also included a combination with a non-NME component and an mAb. Can you provide a more precise definition of your inclusion and exclusion criteria? (such as – all NME	We wished to examine all antibiotics approved by the FDA between October 2016 and December 2019. We used the FDA’s annual novel drug approvals list to identify these antibiotics, and excluded any that were approved solely based on

antibacterials J01 & J06…etc., systemic, as you think best). Otherwise, it is difficult to replicate the study for longer periods of time. This request may require you to re-run the results, or to give estimates of the impact of inclusion and exclusion criteria on the results and conclusions.	animal testing, including one for anthrax. This was the inclusion criteria for our cohort. We have edited lines 93-94 on page 6 of the tracked draft to address this.
6/52 If most of the NI margins were 10-15%, please report this in addition to the largest margin of 20%. What was the reason the regulatory allowed the wider margin?	In regards to the cefiderocol pivotal trial with a margin of 20%, the sponsor planned to sequentially test a 15% margin if the 20% margin was met. However, it was not used in the final efficacy analysis because the trial met criteria for declaring cefiderocol's statistical superiority. We based our analysis on the design of the trial and margin allowed at the time of patient enrollment and informed consent. We edited language in lines 234-236 on page 12 of the tracked draft to address this point.
7/40 There is a literature on NI trials in antibiotics, some of which could be cited in this discussion paragraph. The reader should know that others have written on this topic.	Added citations 23 and 24.
8/6 HPV is another example of indirect endpoints being useful and necessary.	Agreed. We did not add this point, cognizant of not expanding the paper's word count too much.
8/11 "unvalidated indirect endpoints" – this implies that the particular endpoints in these antibiotic pivotal trials were "unvalidated". If that is true, please explain.	We have added additional language in line 320-322 of page 16 of the tracked draft to address this.
8/20 "limited evidence of questionable rigor" – what has been shown about pretomanid after approval? What has been the experience at MSF, the Global Fund, TB Allliance, and others post-approval? You cast negative aspersions on the drug, which may or may not have been confirmed by subsequent evidence. See, e.g., https://pubmed.ncbi.nlm.nih.gov/34400340/.	Our intent with pretomanid was to highlight that evidence on approval was based on one single-arm trial that compared to a historical control. We agree that sometimes post-approval evidence does demonstrate the drug's effectiveness, but our study was limited to the nature of an antibiotic's evidence at the time of approval. Post-approval effectiveness data is sometimes not

	collected for many years after a drug is approved, so the level of evidence at the time of approval is a useful outcome measure.
8/42 Are you saying that plazomicin is more dangerous than other aminoglycosides that are effective?	The pre-approval evidence showed greater incidence of renal insufficiency with plazomicin compared to the control drug, not in relation to other aminoglycosides.
9/7 “new antibiotics meant to fill unmet medical needs lack confirmatory evidence that they do so” – you only present data from the pivotal trial, not the evidence present today. The FDA saw enough evidence to approve the drugs, so what do you mean by “confirmatory evidence”? Are you saying the FDA standards are too low?	No, we meant to refer to evidence from pivotal trials. We also highlight a concerning shift in evidence collection from the pre-approval to the post-approval period. We have updated language in lines 378-380 of page 18 and 19 of the tracked draft to address this.
9/12 Agreed that any new incentives must have stronger indicia of quality (satisfying unmet medical needs)	Thank you. No changes needed.

Reviewer #2

Overarching Comments: In this follow-on study to their prior analysis published in 2016, the authors characterized the pivotal clinical trials of antibiotics approved by the FDA between October 2016 and December 2019. In addition to extracting specific characteristics of these trials, the authors examined the launch prices for these antibiotics as well as the status of meeting post-marketing requirements or commitments requested by the FDA upon approval. The latter piece is a new addition to their previous work.

Like the previous article, the authors find that the pivotal trials for newly-approved antimicrobials have largely demonstrated uncertain efficacy and a majority are non-inferiority trials. They also found that no drug within their study cohort has completed or submitted its post-marketing requirements or commitments. Finally, they also found wide price variation for these drugs with some drugs being priced quite high. While the study is of a timely and interesting topic, there are areas for methodological improvement as detailed in the comments below.

Additionally, the lack of summary tables of the study’s key findings makes it difficult to follow the results and understand how the findings relate to each other. The study would be significantly strengthened if the authors make clear the connections between evidentiary standards for approval and the cost analyses for the drug. For instance, are there clear differences in the costs and costs ratios between the drugs that were approved based on non-inferiority margins versus those tested in a superiority trial? Additionally, did those drugs with direct versus indirect endpoints also have differences in terms of costs and cost ratios? Comparative conclusions are also drawn of these study findings, but it is unclear to what cohort the study findings are made to. Were these drugs approved based on fewer pivotal trials than all FDA approved drugs or the previous cohort studied?

Author Response: *Thank you for this feedback. We agree that further subanalyses would be interesting. However, the small sample size of our cohort may limit the usefulness of findings and comparisons based on sub-analyses. This concerned is heightened by the large number of noninferiority trials relative to superiority trials in our cohort, and also by the fact that all drugs were approved based on surrogate measures.*

There were more drugs in our cohort than in the previous cohort (2009-2015) studies, despite a shorter timeframe for approvals. This indicates an accelerated rate of antibiotic approval. Otherwise, both cohorts illustrate similar concerns in pre-approval testing that cast doubt on their clinical value.

Comment	**Author Response**
(Abstract, page 2, lines 19-20): The authors write that all drugs were approved based on changes to surrogate measures, but this does not seem to be reflected in the tables in the manuscript. In the actual manuscript (Results), it seems that this has been clarified further but should be edited in the abstract to reflect this finding.	Please see updates to Table 3 that address this.
(What this Study Adds, page 3, lines 17-18): Can we say for sure that policy incentives led to a steady increase of novel antibiotics without examining the prior period or is this a possible suggestion? Given the duration needed for antibiotic drug development, would the initiation of development for drugs within this cohort predated the incentives?	We have added language in lines 27 and 28 on page 5 of the tracked draft noting the number of antibiotics approved in previous decades to provide a better comparison.
(Methods, page 5, Lines 26-27): In reviewing the FDA Orange Book and the data descriptions (https://www.fda.gov/drugs/drug-approvals-and-databases/orange-book-data-files), it is unclear to me where there is information regarding Post-Marketing Requirements and Commitments that were used to confirm these for each antibacterial product. Could the authors point to more specifically where they extracted this information?	Each drug's PMR/PMCs were first identified using the drug's original approval letter. Status was pulled from the FDA's online database of PMRs/PMCs. See edited language in lines 148 -151 of pages 8 and 9 of the tracked draft.
(Methods, page 5, lines 28-29): Further details on how the authors categorized Post-Marketing Requirements and Commitments as "open" and "closed" as well as the other subcategories should be included as well as what database was used to do so. Did the authors verify this through ClinicalTrials.gov and through publication searches as well?	We used the status listed on the FDA database to categorize the PMRs/PMCs. We have edited lines 152-153 on page 9 of the tracked draft.

(Methods, page 5, lines 55-56): Could the authors clarify why the cost of treatment calculated based on dosage and administration used in the pivotal trials rather than the labels as the latter would more likely inform clinical practice? Were there significant differences between the comparator regimens used within the clinical trials and those in the label and/or as recommended in clinical practice? And to further clarify, was the regimen used within the pivotal trials or the label used to calculate the cost of approved antibacterial?	We used the comparators in the pivotal trials in the cost analysis for simplicity. When the trials compared to placebo or to a historical control, we used clinical recommendations to select an appropriate comparator. We did not systematically compare the comparator in the pivotal trials to those recommended in clinical practice. For the approved antibacterial, we used the regimen in the label – which was also frequently the regimen in the trial – to calculate cost of treatment.
(Results, page 6, lines 44-45): The authors should consider determining the median number of patients across trial arms for those pivotal trials that did have active or placebo comparators.	We chose to represent this as a range rather than as medians given the variability in the type of drug and indications studied.
(Results, page 6, lines 52-54): Further clarity on how the authors determined whether the trial design was that of a non-inferiority or superiority trial would be helpful – was this based on FDA's determination or that of the manufacturer? For example, with Recarbrio, the FDA stated that the trial for the cUTI indication were "not powered to demonstrate non-inferiority in the appropriate patient population" (Multi-Discipline Review, page 22).	We used the sponsor's stated trial hypothesis, pulled from the review documents on Drugs@FDA. Even when hypotheses were not clearly stated, FDA interpreted trial results as if the trial was designed as a non-inferiority trial.
(Results, page 6, lines 51-55): Did the authors ascertain whether the appropriate active comparator (as recommended in clinical practice) was used for these pivotal trials? It may be worthwhile to also include this if there is a discordance between the comparator chosen by the sponsor and what is recommended for use for these indications.	While the possibility that manufacturers would use less effective comparators to make their drug looks better is a well-known drug development strategy, we were less concerned about that possibility in this case because many antibiotics are tested in non-inferiority-design trials in which the comparator is therefore assumed to be highly effective.
(Results, page 7, lines 3-5): The coding for the primary endpoints to these outcomes should be included within Table 3.	Added to Table 3.

(Results, page 7, lines 10-13 and lines 15-17): Is this supposed to sum up to 52? It appears there are 53 here across the 4 categories (25+21+2+5) and across their statuses (27+7+3+1+8+4+3). Also, it is unclear from this section which are PMRs and PMCs – aggregate figures for each category should be listed here to make clear which are mandatory (PMRs) and which are voluntary (PMCs).	There was an error that we have now rectified. Updated numbers are on lines 246 – 257 on page 13 of the tracked draft.
(Discussion, page 7, lines 41-45): Without a clear comparison to the number of antibiotics prior to policy incentives, it seems these sentences should be modified. For this cohort, did development of these drugs begin or move more quickly from phase to another after certain policy incentives were introduced? Were special regulatory pathways sought after policy incentives were introduced? It's unclear in looking at this list if there is a clear causal pathway between the policy incentives and the findings in this piece. Similarly, it is unclear what the authors are comparing these findings to in saying "fewer pivotal trials,…,greater allowable losses." Was this cohort compared to the previously studied cohort from 2010-2015 and found to have fewer pivotal trials?	We wished to illustrate that policy incentives may have accelerated the rate of new antibiotics to market, but do not ensure that these antibiotics will add significant clinical value. Other than this, our current cohort was similar to the 2009-2015 cohort in that all drugs were approved based on at least one surrogate measure, several were approved based on only one pivotal trial, and the majority of trials were noninferiority trials. We have edited text in lines 280 – 290 on page 14 and 15 of the tracked draft to address this.
(Discussion, page 8, lines 7-8): Further explanation is needed here from the authors, perhaps using specific examples from the cohort studied, of why validated indirect endpoints including clinician-reported or observer-reported outcomes is questionable in acute diseases. Are there cases where this may be appropriate? If not, why?	Indirect endpoints are particularly useful when clinical outcomes may take a long time to study. This is particularly useful for indications such as oncology or other chronic conditions where the clinical endpoint, such as survival, occurs on a long timescale. Indirect endpoints are also useful when the surrogate has been shown to strongly correlate with clinical benefit. For acute diseases such as bacterial infections, clinical outcomes can be observed within a short period of time. Furthermore, surrogates for acute indications do not always reflect direct patient benefit. For example, we have published previously on how use of indirect assessment/biomarker of urine culture gives misleading "superior" results in trials when there is no demonstrated added benefit on patient centered outcomes of survival or symptoms.

	We have inserted this text into lines 305-311 on page 15 and 315-318 on page 16 of the tracked draft for clarity.
(Discussion, page 8, lines 11-13): Table 3 currently does not include the categorization of direct or indirect outcomes or the further subcategories so it's unclear how costs relate to those drugs approved based on indirect outcomes. The authors might consider a stratified analysis looking at median costs or cost ratios for those drugs approved based on indirect outcomes versus those based on direct outcomes.	We added this information to Table 3. As all drugs were based on indirect outcomes, we were not able to do a cost analysis comparing drugs approved on direct endpoints to those based on indirect endpoints.
(Discussion, page 9, lines 9-12): Theoretically, in reading through the statute and the authors' description of the PASTEUR Act, wouldn't the committee determining Critical Need Antimicrobials serve as an entity to deny this designation and therefore, such contracts if the drug does not have public health value or is the authors' recommendation that additional criteria regarding evidentiary standards be used in making this additional designation?	A drug for a high-need indication may have significant potential for public health impact, but its actual value depends on how rigorously it is tested and the results from those trials. Our point is that only demonstrated added benefits compared to available standards of care on direct patient outcomes should be the standard for payment of the bonus under the PASTEUR Act and should be demonstrated in the pivotal randomized trials forming FDA approval. No committee is needed if the evidence is clear. We've added some text in lines 378-380 on page 18 and 19 of the tracked draft to clarify this.
(Introduction, page 4, lines 16-18): The authors should consider including mention of the other incentives made possible through the GAIN Act including one that relates to one of the outcome measures studied – eligibility to receive Fast Track designation if requested as well as Priority Review for a new qualified infectious disease product.	Added.
(Methods, page 3, line 30): The word "approval" seems to be missing after FDA.	Addressed.
(Methods, page 5, lines 20-21): This sentence should be modified to clarify that Post-Marketing	Addressed.

Commitments are not “imposed” by the FDA as these are voluntary as explained later in the paragraph.	
(Methods, page 5, lines 44-50): To further clarify, was it only for pretomanid and secnidazole that an active comparator was not used within the pivotal trial? Noting that this was in the table, the authors should make clear that the total cost of the pretomanid, bedaquiline, and linezolid regimen as tested in the clinical trial was calculated.	No, several of the drugs’ pivotal trials used placebo controls. For these, we used clinical recommendations to determine an appropriate cost comparator. To address the second point, we have made edits in lines 169 – 171 on page 9 of the tracked draft.
(Methods, page 6, lines 8-10): Were costs adjusted for inflation to one year? The authors should consider including mention of this within the main article.	We did not adjust for inflation, which was minimal during this short time period.
(Results, page 6, lines 20-32 and 33-41): Might the authors consider constructing tables to demonstrate these aggregated results, particularly for the second paragraph? Related to the results presented in the first paragraph, it may also be worthwhile to point out the number of antibacterial drugs that received simultaneous approval for multiple indications (e.g. cUTI and cIAI for instance for a few).	To the first point, Table 2 (Regulatory Overview) illustrates drug development milestones, with the text serving as the analysis of development timelines. To the second point, we agree, and have added text in lines 206 – 208 on page 11 of the tracked draft.
(Results, page 6 and 7, lines 44-56, 3-5): The authors should consider summary tables to depict these results.	We hope that Table 2 adequately summarizes our results, with analysis in the text.
(Results, page 7, line 8): The title should be changed to “Post-Marketing Requirements and Commitments.”	Addressed.
(Results, page 7, lines 21-23): Might the authors consider calculating the median cost and/or cost ratios for these antibacterial drugs (possibly excluding some outliers) or perhaps using a stratified approach in doing so with interquartile ranges? Additionally, including that these are all 2021 prices would be useful.	We have added text on line 168 on page 9 of the tracked draft to clarify that we used 2020 prices for our cost analysis. Costs and cost ratios are shown as ranges from the minimum and maximum days that a patient should take the drug. Given that these data are given as ranges, we did not feel it would be useful to the reader to provide medians.

(Discussion, page 7, lines 49-51): The authors should consider including the non-inferiority range determined in the prior study.	Added in line 297 of page 15 of the tracked draft.

VERSION 2 – REVIEW

REVIEWER 1	Outterson, Kevin; Boston University. Competing Interest: I am the principal investigator for CARB-X, which makes grants to antibacterial R&D world wide. We are funded by three governments (US, UK, Germany) and two foundations (Wellcome Trust and the Bill & Melinda Gates Foundation). We have no financial conflict of interest in any antibacterial company whatsoever.
REVIEW RETURNED	20-Apr-2022

GENERAL COMMENTS	This is my second review of this paper. The authors have carefully considered my prior comments and I am satisfied with their responses for the purposes of peer review. I will limit my comments accordingly. Abstract, line 36 (using the cumulative line numbers): instead of "allowing lesser efficacy of up to 20%," I suggest "allowing non-inferiority margins ranging from X% to Y%." In many cases, the actual efficacy data from the pivotal study includes a range of positive (more effective than comparator) and negative (less effective than comparator) values, with a range required due to the sample size. Line 64: One could update the 2019 CDC numbers with the more comprehensive estimates published in The Lancet earlier this year by IHMA. Line 84: "large up-front payments for new antibiotics in addition to per-prescription payments." The Ref is to the PASTEUR Act, which calls for annual subscription payments over a decade in lieu of any per-prescription payment by any US government buyer. Line 210: median of 8.2 years IND-->FDA approval is the same result obtained by Dheman et al in CID 2020. Worth a reference to their paper as confirmation. Interesting that Dheman et al report shorter development times in prior decades, before the various Priority Review, Fast Track, QIDP and other designations you mention. Lines 251-252: "No study drug has either submitted or fulfilled all its PMCs." It would be helpful to know if the FDA expected these recently approved drugs to have fulfilled all of their PMCs by now, so we could understand if this was poor or acceptable performance. Given the natural time it might take for PMCs, it might be more instructive to know whether the PMCs from your previous study (drugs approved 2010-2015) have been completed by now. Line 337: Typo - "as such" is repeated.

REVIEWER 2	Morris, Julie. Competing interest: None

REVIEW RETURNED	24-Apr-2022

GENERAL COMMENTS	This descriptive study outlines the published clinical evidence relating to antibiotics approved by the FDA during the period Oct 2016-Dec 2019. The details of the regulatory background of the 15 new antibiotics identified, and the associated pivotal trials are relatively well presented. However, there are few issues relating to the comparators and costing process, and the interpretation of the data that should be addressed. 1. The comparator regimes in the pivotal trials were used in the cost comparisons (rather than the clinically recommended regimes). Would these differ greatly? More details should be provided. 2. The definition of 'direct' vs 'indirect' outcomes needs to be clarified, and also how this is applied to the pivotal trials (Table 3). On Page 8, it is stated that direct endpoints "…include survival and patient-reported outcomes (PROs)". But, in Table 3, some primary endpoints are described as survival or PROs, and these are designated 'indirect'. Also, in the Abstract reference to 'surrogate measures' are made, which again needs to be clarified. 3. Page 15. "Non-inferiority hypotheses enable smaller trials…" This rather general statement requires clarification. Sample size for non-inferiority trials is heavily dependent on the size of the non-inferiority margin. It is the choice of this margin, in comparison with a detectable difference for a superiority study, which determines the relative difference in study size. In fact, I would say that non-inferiority trials are more likely to have large sample sizes than placebo-controlled trials, if designed correctly. 4. Some of the Tables are rather complex and unwieldy. Would it not be more appropriate to include some of these in an Appendix and have a more succinct summary Table in the main paper? 5. Abstract. "New antibiotics have been approved in recent years mostly based on fewer, smaller, and non-inferiority pivotal trials….". Greater justification for the comparative element of this statement needs to be provided. Is a comparison with the results of the 2010-2015 cohort of 8 antibiotics is being made here? Reference to this very small cohort is made in the Discussion, but few details are given. In fact, a greater proportion of the pivotal trials for the earlier cohort were non-inferiority trials (7 of the 8 drugs as compared to 15 of the 27 drugs in the present cohort), with one of the 7 drugs in the 2010-2015 cohort having pivotal trials with, "…large margins of inferiority exceeding 15%-20%". Also, "….None used patient mortality or direct measures of patient disability as primary end points….." Thus, there appears to be little difference between the two cohorts. Hence is this a comparison with an even earlier cohort?

REVIEWER 3	Ramachandran, Reshma; Yale School of Medicine, Yale National Clinician Scholars Program. Competing Interest: I serve on the boards in an unpaid position for Universities Allied for Essential Medicines North America and the Ameican Medical Student Association Foundation. I also serve as the Chair of the DFA FDA

	Task Force as an unpaid volunteer - the work of the Task Force is funded by Arnold Ventures. I have also collaborated with one of the co-authors, Dr. John Powers on various advocacy and policy activities related to the FDA and the approval of novel antimicrobials through the DFA FDA Task Force.
REVIEW RETURNED	10-May-2022

GENERAL COMMENTS	In this follow-on study to their prior analysis published in 2016, the authors characterized the pivotal clinical trials of antibiotics approved by the FDA between October 2016 and December 2019 as an addition to their previous work examining approvals of antibiotics from 2010 to 2015. Extracting the characteristics of pivotal clinical trials, the authors found more than half used an active control, non-inferiority design. In reviewing the post-market requirements and commitments of these antibiotics upon approval, they found that very few were fulfilled or submitted. Moreover, in examining the launch prices of these drugs, they found wide price variation with some drugs being priced quite high. I had reviewed this article previously for the BMJ and am pleased to review a revision to this manuscript for BMJ Medicine. I appreciate the authors' attention to my prior comments in their revisions, many of which have been addressed in this current version. Overall, while an important study addressing a timely area of interest, particular areas of the study could be strengthened particularly around demonstrating evidence to support the conclusions made in comparing this cohort's characteristics to that of the previously studied cohort of antimicrobials. More specific comments are as follows: Major Comments: (Abstract, page 2, lines 36-37): While the authors state that this has been addressed with changes in Table 3, the previous comment raised where they wrote that all drugs were approved based on changes to surrogate measures does not seem accurate. For instance, for the antibiotic, delafloxacin, two pivotal trials are noted that supported FDA approval and are characterized as clinician reported outcomes that are categorized in the manuscript to be indirect endpoints, which seems to be separate from surrogate measures. Or are the authors stating that all indirect endpoints are being considered surrogate measures? Given the specific definition of surrogate by the FDA, the authors should consider clarifying this to be indirect endpoints within the manuscript. (Abstract, page 2, lines 43-45): The authors conclude that "new antibiotics have been approved in recent years mostly based on fewer, smaller, and non-inferiority pivotal trials that commonly use surrogate measures", implying there is a comparison group of antibiotics with a larger number of pivotal trials enrolling a larger number of patients. Is this meant to be compared to the antibiotics previously approved? (Results, page 12, lines 231-234): Regarding the specific example of Recarbrio – as FDA stated that both pivotal trials for cUTI and cIAI were "not powered to demonstrate non-inferiority", they did not consider these pivotal trials in making their approval decision. For cIAI, the drug also failed to meet the sponsored-designated non-inferiority margin. A caveat regarding this approval (and others like	

	this) should be considered as although FDA examined the non-inferiority margins, they did not consider these studies as supporting FDA approval. (Discussion, page 14, lines 279-281): Further clarity on the findings of the prior study is needed to demonstrate the similarities between the pivotal trial characteristics from the prior study to this present cohort. (Discussion, pages 14, lines 282-285, 289-291): These sentences now compare the results of the evidentiary basis for approval of this cohort to the findings from the previous study, but no comparison has been conducted as part of the analysis to demonstrate this conclusively. Moreover, it is unclear how the authors came to the conclusion that they found non-inferiority trials being used "often in serious diseases" as this was not mentioned in the Results that certain indications were considered serious or not. (Discussion, page 15, line 302): The authors seem to be using indirect and surrogate endpoints interchangeably, which veers away from how FDA defines what a surrogate endpoint is. This may introduce confusion and could mislead readers on the specific endpoints used in this cohort's pivotal trials. Tables – While the Tables include valuable, detailed information about the antimicrobials within this recent cohort across a number of areas of analysis, a summary table of key characteristics for the pivotal trails as well as comparisons detailed in the Discussion between this cohort and the previously studied cohort would be valuable. Minor Comments: (Introduction, page 5, lines 70-72): The authors could add a clause here that it is not only the cost of drug development, but the also the limited and potentially, unexpected returns especially as these are drugs meant to be conserved. (Introduction, pages 5 and 6, lines 82-83): CMS has already implemented the rule of as 2019 to provide additional supplemental payments for new antimicrobials and remove the criteria of "substantial clinical improvement." As the authors are aware, efforts are underway to codify this agency change in legislation and increase the supplemental payments. This might be revised to note this here. (Methods, page 7, lines 114): Technically, the Orphan Drug Act is not an expedited regulatory designation and is a separate designation qualifying for additional incentives.	

VERSION 2 – AUTHOR RESPONSE

Reviewer #1:

Comment	Author Response
Abstract, line 36 (using the cumulative line numbers): instead of "allowing lesser efficacy of up to 20%," I suggest "allowing non-inferiority margins ranging from X% to Y%." In many cases, the actual	In the abstract, we are referring to the non-inferiority hypothesis, not the results of the

efficacy data from the pivotal study includes a range of positive (more effective than comparator) and negative (less effective than comparator) values, with a range required due to the sample size.	study, which we agree could range widely. To help avoid confusion, we deleted the phrase.
Line 64: One could update the 2019 CDC numbers with the more comprehensive estimates published in The Lancet earlier this year by IHMA.	Thank you for the suggestion. We are aware of the Lancet paper, but believe that it integrated too many assumptions into its modeling and definitions of key terms like 'resistance' to be reliable for our purposes. We believe the CDC reference is still the proper one for this circumstance.
Line 84: "large up-front payments for new antibiotics in addition to per-prescription payments." The Ref is to the PASTEUR Act, which calls for annual subscription payments over a decade in lieu of any per-prescription payment by any US government buyer.	We have now included a more explicit reference to the PASTEUR Act and its current status in lines 88-91 of the tracked draft.
Line 210: median of 8.2 years IND-->FDA approval is the same result obtained by Dheman et al in CID 2020. Worth a reference to their paper as confirmation. Interesting that Dheman et al report shorter development times in prior decades, before the various Priority Review, Fast Track, QIDP and other designations you mention.	We have now referenced the Dheman et al. study on page 15 of the tracked draft.
Lines 251-252: "No study drug has either submitted or fulfilled all its PMCs." It would be helpful to know if the FDA expected these recently approved drugs to have fulfilled all of their PMCs by now, so we could understand if this was poor or acceptable performance. Given the natural time it might take for PMCs, it might be more instructive to know whether the PMCs from your previous study (drugs approved 2010-2015) have been completed by now.	It's difficult to understand the trajectory of PMRs and PMCs without obtaining status reports from the FDA through Freedom of Information Act (FOIA) requests. They are removed from the online data one year after they are either released or fulfilled. This is a great topic for a follow-up investigation, especially given that the pre-approval evidence for recent antibiotics leaves evidence gaps that postmarket studies can answer.
Line 337: Typo - "as such" is repeated.	We have made this edit.

Reviewer #2:

Comment	Author Response
The comparator regimes in the pivotal trials were used in the cost comparisons (rather than the clinically recommended regimes). Would these differ greatly? More details should be provided.	Agreed. We discuss this in lines 172 – 185 of the tracked draft. The trial sponsor's rationale for choosing a given comparator was not always evident from approval packages on Drugs@FDA.
The definition of 'direct' vs 'indirect' outcomes needs to be clarified, and also how this is applied to the pivotal trials (Table 3). On Page 8, it is stated that direct endpoints "…include survival and patient-reported outcomes (PROs)". But, in Table 3, some primary endpoints are described as survival or PROs, and these are designated 'indirect'. Also, in the Abstract reference to 'surrogate measures' are made, which again needs to be clarified.	We have included definitions of direct and indirect outcomes in the methods section (lines 128-145 of the tracked draft) as well as in the legend of Table 3. Patient Reported Outcomes that measure signs of disease (e.g. number of episodes of diarrhea but not abdominal pain or cramping in diarrheal illness) are still indirect measures even though information captured directly from patients. (PROs refer to a method of how information is obtained and not necessarily what is measured.) We have noted in Table 3 when the PRO was a measure of signs of disease. We have added a citation to support these definitions (new reference 16).
Page 15. "Non-inferiority hypotheses enable smaller trials…" This rather general statement requires clarification. Sample size for non-inferiority trials is heavily dependent on the size of the non-inferiority margin. It is the choice of this margin, in comparison with a detectable difference for a superiority study, which determines the relative difference in study size. In fact, I would say that non-inferiority trials are more likely to have large sample sizes than placebo-controlled trials, if designed correctly.	This is a good point. We have edited the text in lines 308-311 of the tracked draft to reflect this point.
Some of the Tables are rather complex and unwieldy. Would it not be more appropriate to include some of these in an Appendix and have a more succinct summary Table in the main paper?	We have included Appendices with expanded information and calculations to accompany our summary tables. We condensed Table 3, cutting down 2-3 pages, and transferring any additional relevant information in another Appendix.
Abstract. "New antibiotics have been approved in recent years mostly based on fewer, smaller, and non-inferiority pivotal trials….". Greater justification for the comparative element of this statement needs	This is a good point. Our intention with this statement was not to compare to the earlier cohort of 2010-2015 antibiotic approvals, but rather to illustrate that the trends identified for

to be provided. Is a comparison with the results of the 2010-2015 cohort of 8 antibiotics is being made here? Reference to this very small cohort is made in the Discussion, but few details are given. In fact, a greater proportion of the pivotal trials for the earlier cohort were non-inferiority trials (7 of the 8 drugs as compared to 15 of the 27 drugs in the present cohort), with one of the 7 drugs in the 2010-2015 cohort having pivotal trials with, "…large margins of inferiority exceeding 15%-20%". Also, "….None used patient mortality or direct measures of patient disability as primary end points….." Thus, there appears to be little difference between the two cohorts. Hence is this a comparison with an even earlier cohort?	that cohort are also observed in our new study.

Reviewer #3:

Comment	**Author Response**
(Abstract, page 2, lines 36-37): While the authors state that this has been addressed with changes in Table 3, the previous comment raised where they wrote that all drugs were approved based on changes to surrogate measures does not seem accurate. For instance, for the antibiotic, delafloxacin, two pivotal trials are noted that supported FDA approval and are characterized as clinician reported outcomes that are categorized in the manuscript to be indirect endpoints, which seems to be separate from surrogate measures. Or are the authors stating that all indirect endpoints are being considered surrogate measures? Given the specific definition of surrogate by the FDA, the authors should consider clarifying this to be indirect endpoints within the manuscript.	Clinician reported outcomes measure observable signs of disease fall under the definition of "surrogate endpoints". The April 1992 Federal Register notice that first articulated accelerated approval includes signs of disease as surrogate endpoints. Signs of disease are not direct measures of how patients feel, function or survive. In this sense, all indirect measures are "surrogates" since the assumption is they are used as replacements to reflect direct measures of patient outcomes. We have further clarified in the methods section (lines 128-145 of the tracked draft).
(Abstract, page 2, lines 43-45): The authors conclude that "new antibiotics have been approved in recent years mostly based on fewer, smaller, and non-inferiority pivotal trials that commonly use surrogate measures", implying there is a comparison group of antibiotics with a larger number of pivotal trials enrolling a larger number of patients. Is this meant to be compared to the antibiotics previously approved?	This is a good point, and one that Reviewer #2 also flagged. Our intention with this statement was not to compare to the earlier cohort of 2010-2015 antibiotic approvals, but rather to illustrate that the trends identified for that cohort are also observed in our new study.

(Results, page 12, lines 231-234): Regarding the specific example of Recarbrio – as FDA stated that both pivotal trials for cUTI and cIAI were “not powered to demonstrate non-inferiority”, they did not consider these pivotal trials in making their approval decision. For cIAI, the drug also failed to meet the sponsored-designated non-inferiority margin. A caveat regarding this approval (and others like this) should be considered as although FDA examined the non-inferiority margins, they did not consider these studies as supporting FDA approval.	We agree the approval of imipenem/cilastatin/relebactam had little evidence to support it and the FDA review states the studies were not adequate and well-controlled, yet still deemed the data sufficiently “adequate” (p 21) for approval. The two studies were dose-ranging studies of high vs. lower doses of relebactam combined with imipenem vs imipenem plus placebo. The FDA review on p. 93 stated the cUTI study was designed with a -15% NI margin by the sponsor and had 87% power for this comparison, even though FDA guidance states a -10% NI margin is “acceptable”. The results show a lower bound of the 95% CI of -18.9% in the FDA analysis. For cIAI, the FDA review stated on p 113 that the study had 80% power and a NI margin of -15% (chosen by the sponsor). The FDA analysis showed a lower bound of the 95% CI of -8.8%. This information has been added to the Table 3 legend.
(Discussion, page 14, lines 279-281): Further clarity on the findings of the prior study is needed to demonstrate the similarities between the pivotal trial characteristics from the prior study to this present cohort.	We have added text in lines 293-297 of the tracked draft to address this point.
(Discussion, pages 14, lines 282-285, 289-291): These sentences now compare the results of the evidentiary basis for approval of this cohort to the findings from the previous study, but no comparison has been conducted as part of the analysis to demonstrate this conclusively. Moreover, it is unclear how the authors came to the conclusion that they found non-inferiority trials being used “often in serious diseases” as this was not mentioned in the Results that certain indications were considered serious or not.	Our hope was to show a continuation in trends that we first identified in the previous cohort, rather than draw a direct comparison between the two. We have edited text in lines 293-297 to address the second point.
(Discussion, page 15, line 302): The authors seem to be using indirect and surrogate endpoints interchangeably, which veers away from how FDA defines what a surrogate endpoint is. This may introduce confusion and could mislead readers on	FDA does define “surrogate” endpoints as both laboratory tests and signs of disease as noted above in the April 1992 FR notice. We agree FDA’s application of these terms is inconsistent (including issuing full approvals

the specific endpoints used in this cohort's pivotal trials.	using unvalidated surrogate endpoints), which is why we have used the terms "direct and indirect" to avoid confusing terminology around types of "surrogates". We have defined these further now in the methods and legend of Table 3.
Tables – While the Tables include valuable, detailed information about the antimicrobials within this recent cohort across a number of areas of analysis, a summary table of key characteristics for the pivotal trails as well as comparisons detailed in the Discussion between this cohort and the previously studied cohort would be valuable.	We will condense the pivotal trial table for the paper, and include additional details in an appendix.
Minor Comments: (Introduction, page 5, lines 70-72): The authors could add a clause here that it is not only the cost of drug development, but the also the limited and potentially, unexpected returns especially as these are drugs meant to be conserved.	Agreed. Added in lines 75-77 of the tracked draft.
(Introduction, pages 5 and 6, lines 82-83): CMS has already implemented the rule of as 2019 to provide additional supplemental payments for new antimicrobials and remove the criteria of "substantial clinical improvement." As the authors are aware, efforts are underway to codify this agency change in legislation and increase the supplemental payments. This might be revised to note this here.	We have edited the text on lines 87-91 of the tracked draft to address this point.
(Methods, page 7, lines 114): Technically, the Orphan Drug Act is not an expedited regulatory designation and is a separate designation qualifying for additional incentives.	Good point. We have adjusted the language at various points in the draft to "special regulatory pathway" to acknowledge this nuance.