We thank the Editorial committee and reviewers for their comments and for their patience while we made amendments. We are very grateful to have received such a thorough, robust review with many detailed comments that we feel have improved the quality of this manuscript.

Please find below a point-by-point response (red).

EDITORIAL COMMITTEE:

1. Can ‘innovative’ be replaced with ‘new’ medical devices or just leave it at medical devices? At certain points, innovative is equated to novel (no previous clinical study) which may be misleading.

We thank the Editorial committee and agree that the term ‘new’ may be more appropriate than ‘innovative’. We have amended throughout the manuscript.

2. Please clarify throughout the paper that 510K is a ‘clearance’ and not approval.

We agree and have clarified throughout the paper that 510(k) is a clearance and not approval.

3. Please elaborate why clearance through 510K pathway is a problem. It could be that this was a minor modification and did not require a clinical study as per FDA norms. We don’t really know how incremental a device may have been and if 510K approval is indeed appropriate in those instances. Do you advocate changing that?

We thank the Editorial committee and now extend our discussion to emphasise the principal findings of our work. We restricted our analysis to clinical studies of new medical devices reported in the biomedical literature. Our analysis is therefore likely favoured more novel devices, which clinicians might have thought warranted publication in the biomedical literature. Nonetheless, even devices that were comparable to predicate devices, which would be appropriate for the 510(k) pathway, often had no published articles supporting their use on their release to the marketplace. This finding dovetails the work of other authors, who found that there is little scientific evidence to support the use of many medical devices, making shared informed decision making difficult or impossible.

Discussion, Line 222: “We identified a multitude of new medical devices in clinical studies, almost half of which received regulatory approval. The 510(k) pathway was most commonly used, and devices often received regulatory clearance before the first published clinical study. The corollary is that many devices cleared for use in patients had no clinical data accessible in the literature to support their use. Published clinical studies were mostly case series’ comprising Level 4 evidence. Without high quality clinical data available, informed shared decision-making on the use of new medical devices is difficult if not impossible.”

Discussion, Line 254: “In keeping with the present study, several other groups have also found limited publically available evidence to support the regulatory clearance and approval of new devices. Zuckerman et al evaluated the types of scientific evidence used to support devices cleared using the 510(k) pathway.[5] Of the 50 devices included, 8 had data to support the claim they were substantially equivalent to a predicate device, and only 3 had data on safety or effectiveness. Chang et al found that even devices approved using the PMA pathway, which require considerably more scientific evidence, often had no published clinical trials.[21] When trials are published, comparators are often absent, and details may differ substantially from the data submitted to the FDA.[21 22]”

4. You started with clinical studies to identify devices that did not receive approval. Might we have more details of non-approved devices? Also, for instance, in the supplementary table listing devices that received regulatory approval, might you instead list all studies/devices included and add a column of whether regulatory approval was received or not?

We agree that more information on the characteristics of medical devices that did not receive regulatory clearance or approval would be helpful, and now provide this in Table 1.

5. It is not clear how you identify some devices cleared through 510K that did not have a prior published clinical study. If you start from clinical studies, how did you

identify the reverse, i.e. approved devices with no prior published clinical study? What is the time period over which you searched FDA databases? Is it even before 2000 and so you identified 'cleared' devices and a clinical study that followed it?

Search results were not limited to a date range, allowing for the identification of regulatory clearance or approval prior to first published clinical study. We have now clarified this in the manuscript.

> Methods, Line 177: "Search results were not limited to a date range, allowing for the identification of regulatory clearance or approval before the first published clinical study. All the searches were performed in August 2015, allowing a minimum of 10 years from publication to regulatory clearance or approval. Discrepancies were resolved by consensus."

6. Is the comparison to drugs approval necessary? We wonder if it may be removed throughout.

We agree this was confusing and have now removed the comparison to drug approval throughout the manuscript.

7. The conclusions are still overstating the importance of working with industry. Industry picks and chooses when to collaborate and they get involved when there is a greater potential for commercialization or when they know they will be successful. This is a serious bias that can't be adjusted for. The conclusions are still not impartial. I think the conclusions need to reflect the limitation of the data not just put in limitations that few people read.

On reflection, we agree that we may have overstated the importance of industry collaboration. Our finding that industry collaboration was associated with a greater rate of regulatory approval – while in our opinion certainly meriting further research – remains subject to bias. To this end, we have completely removed this analysis from the abstract, and from the conclusions of the manuscript itself.

REVIEWER 1

1. The authors have satisfactorily addressed all of my comments and the revised

manuscript is significantly improved.

We thank Dr. Drozda for their kind comments.

REVIEWER 2:

1. This paper provides a descriptive analysis of the early clinical testing reported in the literature for novel medical devices. This is an interesting and well-designed study, although the authors could better showcase the findings of greatest clinical and policy import and better contextualize these findings. Below are a few major and specific comments that the authors could consider in revising their manuscript.

We thank Dr. Kesselheim for their kind comments and have addressed their suggestions below.

2. The authors report that the (limited) clinical testing reported in the literature is often not accessible before new devices are cleared or approved by the US FDA. This is a novel finding that merits more attention in the text and that extends the work by Zuckerman et al. (JAMA Intern Med 2014;174:1781), which found that scientific data to support 510(k) device clearances is frequently unavailable in the mandatory FDA summaries, and Chang et al. (BMJ 2015;350:h2613), which found that the key study results for many of the highest-risk cardiovascular devices remain unpublished post-approval.

We thank the reviewer and agree that the lack of publically available data before new devices are cleared or approved by the FDA is among the principal findings of our work. To this end, we have extended the discussion to emphasize this point, including the aforementioned references. (See Response to Editorial Committee, Point 3)

3. The finding mentioned in point 1 would be made stronger if the authors could characterize the type of studies they identified for each of the approved devices. How many patients did these studies enroll on average? What was the study design (randomization, masking, controls)? Were larger, higher-quality studies conducted after approval, or were the initial studies the best that clinicians got? This would provide some sense for the quality of the evidence available when these devices enter the market. Some relevant cites: Dhruva et al. (JAMA 2009;302:2679); Hwang et al.

(BMJ 2014;348:g217); Rathi et al. (JAMA 2015;314:604)

We thank the reviewer for this important comment. We have now characterised the type of studies identified for each of the approved items in the supplement, and also mention these in the Results and Discussion section. As might be expected, published clinical studies were mostly case series' comprising Level 4 evidence.

> Results, Line 209: "Published clinical studies of devices that received regulatory clearance or approval were mostly case series' comprising Level 4 evidence (89/99; 89.9%)."
>
> Discussion, Line 222: "We identified a multitude of new medical devices in clinical studies, almost half of which received regulatory approval. The 510(k) pathway was most commonly used, and devices often received regulatory clearance before the first published clinical study. The corollary is that many devices cleared for use in patients had no clinical data accessible in the literature to support their use. Published clinical studies were mostly case series' comprising Level 4 evidence. Without high quality clinical data available, informed shared decision-making on the use of new medical devices is difficult if not impossible."

4. The number of studies collected for PMA devices (n=17) seems very low. Chang et al. report a publication rate for cardiovascular PMA devices of 57% (60/106). During the study period (2000-2004), the FDA approved 216 new PMA devices, which suggests up to 100+ trials should have appeared in the literature. Even if some of these are excluded due to studies published before 2000 or after 2004, and even if there are varying rates of publication in non-cardiovascular specialties, there is still quite a big gap between the authors' figure and the back-of-the-envelope calculation above. To validate their search strategy, the authors could cross-check their results against 1 or 2 years-worth of PMA approvals (i.e., take the ~50 PMA devices approved in 2002, search each for publications, and determine whether any were missed)

We thank the reviewer for their perspicacity, and have spent some time investigating this. While the number of devices that received PMA approval did significantly

exceed the number of relevant publications identified, many of these devices were unlikely to warrant publication in a peer-reviewed journal because they were either not considered interesting or important (e.g., contact lenses), or represented part of a larger system or device. We do acknowledge this bias in the limitations section of the manuscript.

> Discussion, Line 275: "Our analysis may also have favoured more novel devices, which clinicians might have thought warranted publication in the biomedical literature. The proportion of devices cleared through the 510(k) pathway was therefore likely to be an underestimate."

5. I would suggest revising the title, which currently does not convey the key aim or results of the study (in particular making clear the focus was early / first human clinical testing)

We agree that the title could be clearer and now use the term 'new medical device' rather than 'innovative medical device' as per the Editorial Committee's suggestion.

6. Page 7, line 112 – suggest rephrasing to emphasize that the US has a central regulatory body for premarket device review, whereas the EU and other jurisdictions lack such authority

We have now highlighted the fact that the USA has a central regulatory body in the discussion.

> Discussion, Line 278: "We determined whether a device had regulatory approval using only the FDA medical device databases. The proportion of medical devices receiving regulatory approval was therefore also undoubtedly an underestimate, in particular it is likely that licenses were granted from the European Union which does not require any evidence of clinical value.[11] The reason for selecting the FDA, rather than other licensing authorities, was because the FDA provides public databases and search engines that allowed for a systematic search strategy, the FDA acts as the central body for all medical devices receiving regulatory approval in the USA, and the USA represents the largest medical device market in the world. We hypothesise that

most of the manufacturers of devices that received regulatory approval from another jurisdiction would have ultimately sought and obtained FDA approval within the timeframe of this study if they were successful."

7. Page 7, line 121 – an important caveat is that other studies have pointed to the importance of academic and public-sector research in transformative drug therapies. See Stevens et al. (NEJM 2011;364:535) and Kesselheim et al. (Health Aff 2015;34:286)

We have now removed the comparison to drug approval as per the Editorial Committee's suggestion (See Response to Editorial Committee, Point 6).

8. Page 8, line 142 – there needs to be more information on how the academic / industry tag or "relationship" was determined. Was it on the basis of affiliation only, any financial disclosure, provision of the device/intervention, or some combination of the above? (repeated for page 9, line 179; page 13, line 290)

We apologise for this oversight and have now clarified that a relationship included relevant author affiliation, financial support or provision of technology.

Methods, Line 159: "For each clinical study of a new medical device we determined the type of device, the target specialty, the involvement of academia, and the involvement of industry (HJM and CJP, checked by AHH and APM). The types of device were based on the FDA definition and the target specialties were drawn from the FDA databases. We considered academia and industry to be involved in the development of a device if relevant author affiliation, financial support, or provision of technology was described in the author affiliations, main text, or acknowledgements of the article. Discrepancies were resolved by consensus."

9. Page 10, line 192 – I presume the authors also checked search engines and other public sources since some devices may have been discontinued, withdrawn, or recalled and would not currently appear in the FDA database

We are grateful to the reviewer for raising this, and can confirm that we did perform supplementary searches using Google for devices that may have been discontinued,

withdrawn, or recalled. This has now been added to the methodology.

> Methods, Line 176: "…We also searched Google™ (Google Inc., California, USA) for devices that may have been discontinued, withdrawn, or recalled."

10. The Results section was somewhat difficult to follow, and the presentation of the findings in this section did not seem to mirror the findings highlighted in the abstract or Discussion. I'd recommend reorganizing around the key points

The subtitles correspond to those in the methodology and we have amended the Results section such that it better highlights the principal findings.

11. Page 10, line 207 – the most frequent journals of publication seemed much less relevant (and could be moved to an appendix) than the most frequent therapeutic areas (cardiovascular, orthopedic, general / plastic surgery, as shown in Table 1)

We agree that the most frequent journals are much less relevant and, in any case, these are now listed in the supplement. We have therefore removed this from the Results section.

12. Page 11, line 217 – should the denominators here be 99 instead of 218?

We agree with the reviewer that representing the number of 510(k), PMA, and HDA as a proportion of all approved devices would be helpful, and have now amended.

13. Page 11, line 217 – what was the range of approval years, or when was the latest approval for one of the studied devices?

We thank the reviewer for highlighting this and have now provided this data in the manuscript.

> Results, Line 204: "Regulatory clearance or approval was granted between April 1997 and September 2014."

14. Page 11, line 219 – what was the lag between approval and publication specifically for the 43 devices that were approved before any clinical studies?

We thank the reviewer for highlighting this and have now provided this data in the manuscript.

> Results, Line 204: "The median lag between publication of the clinical study and regulatory clearance or approval was 2 months (interquartile range -10.8 months to 26.3 months). Of these, 43 devices (43/99; 43.4%) were actually cleared or approved before a clinical study was published; the median lag in these devices was -12.5 months (interquartile range -23.3 months to -6.3 months)."

15. Page 11, line 222 – do you mean "approved / cleared by FDA"? the use of "translation" and "approval" interchangeably here, and elsewhere in the manuscript, is confusing, and if this terminology is retained, it should be better defined upfront

We agree this is confusing and have now removed the term 'translation' from the manuscript.

16. Page 12, line 245 – the IDEAL model could be explained for readers unfamiliar with this concept

We have now expanded to further explain the IDEAL model.

> Discussion, Line 239: "…The Balliol Collaboration has proposed the IDEAL model for safe innovation to address this shortfall, the central tenet being that innovation and evaluation can and should proceed together in an ordered and logical manner."

17. Page 12, line 258 – I am not convinced that this paper could be read-across to the Contopoulos-Ioannidis paper, which appears primarily to include preclinical research (vs. the exclusively in-human studies included here). This would bias the current results to a higher % resulting in approved products relative to the C_I paper

We agree with the reviewer that the comparison with the Contopoulos-Ionnidis paper was not very helpful, and it has now been removed.

18. References – I would recommend strengthening the references list. The BMJ has

published quite a bit recently on devices and transparency, for example, and most of those studies are not cited. Other suggestions are listed above

We thank the reviewer and have now added several further references to the manuscript.

19. Table 2 – percentages / p-values for the tests reported in Results would be helpful here

We agree with the reviewer that presenting the percentages in Table 2 would be helpful, and did so in earlier drafts. However, we found that this was confusing to readers, who often misunderstood what the percentages represented (even when stated in the legend). We would be happy to defer to the Editorial Committee.