Summary of revision for manuscript (MBJ-2020-063318), "Analysing Privacy Issues of Android Mobile Health and Medical Applications"

Dear Editor,

We thank the reviewers and associate editor for their valuable comments that have helped us significantly improve our manuscript. In the following, we describe the changes we have made in our manuscript against points raised in the meta review and the individual comments from the reviewers. For readability and easy cross-referencing, we have included the original comments from the reviewers. The list c**ontains each point raised in the individual reviews by all reviewers and the associate editor. The comments have resulted in significant updates to the document. We have updated (removed/added some) figures**, **tables**. We have also performed additional analysis to further elaborate our findings and answer queries by the reviewers. For the edits corresponding to each of the points, we have mentioned and highlighted the section number and paragraph where the changes have been made. We comprehensively reviewed the manuscript for language issues (typos, syntax, repeated words, and phrases, etc). Changes have been incorporated (with blue color) in the manuscript.

If you have any queries, please do not hesitate to contact us.

Best Regards,

Gioacchino Tangari
Muhammad Ikram
Kiran Ijaz
Mohamed Ali Kaafar
Shlomo Berkovsky

Comments from committee

C.1. *The problems of data privacy in medical applications are of course important and your analysis gives us all something to think about and discuss. The paper is an unusual one for a clinical journal like the BMJ and many of the technical aspects will be beyond our readers. We will commission an accompanying editorial to go with your research and we also ask you to explain some of the technical terms where necessary. Perhaps you could ask a clinician who knows nothing of the subject to read and give feedback during the revision process.*

Response to C.1. By explaining the technical terms and expanding the discussion, we made significant efforts to improve the readability of our paper for non-tech savvy readers. We highlighted the edits in blue text in our paper.

C.2. *Please make a clear distinction between legal and illegal data breaches. At the moment it is very difficult to get a feel for how "wrong" some of the detected behaviours are.*

Response to C.2. As per your suggestion, we clarified what app behaviours are in breach of privacy regulations (such as GDPR) and thus can be considered illegal. In particular, we highlight in Section 2.2 (Analysis Methodology) that the disclosure of privacy practices of apps is a legal requirement set by privacy regulations (e.g. GDPR).

Moreover, we elaborate on the cases where a user data transmission is potentially illegal in Section 3.3, par. “Non compliance with privacy policies”, which reads as:

Specifically, we tag each data transmission as *complying* if the associated data collection practice was disclosed in the privacy policy, *violating* if the app has a privacy policy but the practice is not disclosed, *no privacy policy* if the app has no privacy policy.
Both the *violating* and *no privacy policy* cases are potentially illegal as they are clear breaches of privacy regulations like the GDPR (which requires informed and unambiguous consent -- https://gdpr.eu/gdpr-consent-requirements).

C.3.a *In the introduction please provide a little context for similarities and difference between Australia and, say, Europe and the USA. Are the same apps available?*

Response to C.3. We address this comment in the sixth paragraph of Introduction which reads as:

The scale of our analysis is orders of magnitude larger than previously reported analyses[11,14–16] and virtually covers all the Google Play store mHealth apps accessible from Australia, as a sound proxy for the world-wide Google Play app marketplace.
Google Play store [3] provides various filters and configurations to developers facilitating to localise and distribute Android apps releases to specific countries or geo locations [1] . We used this to validate that the majority of the collected (91%) and analysed (76%) mHealth

apps are not specific to Australia, but also present and available in other locations such as Europe and U.S.

C.3.b *Are the regulatory frameworks and privacy laws comparable?*

Response to C.3. We address this comment in second, third, and forth paragraphs of Introduction which reads as:

For example, EU General Data Protection Regulation[2] (GDPR) defines eight rights of individual users and several rules implemented under the U.S. Health Insurance Portability and Accountability Act[3] (HIPAA) establishes a baseline of privacy protection and patient rights.

In line with the HIPAA Act, the FDA (U.S. Food and Drug Administration) released in 2016 a Guidance for Postmarket Management of Cybersecurity in Medical Devices[4] . In this Guidance, FDA recommended medical device manufacturers, i.e., app developers, to incorporate risk management into the life cycle of their products and implement controls that ensure the devices are secure and protect patients. Specifically, it covers cybersecurity and privacy aspects and stipulates risk management programs that "*address vulnerabilities which may permit the unauthorized access, modification, misuse, or the unauthorized use of information that is stored, accessed, or transferred from a medical device to an external recipient, and may result in patient harm*".

However, the above Regulation and Guidance are hard to enforce in practice. Several recent episodes shed light on the problem of apps collecting and sharing data in an unauthorized manner. For instance, a Norwegian non-profit showed[5] that 10 popular apps, including a popular Health & Fitness one, leaked data to advertising companies without informed user consent, in a clear breach of GDPR. 41 popular apps, some developed by leading technology companies, have been called out[6] by the Chinese Ministry of Industry and Information Technology for illegal data collection. A 2019 decision by CNIL, the French data protection authority, deemed Google to be in breach of the principle of transparency[7] , because the information on the use of personal data was presented in a vague and difficult to comprehend manner.

C.4*. Please at least comment on likely similarities or differences with apps for the Apple OS.*

Response to C.4. In Appendix I (OS and Android Apps), we briefly discuss the difference between iOS and Android apps and their architecture.

Most popular smartphone operating systems worldwide are open-source Google's Android that comes with a wide variety of smartphones and Apple's proprietary iOS that is used in iPhones. Google provides free services such as analytics and ads served via Android built-in modules or application program interface. Android's apps can be developed by anyone and can be distributed either Google Play store or other marketplaces such as Tencent MyApp and Xiaomi Market[30].

In contrast, iOS apps can only be developed by subscribers to the iOS Developer Program, and can only be distributed through the official App Store.

App developers for either platform can monetise their apps by integrating third-party services for analytics, ads, social interaction or development aids into their apps. In particular, Google-owned services were the most recurrent in the analysed app set. This is likely due to the dominant position of Google's analytics and ad services, but it also reflects the choice of Google Play Store for our app dataset. Android apps leverage support tools (e.g. for bug reports) that directly report to Google, which may share additional device information. Therefore, we would expect a (slightly) less pronounced role of Google for mHealth apps in the Apple store

Recent versions of iOS (version 14 and above) and Android (version 11 and above) inform users about the capabilities of apps using run-time permissions to request phones' resources such as Camera and Contacts details. Although permissions may limit access to users' sensitive data, research studies [13, 22] have shown that the majority of apps' users do not quite interpret the permission dialogs and warnings.

C.5. *Please be aware of how results may be interpreted by people less familiar with the subtleties of the inner workings of apps. For example, potentially alarming charts or statements may deter patients from enrolling in COVID-19 contact tracing apps, which may be detrimental to their own and population health. It is also possible that conspiracy theorists may misuse some results to attract people to their cause.*

Response to C.5.

Thanks for your suggestion, we revised the text to avoid any alarming, potential mis-interpretations of our results.

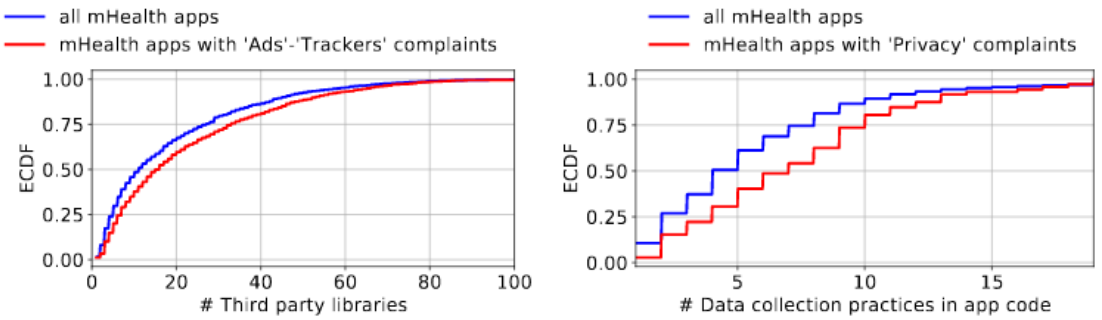
C.6. *Our statistician offered the following observations*

-The cumulative distributions in Figures 2 and 7 (Please define ECDF) would be better as density distributions. -- The barcharts would be better with the total bars omitted.

Response to C.6. As we presented the users' reviews ratings in Table 1, we removed Figure 2 (was captioned: Empirical Cumulative Distribution Function (ECDF) of the number of users reviews--overall and negative ones--per mHealth app) from this version of our paper. Snapshot of Table 1 with users review is as follow:

Fraction of Negative Reviews $(100\% \times \frac{\#\ of\ Negative\ Reviews}{\#\ of\ All\ Reviews})$	
0% – 20%	10,371 (49%)
20% – 40%	4,157 (20%)
40% – 60%	2,663 (13%)
60% – 80%	1,474 (7%)
80% – 100%	2,326 (11%)

We also replotted sub-figures in Figure 7 and expanded ECDF in the (sub)captions.



(a) Empirical Cumulative Distribution Function (ECDF) of the number of 3rd-party libraries in apps code.

(b) Empirical Cumulative Distribution Function (ECDF) of the number of data collection practices in apps code.

Figure 7. Relation between user complaints and the privacy conduct of mHealth apps, expressed in terms of third-party presence and data collection operations in each mHealth app.

***C.7**. Please present percentages as whole numbers. Decimal places add nothing useful.*

Response to C.7. Thanks for your suggestion. We agree as it improves the readability of our paper, we now present the percentages as whole numbers not only in tables (e.g., Table 1 and Table 10) but also in the text.

***C.8.** Our patient editor asked for the following:*
The authors need to add in their PPI declaration in their own words and if there was no PPI please have them share why so the statement is meaningful e.g. No funding, COVID restrictions, limited data access etc. The paper is also missing dissemination statements, this can be a good opportunity to involve one of the patient reviewers or someone in the community to co-author an Opinion piece about your research for The BMJ. The Link may be helpful https://drive.google.com/file/d/14vnXwTJ2CDn2KQsuNpuEnSwad69gc7dR/view This is a topic of great interest to the public and many have found the privacy limitations to be harmful or intrusive.

Response to C.8. In section 2.3, we provide PPI statement:

Patients and public were not directly involved in the study. The subject of the study were mHealth mobile apps publicly available on Google Play. The data collection and analysis methods leveraged an automated testing platform designed by the authors, not requiring the involvement of mHealth app users or developers. Likewise, we analyzed public app reviews from Google Play, which were voluntarily contributed by the app users. We share the collected datasets, the analyses library, and our findings with clinicians, patients, app developers, and the general public, aiming to raise awareness of privacy risks in mHealth.

***C.9.** Dissemination*
Please confirm when and how results were (or will be) disseminated. Guidance for best practice in dissemination is set out in the following link and gives examples: https://www.nihr.ac.uk/funding-and-support/documents/funding-for-research-studies /manage-my-study/How-to-disseminate-your-research/dissemination-guidance.pdf

Response to C.9. We will share and disseminate the collected dataset and results of our analysis at https://mhealthapps2020.github.io/. We updated our paper with the following statement in Section 'Addenda' which reads as:

Disseminating and sharing data and results: A sample of our dataset is available at https://mhealthapps2020.github.io/. Note that, upon publication, we will release all our dataset and analysis script for further research.

Comments from Reviewer 1

R1.1 *This paper may be rather esoteric for all but a small number of BMJ readers but that is not a reason to reject it.* ***A lot of the tables and graphs should be in an appendix to reduce the complexity for the non-technical reader trying to follow the arguments.*** *They need to be there but only for people who wish to look at the background information. Putting them online and not in the paper publication is obviously an option. It uses techniques that I have not seen before for analysis of a medical application. Mobile healthcare applications are becoming important and their potential impact is unknown. This paper uses ways of analysing the software and produces an interesting output that deserves to be more widely known. There has been significant disquiet about data collection from mobile apps and what it could be used for. This paper makes a contribution to defining the scope of the problem in health care applications. The objectives, design, interventions, and outcomes are reasonable and have been fulfilled. The results and conclusions, with minor changes, are acceptable and in my opinion an important addition to this area and may be considered groundbreaking. In conclusion I feel that this should be published with some corrections as outlined.*

Response to R1.1. Thanks for your suggestion. To improve the readability and reduce the complexity for non-technical readers, we have made the following changes":

- We have removed the Figure "Distribution of users reviews (overall and negatives) per mHealth app. "
- We have moved to Appendices the Table "Top-10 most popular mHealth apps (1M+ installs) with strong presence of advertisements"
- We have moved to Appendices the Table "Leaks of user data in HTTP and HTTPS traffic"
- We have moved to Appendices the Table: "Consistency of privacy policy (PP) with user data leaks in popular mHealth apps."
- We have moved to Appendices the Figure: "Relation between user complaints and the behaviour of mHealth apps"

R1.1. *Page No. 2 Line 14 (e.g., the.....*
Unnecessary comma should this be (For example, the....

Response to R1.1. We have removed unnecessary commas in all occurrences found in the paper.

R.1.2. *Page No. 4 Line 52 while paid and geoblocked apps were excluded. Can understand why but may this have introduced an unknown bias? Either the free apps captured more data to und them, or the paid for apps were more complex, had better security and were not selling data as they did not need to sell data to make it commercially viable. Difficult to unravel and would require significant funding. However, they analysed 75% of the 20,000 more than sufficient. Overall, I feel that this is not a major issue in the context of the issues revealed in this paper.*

Response to R1.3. To scale up the analysis to tens of thousands mHealth apps, we restricted the study to free apps only. In Section 4.2 of the revised manuscript, we acknowledge that this might introduce a bias, as paid apps have business models less dependent on selling user data, hence they may retrieve less user data and exhibit a reduced presence of ads and trackers. However, we conjecture this does not strongly penalise the generality of our findings, since no more than 15% of mHealth apps found on Google Play were not free.

R.1.4. *Page 12 Line 53 Insecure transmission of user data: I think most of this paragraph may be unnecessary as this is a well known problem. Could be reduced to a single sentence, "Analysing the communication leaking personal data, we observe that as much as 23% of leaks are in unencrypted HTTP traffic a known security problem."*

Response to R1.4. We have revised the paragraph "Insecure transmission of user data", which we reduced in size. The table with the detailed breakdown of user data transmissions using HTTP and HTTPS is now in the Appendices.

R.1.5 *Page 13 Line 10 et sequae. 3.4 User Perceptions of mHealth Apps This analysis is in my opinion interesting but not valid. User feedback is very variable and does not reflect the full picture. There is a bias in the data, negative feedback is much more common than positive. I do not think this can be included other than a simple statement that analysis of this area was tried but the results could not be validated. A variation on the final paragraph should suffice I would suggest "Analysis of user reviews was undertaken but because of the bias introduced by negative feedback being far more common than positive feedback, we conclude that while mHealth app users have a limited interest in (or awareness of) the apps' privacy conduct and the presence of ads/trackers and the inclusion of user data collection operations the significance is not clear."*

Response to R1.5. In accordance with this comment, and based on the observations from Reviewer 5 (comment 5.11), we revised section 3.4 (Review Analysis). We reduced this section in size, and we revised our statements about the users perceptions on mHealth apps and their privacy, given the uncertain validity of the analysis. Moreover, we explicitly acknowledge that what is measured in the negative app reviews is difficult to untangle, as app reviews may not be the only destination for user concerns on privacy.

Comments from Reviewer 2

R2.1 *The study presented the data collection practices and analysed the current state of mHealth apps on google play. As indicated in the paper, Google requires the app developers*

to disclose the collection and sharing of user data. It will be informative to present the correlation between third party presence in app resources, access to personal data in the app code and privacy policy.

Response to R2.1 The results reported in Section 3.2 show that those third parties whose tools/libraries are most frequently included in the app resources (e.g., Google Analytics, Facebook), are also prevalent in the data collection operations identified in the app code. We have further stressed this relation in Section 3.2 (under "Third Party Data Recipients"). Concerning the app privacy policies, it is difficult to measure the disclosure of data sharing practices towards specific third-parties, since the privacy practices are often presented using broad terms (for instance, "the app shares personal information with selected third-party advertisers, partners, etc.").

R2.2. *As indicated by the authors, the analysis is based on automated testing platform, not involving users or developers. Thus, providing user perceptions of mHealth Apps based on negative reviews only would not represent the user perception.*

Response to R2.2. We agree with the Reviewer that the analysis of negative user reviews is not enough to unveil the users' perceptions on the privacy conduct of mHealth apps. This point has also been raised by Reviewer 1 and 5. To address it, we have modified Section 3.4 (User Complaints in App Reviews) and revised our considerations/conclusions from the inspection of user reviews.

R2.3 *Having said that, the study can be indicated as the results of automated testing and evaluation but to draw the conclusion of user awareness. Privacy and confidentiality concerns are for any online forum*
or applications and comparison of mHealth apps and other apps would be informative.It is noted that the apps have been categorised as medical and health and fitness in the manuscript. Could you please clarify the definitions use to categorise medical apps in the manuscript. Are these the medical devices that are identified by FDA or TGAs? Privacy and data collection of these would be of interest to clinicians and patients and that would be an important contribution to the literature and also reflecting of the privacy requirements from NHMRC guidelines or the Privacy guideline in Australia or the specific health privacy requirements from the jurisdiction.

Response to R2.3. The categorisation of mHealth apps (in Medical and Health & Fitness classes) comes from Google Play, which is currently the most prominent mobile apps store. Only a small portion of Medical apps are actually approved by FDA/TGA (e.g. included in the FDA "510(k) Premarket Notification" database).

Differently from existing research work in the area (which has generally targeted specific, small-sized, segments of health-related applications), our study aims to provide a large-scale analysis of mHealth privacy aspects. To keep the scale of the analysis, we just followed Google's categorisation. This categorisation (Medical and Health & Fitness) is also consistent with the other main app platform, the Apple App Store.

R2.3. *There would be some health and fitness app that access personal information and monitor their health progress with the consumer's consent and share information to the third party with their consent. Thus, the purpose of the apps and the privacy policy needs to be considered. The study concluded that the mHealth apps are far from transparent when dealing with user data. However, it has not been presented in the article the correlation between them. Several fitness apps and several other apps are under the category of health and wellness application in the google play store but they are not the decision support tools for clinicians as indicated in the article section 5, line40. The paper could be expanded to discuss comprehensively on privacy issues or presented as the comparison of data collections providing the open question.*

Response to R2.3. In Section 3, our paper presents comparative results for mobile applications of different categories (Medical and Health & Fitness), as well as baseline (i.e., non health related) apps. The results (Table 2, Fig 3,5,6) show that Medical apps (more likely to be “candidate” for approval/accreditation as medical devices) are generally less prone to retrieve/share user data compared to Health and Fitness ones, and that overall mHealth apps are more conservative in data collection practices compared to non-mHealth ones. We have made these considerations more explicit in the Discussion section.

Comments from Reviewer 3

R3.1 *1.Fix grammatical errors in the text. Last paragraph in section 1 Introduction, Insert a “to” after order.*

Response to R3.1. We have further revised the manuscript to correct grammatical errors.

R3.2 *Table 1. Consider add “Yes” and “ No” above columns in rows “Contains Ads...” and “ Includes privacy link .....*

Response to R3.2. We revised Table 1 with “Yes” and “No” above the columns in rows “Contains Ads …” and “Includes privacy link …”.

R3.3 *Table 1. There is a duplicate “privacy’ word in the row “includes privacy link...”*
has been adjusted according to the Reviewer’s comments in the revised version of the manuscript.
Response to R3.3. Thanks for pointing to it. We have removed the duplicate and revised Table 1 accordingly.

R3.4*. Define ECDF in Figure. 2*

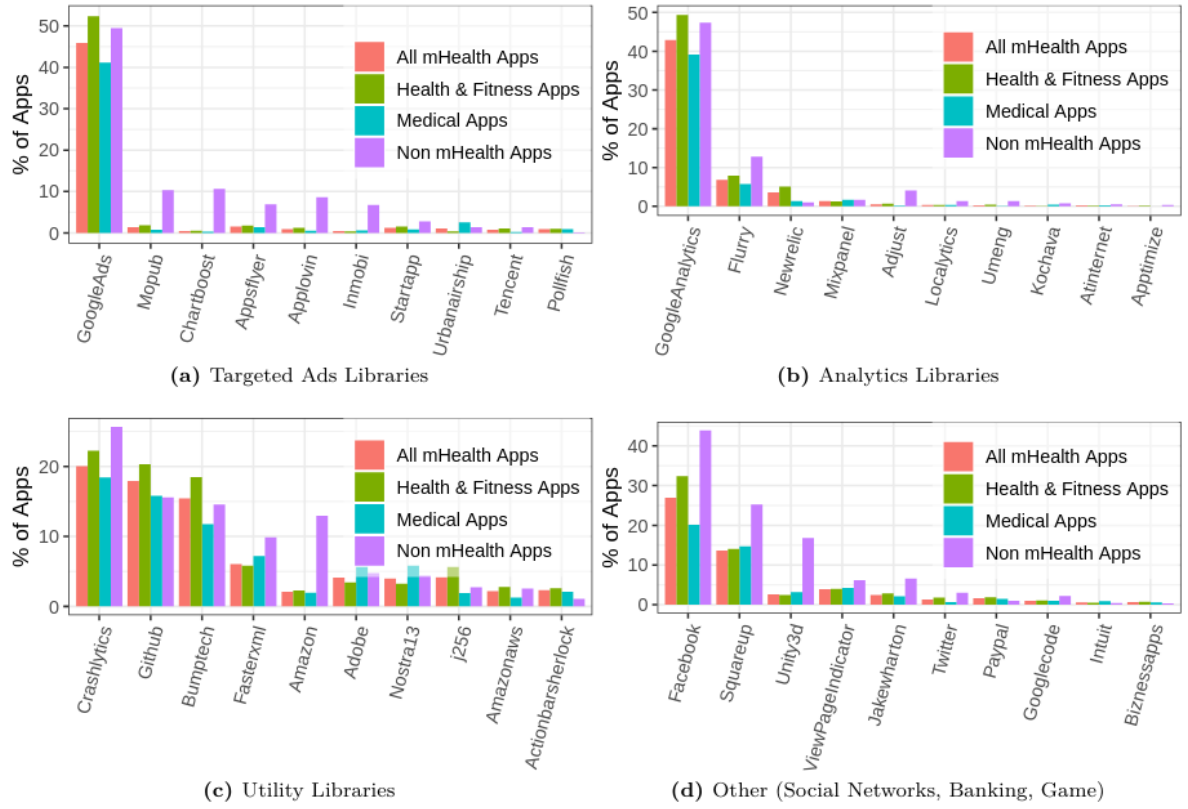
Response to R3.4. We now define the ECDF (Empirical Cumulative Distribution Function) in the paper. For conciseness, we have removed Figure 2, as the overall results on % negative reviews are already included in the dataset summary table.

R3.5 *5. In Figure 4 define All. Is this all mHealth apps?*

Response to R3.5. In Figure 4 caption, we have clarified that “All” stands for “all practices in the app code” in Fig 4.A, all “all data transmissions” in Fig 4.B.

R3.6 *In Figure 5 consider the first 10 sets of libraries as for others the results are negligible. If its important to list all, the results for the remainder can be added in an appendix.*

Response to R3.6. As per your suggestion, we improved the readability of Figure 5 by increasing the font size of legends and x/y-labels as well as limited the number of libraries to 10. The replotted Figure 5 is:



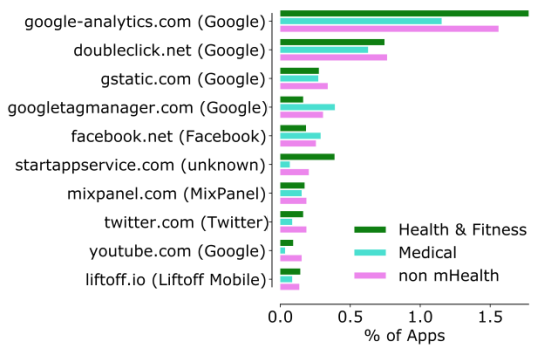
(a) Targeted Ads Libraries (b) Analytics Libraries

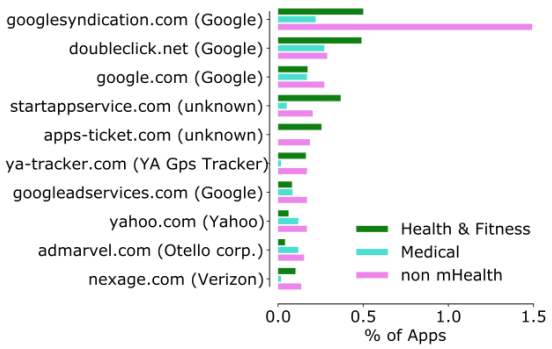
(c) Utility Libraries (d) Other (Social Networks, Banking, Game)

Figure 5. Third-party libraries found in various mHealth app categories and non-mHealth apps.

R3.7. *In Figure 6 it is difficult to differentiate the difference between various add and tracker domains. Suggest only listing the major ones.*

Response to R3.7. We have replotted Figure 6. In the new version, we only show the top-10 ad and tracking domains. In addition, we have specified the ‘entity’ to which each domain belongs.





R3.8. *In Table 5 it seems that no privacy policy column defines where privacy policy exists but does not cover certain data leaks. Reword this column. However if there is no privacy policy, how can we say there is compliance or violation regarding data leaks?*

Response to R3.8. We now provide a more informative description for Table 5 to clarify the different cases: no privacy policy, complying, violating.

R3.9 *Table 6. What is the basis of the PP violation %.*
Response to R3.9. We have modified the labels in Table “Consistency of privacy policy (PP) with user data sharing in popular mHealth apps” for clarity. The label “PP violation [%]” has been renamed as “Transmissions violating the PP [%]”, which indicates the fraction of user data transmissions that are not disclosed in the respective app privacy policies.

R3.10. *It is great that the authors have highlighted a problem, however, it would be useful to incorporate a comprehensive section of what steps patients and clinicians can carry out to ensure that they use apps that are compliant with national data privacy guidelines such as GDPR. In addition what action can Google take to reduce lack of privacy policy implementations and to increase compliance with stored privacy policies and to national Guidelines?*

Response to R310. In the revised version of the manuscript, we have incorporated a subsection “Recommendations” (4.4), where we provide recommendations and suggestions for patients and practitioners on how to better understand and articulate the privacy aspects of mobile health apps. In particular, we point to the existing resources available to doctors to familiarise with the apps privacy aspects, and we suggest the adoption of tools that can help understanding long and complex privacy policy documents.

We also discuss the need for app stores (Google Play, Apple Store) to better review the app privacy disclosures upon publication, as we found that a significant portion of mHealth apps offer no privacy disclosure at all.

Comments from Reviewer 4

I enjoyed reading this important paper that analyzed the privacy issues of Android mobile

health applications. This is study is expansive in scope, and employed rigorous methods to elucidate interesting insights.
Here are my thoughts and suggestions for consideration:

R4.1 *Consider articulating how the android apps are positioned within the larger 'market' that consists of apps from other platforms (e.g., Apple). In particular, what is the market capture of android health/wellness apps as compared to Apple? Do Android health/wellness app make up just 10% or perhaps 60%? Presenting the 'market' context will help articulate the representativeness of the data.*

Response to R4.1 Thanks for your comment; as per your suggestion, we have included the following text in Section 2:

Since 2015, app marketplaces such as Google Play and Apple Store experienced approximately 38% growth and it is expected to generate 111.1 billion by 2025[2]. While the number of mHealth apps available in the appstores is constantly increasing[21], it has been recently estimated that, out of the 2.8 million apps on GooglePlay and the 1.96 million apps on Apple Store, 99,366 belong to medical and health & fitness categories.Specifically, these apps account for the 2% (47,890) of Google Play and the 3% (51,476) of Apple Store[4,6].Our analysis focuses on Google Play, the largest app store, and it virtually covers all the Google Play mHealthapps accessible from Australia, as a sound proxy for the world-wide Google Play app marketplace.

R4.2 *Consider stating explicitly the paper's conceptual significance. From my reading, the key contribution is providing a framework to analyze privacy of mobile apps. Consider expounding on this.*

Response to R4.2. By leveraging hybrid--static and dynamic--analysis, published in parts in previous papers, our key contribution is the (large-scale) analysis of mHealth apps. To foster future research, upon publication, we provide our data collection and analysis scripts as well as the collected dataset to the research community.

R4.3 *The authors made a good point that theirs is one of the first studies to analyze privacy metrices in Android mobile health apps. In the introduction, consider stating how serious is the privacy issue? This adds to the conceptual significance of the paper.*

Response to R4.3. We believe it is hard to quantify how serious the privacy issue is in the mobile health ecosystem. This has been a key motivation for our study.

Recently, several episodes have shed light on the problem of mobile apps collecting and sharing data illegally, for instance:

- A recent report commissioned by the Norwegian Consumer Council (https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf) found that 10 popular apps (including a popular Health & Fitness one) leak data to advertising companies without informed and unambiguous consent, which constitutes a breach of the European GDPR regulation

- 41 popular apps from Chinese companies have been recently called out for illegal data collection https://technode.com/2019/12/19/tencent-xiaomi-apps-called-out-for-illegal-data-collection/
- A 2019 decision by the French data protection authority CNIL deemed Google to be in breach of this principle of transparency, because information about processing operations was spread across multiple documents, and because the information on how personal data may be used was presented in a way that was vague and difficult for the user to comprehend.

These episodes, which show spread of user data as well as lack of transparency in disclosure, motivated our large-scale analysis of the privacy practices of mHealth apps and their associated risks.

We have modified Section 1 (Introduction) in order to better highlight the privacy issue.

R4.4 *The practical significance could also be discussed. For example, what do the findings mean for policy? What are the implications for data security and protection? How could Google use these findings to improve their privacy policies? How do clinicians better articulate the benefits/risk of apps to patients based on the findings of this important study?*

Response to R4.4. Our analysis, conducted on a large number of mHealth apps, demonstrated that the collection of personal user information is a pervasive practice, albeit not always transparent and secure. While we cannot directly affect policies, we believe that, by bringing up our findings (and sharing our analysis data), we can contribute to making doctors and policy makers more aware of the privacy aspects of these apps and their inherent risks.

In the revised manuscript, we have included a Recommendation section (Section 4.4), which provides recommendations for patients and practitioners on how to better understand and articulate the privacy aspects of mobile health apps. In Section 4.4., we also encourage app stores (Google Play, Apple Store) to take a more thorough review of the app privacy disclosure, as we found a significant fraction of mHealth apps providing no privacy policy at all.

R4.5 *The paper excels in detail and rigor, however I found that the key topic of privacy is somewhat distracted by the many sub-topics and figures. Consider a sharper focus in the main text and its figures, and relegating the other impressive, albeit tangential, figures to the supplementary document.*

Response to R4.5. In the revised version of the manuscript, we have made the following changes to keep a sharper focus on the main topics:

- We have removed the Figure “Distribution of users reviews (overall and negatives) per mHealth app. ”
- We have moved to Appendices the Table “Top-10 most popular mHealth apps (1M+ installs) with strong presence of advertisements”

- We have moved to Appendices the Table “Leaks of user data in HTTP and HTTPS traffic”
- We have moved to Appendices the Table: “Consistency of privacy policy (PP) with user data leaks in popular mHealth apps.”
- We have moved to Appendices the Figure: “Relation between user complaints and the behaviour of mHealth apps”
- We have moved the details of “Insecure Transmission of User Data” (Section 3.3) to Appendices
- We have moved part of Section 3.4 (review analysis) to Appendices
- We have reduced the technical details in the Analysis Methodology (Section 2.2)

While de-prioritising these details, we have elaborated more on the paper contributions and significance (Section 1) and on the discussion (Section 4), where we also included a Recommendations subsection (addressing comment R.3.10)

R4.6 *A small typo in the Abstract (“objective”): “To investigate whether and what user data is...” It should be “data are” as the singular form is datum.*
We have further revised the manuscript to correct grammatical errors.

Comments from Reviewer 5

Thank you for the opportunity to review this manuscript. The authors conducted a large-scale analysis of the privacy and data sharing practices of a sample of more than 20,000 medical, health and fitness apps available in the Australian Google Play store for the Android platform. While other analyses of data sharing practices have been conducted at scale, this analysis focuses on health apps.

The authors confirm previous analyses of health apps in finding that the majority of sampled apps collect and share user information with third-parties that provide analytics and advertising services. They further analyse data sharing practices against stated privacy policies, confirming a large literature that suggest that privacy policies are lacking detail and transparency. The most novel aspect of the study was to analyse public app reviews through a privacy lens using machine learning methods, noting that user complaints and awareness of privacy risks are rare.

Please note that I am not qualified to critique the machine learning or other code/traffic analysis methods in detail, but did note that further information on the sensitivity/specificity of the approach may be useful in a supplementary file.
Overall, the authors have conducted a large-scale and comprehensive analysis of mobile health apps in the Australian Google Play store, triangulating data from multiple unique analyses. The paper could be strengthened by highlighting the key gaps in the literature – much of what the authors do replicates and confirms smaller-scale analyses – and emphasising what the ‘scale’ adds. The paper is long for a clinician audience in its current form and so really emphasising what these analyses add, what is novel and ensuring that all analyses are well-integrated and justified would make this more impactful.

R5.1 *I would suggest that the introduction focus squarely on the state of mHealth in relation to privacy. For example, while the authors discuss issues related to efficacy and clinical safety, greater focus could be had from explicating what is known about mHealth data privacy. Further, developments like the FDA guidance should be described in terms of how they address privacy. The summary of key findings may be better placed in the first paragraph of the Discussion; the Introduction should instead explicitly state the gaps that this paper will fill given that there have been numerous recent analyses of data sharing practices of mHealth apps. Specifically, what does a large-scale analysis add to our understanding of risks and benefits of mHealth? Much of what is currently located in the 'Comparison to other studies section' could be beneficial in the introduction to highlight what this study adds.*

Response to R.5.1. Thanks for your comments. We revised Section 1 (Introduction) to clarify the contributions of our paper. For details, please refer to **Response to C.3.**

R.5.2*The authors present information in Figure 3 contrasting the practices between health apps and non health apps. This strikes me as particularly novel and a key contribution to the literature, but it is given little attention in the introduction/study aims and could be made more of in the results/discussion. This also strikes me as the key contribution of a large-scale analysis.*

Response to R.5.2. We provide a note in the paper stressing this key contribution. The text in Introduction now reads as follow:

"as a sound proxy for the world-wide Google Play app marketplace. Not only the scale of our study orders of magnitude broader than previous analyses that considered tens of apps, but we also refine the granularity and depth of our analysis. For example, Dehling et al. categorised mHealth apps into the low, medium, and high privacy risk groups[12], disregarding the type of user information being leaked, the recipients of the information, whether this was disclosed in the app's privacy policy. Moreover, we consider the security of the communication protocols used by the apps, the presence of ad and tracking libraries in the app code, and the users' reviews on the app's privacy conduct. Hence, our analysis offers a more comprehensive view of the privacy aspects of mHealth apps than previous works, also relating the findings to the broader range of baseline non-mHealth apps.."

R.5.3. *A strength of this analysis is that the authors employ three distinct methods of analysing privacy/data sharing practices: static file/code analysis, dynamic traffic analysis, and privacy policy analysis. However, greater clarity around their respective strengths/limitations and integration of these findings is required.*

Response R.5.3. We now clarify in Section 2.2 (Analysis Methodology) the main limitations of app code and traffic analysis, especially when used to characterize the user data collection by the apps.

In particular, in paragraph "Data collection operations in app code", we stress that the result of app code analysis is a "superset" of the actual user-data collection, as some parts of the code may never be used in practice (e.g. unused external libraries):

“[...]The final set of functions represent all the potential data collection in the app: in practice,it is a superset of the actual user-data collection, as some parts of the app code may rarely (or never)be triggered during app execution”.

At the same time, in paragraph “Personal data transmission in app traffic”, we stress that the set of user data transmissions extracted from the traffic may not be complete, because of coverage limitations of the automated app testing:

“[...] The result only includes data collection practices that are “actually” performed during app execution; however, this set is not complete due to coverage limitations of dynamic app testing – which may not trigger some menus, views, or functionalities of the app”

R.5.4. *First, the authors variably use the terms “data collection practice,” “leaks” and “operations involving personal data.” This needs to be clarified right up front and consistent terminology used. It is particularly unclear what constitutes a ‘leak’ (ie does this apply to data sharing intended by the developer?) This is particularly important for the analysis of privacy policies as it is unclear exactly what the authors measured in terms of comparing “data leaks” to privacy policies. Thus, I was unsure what exactly constituted a “violation.” Further, I did not understand how the proportion of violations was calculated – what constitutes a ‘single’ privacy leak for example? What is the denominator?*

Response to R5.4.
We have modified our terminology on “data collection” for clarity:

- We do not generically refer to static analysis results and dynamic analysis results as “data collection”. We have left the title of Section 3.1 as is (“Personal Data Collection Practices”), but we made sure that in the text the data collection-related functions in the apps code are clearly distinguished from the actual transmissions in the apps traffic
- We have removed the terms leak(s) and leakage. We now specifically refer to traffic analysis results as to “user data transmissions in the app traffic”.

We have modified the description of Table “Consistency of data collection disclosure in the privacy policy with the user data transmissions in apps traffic” for clarity. The label “PP violation [%]” has been renamed as “Transmissions violating the PP [%]”, which indicates the fraction of user data transmissions that are in violation of the privacy policy.

We have clarified the categorisation the user data transmissions:

“Specifically, we tag each data transmission as *complying* if the associated data collection practice was disclosed in the privacy policy, *violating* if the app hasa privacy policy but the practice is not disclosed, *no privacy policy* if the app has no privacy policy. Both the *violating* and *no privacy policy* cases are potentially illegal as they are clear breaches of privacy regulations like the GDPR (which requires informed and unambiguous consent).”

R5.5. *The authors do not distinguish between actual and potential data sharing, which I think may be a more accurate representation of what is measured. Previous analyses of apps found that static code analysis can detect possible or potential data sharing, but that often*

embedded ad libraries, for example, often go unused. The dynamic traffic analysis is a point in time analysis of actual data sharing. The authors should consider how these measurements can be compared and contrasted to provide a more nuanced picture, but I

think it is inappropriate to simply combine these findings as "data collection practices."

Response to R5.5. Thanks for your comments. We agree with your recommendation, and now we better distinguish between the results from (static) app code analysis and (dynamic) traffic analysis.. Overall, In the new text of Section 3.1 and 3.2, we clearly decouple the results of *potential* data collection in the apps code (Figure 2.a, 3.a; Table 2 (top), 3 (top)) from the ones on *actual* data transmission in the apps traffic analysis (Figure 2.b, 3.b; Table 2 (bottom rows), 3 (bottom)).

Section 3.1 now reads as follows:

The analysis of apps files/codes yielded a total of 65,068 data collection operations, on average 4 per app. *This result provides the broad set of all information apps can potentially access and share third-parties.* At the same time, in apps traffic analysis we identified 3,148 transmissions of user data across 616 different apps
[...]
Focusing on the app traffic, we observe user data transmissions in 4% of mHealth apps, and mostly for Health & Fitness ones. *This percentage is significant, and should be taken as a lower-bound for the real data transmissions performed by the apps, as some transmissions may not be triggered in automated app testing*

In addition, Section 3.2 now reads as follows:

We present the main third-parties involved i*n data collections operations in the app code, and the ones most frequently receiving user data in the app traffic.* [...] Considering the data collection operations in the app code, a substantial fraction is associated with Google services, besides which we note a significant presence of Facebook (14% of apps embed Facebook cookies), Flurry analytics (6.3% of apps) and PayPal payment service. *Unsurprisingly, the services that are most frequently included in the app resources* (e.g. Google and Facebook libraries), *are also prevalent in the data collection operations identified in the app code. As to the user data transmissions in app traffi*c, we observe that Contact data is mainly shared with analytics services (e.g. Google's crashalytics.com), while the Location and DeviceID transmissions are mainly towards ads (e.g. Liftoff app marketing) and smartphone notification services (e.g. Pushwoosh).

R5.6. *Similarly, in characterizing sharing with third-parties, the findings of third-party sharing in the static file code should be separated from those in the dynamic traffic analysis in terms of possible vs actual (or as the authors state, integrated vs interacting). For example, the proportion sharing with third parties is much lower in the dynamic traffic analysis.*

Response to R.5.6. See our response to **R.5.5.**

R5.7.*It is not surprising that among Android apps, the vast majority have Google services*

embedded in their code – could this be an artefact of Google's developer services? Further, without analysis of iOS apps, the role of Google within the greater mobile ecosystem should not be overstated. I wonder if the analysis would be more meaningful with Google's services removed? Or to do a sensitivity type analysis of third party entities without Google's services?

Response R5.7. We agree that there is a over-presence of google.com as a user data recipient (Table 3) and this is due to the fact that (1) apps incorporate Google's analytics and ad services, and (2) Android apps leverage Android support tools (e.g., for bug reports) many of which report to Google and may be sharing user information (e.g., device ID). We acknowledge that this could partly be an effect of choosing Android apps in Section 4.1, which reads as follows:

[...]Notably, 50 prominent services are responsible for roughly 70% of the data collection operations in apps code and data transmissions in apps traffic. In particular, Google-owned services were the most recurrent in the analysed app set. This is likely due to the dominant position of Google's analytics and ad services, but it also reflects the choice of Google Play Store for our app dataset. Android apps leverage support tools (e.g. for bug reports) that directly report to Google, which may share additional device information. Hence, we would expect a (slightly) less pronounced role of Google for mHealth apps in the Apple store.

R5.8. *On page 10, line 45-48, the authors analyse the types of user data against the categories of third-party entities; this seems especially interesting and important for understanding the nature and level of privacy risk. This could perhaps be highlighted or further analysed. For for example, in Table 3, could the types of data be analysed against categories of third-party?*

Response to R5.8. Thanks for your suggestion. We updated Table 3, which now also shows the category of third-party (ads, analytics, social media, development aid, payments, utility).

R5.9 *Most of these company names will not be recognizable by the average clinician reader. In describing the integration of third-party libraries, I think the authors are referring to ad libraries, but this should be described very explicitly for a generalist, clinician audience (ie what they are and how they work). The findings in Table 2 underscore why the static code analysis and traffic analysis results should be treated separately and triangulated rather than combined – there are discrepancies between what trackers exist in the code and which are deployed.*

Response to R5.9. See Response to R5.5 and R5.8.

R5.10 *The description and rationale for the analysis of HTTPS transmission of user data was missing from the introduction and methods. To better integrate with the other analyses, the authors should first explain the significance of HTTPS in relation to the privacy practices Analyses.*

Response to R5.10. We have included a paragraph "Secure transmission of user data" in Section 2.2 (Analysis Methodology), which reads as follows:

Secure transmission of user data: We measure the fractions of user data transmissions using the HTTP and the HTTPS protocol. While HTTP-based communications are unencrypted, HTTPS encrypts all messages to protect app users from malicious data interception and content tampering. In the light of recent reports of widespread Internet surveillance[7] and legislation permitting internet service providers to sell user information extracted from network traffic[8], HTTPS adoption is essential for user privacy protection[26].

R5.11 *In analysing the privacy-related user complaints, more detail on what was considered 'privacy-related' would be useful for the reader. For example, the presence of ads may be "annoying" to app users (without explicitly identifying a privacy problem) and yet, still pose a privacy problem. Would these be considered privacy-related complaints? Further, in Australia, where there is a Privacy Commissioner and privacy principles, apps users may have other resources for privacy complaints and compliant privacy policies should identify the contact information for the person responsible for handling complaints. Was this detected in the privacy policy analysis? App reviews may thus not be representative in terms of the destination for privacy-related concerns, so I would suggest taking care with interpretations/conclusions that app users are "uninterested" or "unaware", but would instead stress the correlation between privacy practices and review content.*

Response to R5.11. We modified Section 3.4 according to the Reviewer's comments and following the recommendations of Reviewer 1. We reduced this section in size, re-organized our results broken down by complaint category, and we revised our statements about the users perceptions on mHealth apps and their privacy, given the uncertain validity of the analysis. In particular, we now explicitly acknowledge that what is measured in the negative app reviews is difficult to untangle, as app reviews may not be the only destination for user concerns on privacy. In the last paragraph of Section 3.4,, we stress on the positive correlation between complaints measured in the app reviews and the actual app behaviour.

R5.12*The abstract could be edited for clarity (e.g. "detected data-collection practices are towards the app developers. . .")*

Response to R5.12. Thanks for your comment. We have revised the paper abstract for clarity.

R5.13*In the abstract, please quantify "a small number of third-parties" received 67.8% of the collected data.*

Response to R5.13. We have clarified that 67.8% refers to top-50 third parties.

5.14*The abstract results refers to data collection and also data leaks – could you define and differentiate these two instances?*

Response to R5.14. In the new abstract, we clearly differentiate between "data collection operations in the app code" and "data transmissions in the app traffic". The Interventions, Main Outcome Measures, and Results sections of the abstract now read as follows:

INTERVENTIONS: Laboratory-based, cross-sectional assessment of each app, including the inspection of the app resources (code/files), and interception and analysis of app-generated

network traffic. Data collection related operations extracted from the app code through automated search, and user data transmissions identified in the app traffic through automated classification of the traffic flow. Analysis of the app privacy policy and user reviews.

MAIN OUTCOME MEASURES: Characterisation of the data collection operations in the apps code and the data transmissions in the apps traffic. Analysis of the primary recipients for each type of userdata. Presence of ads and trackers in the app traffic. Audit of the app privacy policy and compliance of the privacy conduct with the policy. Analysis of complaints in negative app reviews.

RESULTS: 88% of mHealth apps include code that can potentially collect user data. 4% of apps were found transmitting user information in their traffic. The majority of data collection operations in apps code and data transmissions in apps traffic involve external service providers (third parties). The top-50third parties are responsible for most of data collection operations in apps code and data transmissions in apps traffic (68%, collectively). 23% of user data transmissions occur on insecure communication protocols. 20% of apps provide no privacy policies, while only 47% of user data transmissions complywith the privacy policy. Less than 2% of user reviews raise privacy concerns.

R5.15*The paper could be edited to reduce the word count quite significantly. The authors helpfully provide several 'road map' type statements throughout to direct the reader, but this might be more succinctly replaced with strategic headings and subheadings.*

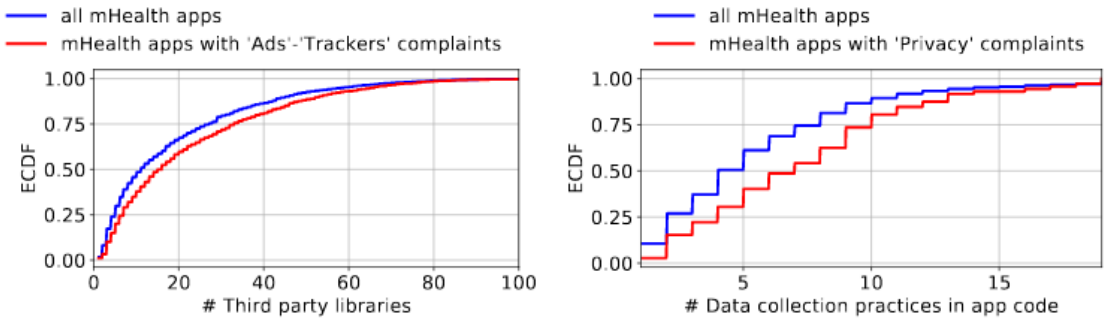
Response to R5.14. Thanks for your comment. We removed the road-map text from Section 3.

R5.16*The scale of Figure 5 made it very difficult to read.*

Response to R5.16. We have refactored Figure 5 for readability. For details, please refer **Response to R3.6.**

R5.17*Figure 7 – what is ECDF? This should be spelled out in the legend. I did not understand*
what units were measured on each of the axes in Figure 7.

Response to R5.17. We now define the ECDF (Empirical Cumulative Distribution Function) in the paper. We have modified the captions of Figure 7 to make them more informative. Please refer to **Response to C.6** for details.



(a) Empirical Cumulative Distribution Function (ECDF) of the number of 3rd-party libraries in apps code.

(b) Empirical Cumulative Distribution Function (ECDF) of the number of data collection practices in apps code.

Figure 7. Relation between user complaints and the privacy conduct of mHealth apps, expressed in terms of third-party presence and data collection operations in each mHealth app.

5.18*In the Discussion, you might need to define IMEI and MAC address for a generalist audience – what are these and what is their significance? (or perhaps simply refer to "persistent identifiers" and explain elsewhere in the paper why this is a category of data of concern).*

Response to R5.18. To make it clearer, we further explain the persistent device identifiers and their significance in Section 3.1, which reads as:

The main types of data mHealth apps can collect include contact information, user location, and several device identifiers. Part of these identifiers (specifically, International Mobile Equipment Identity(IMEI), Media Access Control (MAC), and International Mobile Subscriber Identity (IMSI)) are unique and persistent and can be used by third-parties to track users across networks and applications

The specific persistent identifiers are detailed in footnotes and in Appendix A.

R5.19*It may be of interest to readers to have some information about the sensitivity/specificity of machine learning methods available in supplementary files. You mention 96% accuracy in the Discussion, but this is the first mention of this in the paper.*

Response to R5.19. We have included more specific information about the accuracy of machine-learning methods in Section 2.2 (Analysis Methodology), including the results on Precision and Recall.

R5.20 *The conclusion in part, emphasises the security of data transmission, which seemed a minor part of the analysis. This should either be included as a key aim and justified as a key analysis, or minimised in the conclusion.*

Response to R5.20. As it is not primary findings of our paper, we revised the Conclusion which reads as:

This work investigated the privacy conduct of mHealth apps, belonging to the Medical and Health & Fitness categories on the Google Play store. To this end, we developed an infrastructure to analyze more than 20,000 apps and found that the majority of apps can collect and potentially share data with third-parties, including advertising and tracking services. Interestingly, the apps collected user data on behalf of hundreds of third-parties, with a small number of service providers accounting for most of the collected data. The analysis also revealed that mHealth apps were far from transparent when dealing with user data, with only about half of the apps found to be compliant with their declared privacy policies (if available at all).

Mobile apps are fast becoming sources of information and decision-support tools for clinicians and patients alike. Given that our analyses uncovered worrisome privacy issues and limited user awareness, we argue that it is important to surface our findings on the potential privacy risks and bring them to the attention of clinicians. They should be cognisant of these risks and consider them carefully, to ascertain that the benefits of an app outweigh its risks. On top of this, it is important to articulate such privacy risks to patients and potentially make this an inherent part of the app usage consent. We believe it is critical to consider the trade-off between the benefits and risks of mHealth apps for any technical and policy discussion surrounding the services provided by such apps.