Mobile health and privacy: cross sectional study

Gioacchino Tangari,[1] Muhammad Ikram,[1] Kiran Ijaz,[2] Mohamed Ali Kaafar,[1] Shlomo Berkovsky[2]

[1]Department of Computing, Macquarie University, Sydney, NSW, Australia

[2]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW, Australia

Correspondence to: M Ikram muhammad.ikram@mq.edu.au (or @midkhan on Twitter: ORCID 0000-0003-2113-3390)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2021;373:n1248
http://dx.doi.org/10.1136/bmj.n1248

Accepted: 16 May 2021

ABSTRACT

OBJECTIVES

To investigate whether and what user data are collected by health related mobile applications (mHealth apps), to characterise the privacy conduct of all the available mHealth apps on Google Play, and to gauge the associated risks to privacy.

DESIGN

Cross sectional study

SETTING

Health related apps developed for the Android mobile platform, available in the Google Play store in Australia and belonging to the medical and health and fitness categories.

PARTICIPANTS

Users of 20 991 mHealth apps (8074 medical and 12 917 health and fitness found in the Google Play store: in-depth analysis was done on 15 838 apps that did not require a download or subscription fee compared with 8468 baseline non-mHealth apps.

MAIN OUTCOME MEASURES

Primary outcomes were characterisation of the data collection operations in the apps code and of the data transmissions in the apps traffic; analysis of the primary recipients for each type of user data; presence of adverts and trackers in the app traffic; audit of the app privacy policy and compliance of the privacy conduct with the policy; and analysis of complaints in negative app reviews.

RESULTS

88.0% (n=18 472) of mHealth apps included code that could potentially collect user data. 3.9% (n=616) of apps transmitted user information in their traffic. Most data collection operations in apps code and data transmissions in apps traffic involved external service providers (third parties). The top 50 third parties were responsible for most of the data collection operations in app code and data transmissions in app traffic (68.0% (2140), collectively). 23.0% (724) of user data transmissions occurred on insecure communication protocols. 28.1% (5903) of apps provided no privacy policies, whereas 47.0% (1479) of user data transmissions complied with the privacy policy. 1.3% (3609) of user reviews raised concerns about privacy.

CONCLUSIONS

This analysis found serious problems with privacy and inconsistent privacy practices in mHealth apps. Clinicians should be aware of these and articulate them to patients when determining the benefits and risks of mHealth apps.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Mobile applications (apps) often collect user data and share it with developers' controlled servers as well as external third party, commercial entities

Mobile health (mHealth) apps pose concerns about privacy owing to the sensitive user information they can access

Inadequate privacy disclosures have been repeatedly identified for top mHealth apps, preventing users from making informed choices around the data

WHAT THIS STUDY ADDS

88% of the 20 991 mHealth apps included in this study could access and potentially share personal data

mHealth apps collected less user data than other types of mobile apps

Data collection in mHealth apps was found to be far from transparent and secure, and often exceeded what is publicly disclosed by app developers

Introduction

With the improved accessibility of smartphone devices, mobile applications (or apps) available through a variety of marketplaces have grown exponentially. As of 2021, almost 2.87 million apps were available on the Google Play store alone.[1] Two popular apps come under the categories of medical and health and fitness. Referred to collectively as mobile health or mHealth apps, such apps encompass a wide range of functions, from the management of health conditions and symptom checking to step and calorie counters and menstruation trackers.[2] Mobile health is a booming market that targets not only patients and clinicians but also those with an interest in health and fitness.

Although the potential of mHealth apps to improve access to real time monitoring and health care resources is well established,[3 4] they pose problems concerning data privacy because of the sensitive information they can access, the use of a business model that is centred on selling subscriptions or sharing user data,[5] and the lack of enforcement of privacy standards around the world. For example, the European Union General Data Protection Regulation[6] (GDPR) defines eight rights of individual users, and several rules implemented under the US Health Insurance Portability and Accountability Act[7] (HIPAA) establish a baseline of privacy protection and patient rights.

In line with the HIPAA, the US Food and Drug Administration released guidance for the postmarket management of cybersecurity in medical devices in 2016.[8] The FDA recommended that manufacturers of medical devices (ie, app developers) should incorporate risk management into the life cycle of their products and implement controls to ensure that the devices were secure and protected patients. Specifically, the guidance covers cybersecurity and privacy factors and stipulates risk management programmes 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, regulation and guidance are difficult to enforce in practice. Several recent episodes have

highlighted the problem of app data being collected and shared in an unauthorised manner. For example, a Norwegian not-for-profit organisation found that 10 popular apps, including one on health and fitness, shared data with advertising companies without informed user consent, in a clear breach of GDPR.[9] Forty one popular apps, some developed by leading technology companies, have been called out by the Chinese Ministry of Industry and Information Technology for illegal data collection.[10] A 2019 decision by CNIL, the French data protection authority, found Google to be in breach of the principle of transparency[11] because the information on the use of personal data was presented in a vague manner that was difficult to understand.

Because of the inadequate privacy disclosures of top mHealth apps,[4 12] we used a suite of app collection and analysis tools to carry out a large scale privacy analysis of mHealth apps and performed a privacy audit of more than 20 000 mHealth apps available in the Google Play store, the largest mobile app marketplace.[13]

Compared with previous analyses,[4 12 14 15] our study covers virtually all the Google Play store mHealth apps accessible from Australia, as a proxy for the worldwide Google Play app marketplace. Google Play store[16] provides various filters and configurations to developers, facilitating the localisation and distribution of releases of Android apps to specific countries or geographical locations.[17] From this information we determined that most of the collected (91.1% (19 101)) and analysed (75.7% (15 983)) mHealth apps were not specific to Australia but are also present and available in other locations such as Europe and the US. Our study was large and we also refined the granularity and depth of our analysis. For example, Dehling et al categorised mHealth apps into the low, medium, and high privacy risk groups,[18] disregarding the type of user information being leaked, the recipients of the information, and whether this was disclosed in the app's privacy policy. We considered the security of the communication protocols used by the apps, the presence of advertising and tracking libraries in the app code, and the users' reviews on the app's privacy conduct.

Methods

Since 2015, app marketplaces such as Google Play and Apple Store have grown by about 38%, and are expected to generate 111.1 billion apps by 2025.[19] The number of mHealth apps available in app stores continues to increase.[20] Of the 2.8 million apps on Google Play and the 1.96 million apps on Apple Store, an estimated 99 366 belong to medical and to health and fitness categories. These apps account for 2% (47 890) of those available through Google Play and 3% (51 476) available through the Apple store.[21 22] Our analysis focused on Google Play, the largest app store, which virtually covers all the Google Play mHealth apps accessible from Australia, as a proxy for the worldwide Google Play app marketplace.

mHealth app dataset

Google Play does not provide a complete list of mHealth apps and its search functionality does not show all the available apps. To overcome this problem and to detect as many mHealth apps as possible, we developed a crawler that interacted directly with the app store's interface.[23] Starting from the top 100 medical and health and fitness apps on Google Play, the crawler systematically searched through other apps considered to be similar by Google Play. For each app, the crawler collected several metadata: app category and price, locations where the app is available, app description, number of installs, developer information, user reviews, and app rating. From 1 October to 15 November 2019, the crawler searched through more than 1.7 million apps.

We selected apps belonging to the medical and health and fitness categories on Google Play. Overall, we identified 20 991 mHealth apps, of which 15 893 (75.7%) were free to download, 3 228 (15.4%) were purchased instore, and 1 872 (8.9%) were geoblocked (that is, could not be downloaded in Australia). In addition, we used the crawler to sample a random set of popular non-mHealth apps to be used as a baseline comparator. This set contained 8 468 apps from the tools, communication, personality, and productivity categories. Table 1 shows the dataset characteristics.

Statistical analysis

We analysed the mHealth app files and source code (static analysis), investigated the network traffic generated during execution of the app (dynamic analysis), and inspected reviews provided by users of the apps (fig 1).

App files and code analysis—of the initial set of 20 991 apps, we downloaded all 15 893 (75.7%) free apps and excluded the instore purchasable and geoblocked ones. To access the apps' resources, we processed the downloaded app packages using apktool, a tool that reverse engineers Android apps and decodes them to nearly their original form.[24] In addition, for all 15 893 mHealth apps, we extracted the app's publicly available privacy policy, which discloses the collection and use of personal data and describes the app's privacy practices. Typically, the link to the privacy policy is included in the app page on Google Play. If the link was broken or directed users to a page with no text, we considered the app to have no privacy policy. We analysed the extracted resources as follows:

Third party presence in app resources—to retrieve and classify all third party libraries included in the app, we performed a dictionary based search of the folder containing the decoded app files and embedded libraries. To achieve this, we used a comprehensive dictionary of third party libraries,[25] which comprises 338 third parties, including adverts (eg, GoogleAds); analytics (eg, GoogleAnalytics); utilities (eg, Github); and other social, banking, and gaming services (eg, Facebook or PayPal).

Data collection operations in the app code—we extracted the set of Android operating system functions

Table 1 | Characteristics of the 20 991 mHealth apps and 8468 baseline (non-mHealth) apps, collected from the Google Play store

Characteristics	No (%) of mHealth apps (n=20 991)	No (%) of non-mHealth apps (n=8468)
Medical	8074 (38)	-
Health and fitness	12 917 (62)	-
Download status:		
Instore purchase	3288 (15.4)	-
Free	15 893 (75.7)	8468 (100)
Geoblocked	1872 (8.9)	-
No of downloads:		
≥500	7481 (35.6)	1394 (17.3)
≥1000	4009 (19.1)	74 (0.9)
≥5000	1683 (8.0)	37 (0.4)
≥10 000	3582 (17.1)	206 (2.4)
≥50 000	1253 (6.0)	206 (2.4)
≥100 000	1882 (9.0)	1625 (19.2)
≥500 000	375 (1.8)	820 (9.7)
≥1 000 000	462 (2.2)	2512 (29.7)
≥5 000 000	127 (0.6)	1527 (18.1)
Contains adverts and includes tracking and analytics services (yes/no):		
All (non) mHealth apps	13 163 (63.0)/7928 (37.0)	7960 (83.2)/508 (6)
Medical apps	4516 (55.9)/3558 (44.1)	-
Health and fitness apps	8547 (66.2)/4370 (33.8)	-
Includes privacy policy link on Google Play's webpage (yes/no):		
All (non) mHealth apps	15 088 (71.9)/5904 (28.1)	6329 (74.7)/2140 (25.3)
Medical apps	5439 (67.4)/2635 (32.6)	-
Health and fitness apps	9649 (74.7)/3269 (25)	-
Users' perception (% range)*:		
0-20	10 371 (49.4)	1437 (17.0)
21-40	4157 (19.8)	30 (4.0)
41-60	2663 (12.7)	337 (4.0)
61-80	1474 (7.0)	2125 (25.1)
81-100	2326 (11.1)	4539 (53.6)

*Determined by 100%×number of negative reviews/total number of reviews.

associated with access to users' personal data. For example, the presence of the function android.telephony.TelephonyManager.getLine1Number in the app code indicates the retrieval of the user's contact phone number. In addition, we extracted the set of permissions requested by the app to access components of the operating system such as contact list or global positioning system (GPS) location. Using the permissions, we checked whether each data collection function had all the required authorisations for execution, and, if not, it was discarded. The final set of functions represented all the potential data collection in the app: in practice, it is a superset of the actual user data collection, because some parts of the

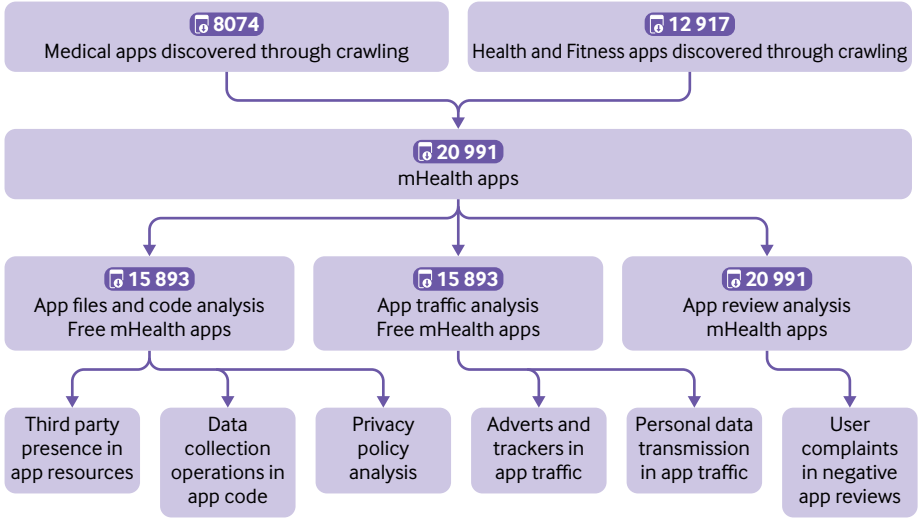

Fig 1 | Privacy analysis of mobile health (mHealth) apps

app code might rarely (or never) be triggered during execution of the app.

Privacy policy analysis—the disclosure of privacy practices is a legal requirement set by privacy regulations (eg, GDPR), and Google Play store has been mandating the inclusion of app privacy policies since 2018. Manually reviewing and annotating the app privacy policies is not feasible owing to the scale of the dataset. To overcome this, we analysed the automatic privacy policy using supervised machine learning to predict the disclosure of personal data in the privacy policy text.[26] We trained the machine learning with a large public dataset of annotated privacy policies, APP-350.[27] This is a set of 350 privacy policies of popular mobile apps annotated by legal experts. The accuracy of this method has been validated at more than 97% for all disclosure types, an average precision of 87%, and an average recall of 77%. Supplementary appendix B presents the detailed prediction performance.

Traffic analysis—we intercepted and analysed all the network traffic generated by the apps during the execution of automated app testing.[28] To achieve this, we built a dedicated testbed composed of a smartphone that connects to the internet through a computer configured as a WiFi access point, which runs a tool[29] intercepting all the traffic transmitted to the internet. Each of the 15 893 downloaded free apps were individually tested (apps purchased in-store or geoblocked were excluded): for each app, on average we performed 35 different activities (eg, opened app, opened menu, clicked on button) in a 180 second test session.

The intercepted traffic was analysed as follows:

Adverts and trackers in app traffic—we extracted the communications with external advert and tracking services—most likely third party recipients of personal data.[30] To isolate the traffic components associated with adverts and trackers, we used two comprehensive filter lists: EasyList,[31] an advert block list, and EasyPrivacy,[32] a supplementary block list for tracking.

Personal data transmission in app traffic—we identified the transmissions of user data performed by the apps during testing. A machine learning method[33] was used to find personally identifiable information in the app traffic considered to be the specific device identifier (eg, Android ID), user identifier (eg, name or email), credentials (eg, password), or location. The machine learning was trained on a large public dataset of annotated mobile app traffic flows[34] and yielded a validation accuracy of 97%, with 97% precision and 96% recall. The result only includes data collection practices that are actually performed when the app is used; this set is, however, not complete owing to coverage limitations of dynamic app testing—which might not trigger some menus, views, or functionalities of the app. For this reason, we studied the user data collection in mHealth apps by leveraging both the app code and the app traffic.

Secure transmission of user data—using the HTTP/HTTPS protocol we measured the fractions of user data transmissions. Whereas 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[35] and legislation permitting internet service providers to sell user information extracted from network traffic,[36] the adoption of the HTTPS protocol is essential to protect users' privacy.[30]

App review analysis—to obtain the complete list of reviews for each app we downloaded the content of the app's page in the Google Play store. After excluding those reviews with no text, we obtained a dataset of 2 130 684 reviews for 6 938 mHealth apps, of which 366 198 (17.2%) referred to medical apps and 1 764 486 (82.8%) to health and fitness apps. We categorised these reviews as positive (4 or 5 stars), negative (1 or 2 stars), or neutral (3 stars), resulting in 1 788 463 (83.9%) positive reviews and 235 210 (11.0%) negative reviews.

Patient and public involvement

No patients or members of the public were directly involved in the study. The subject of the study was 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 analysed public app reviews from Google Play, which were voluntarily contributed by the app users. To raise awareness of privacy risks in mHealth, we plan on sharing the collected datasets, the analyses library, and our findings with clinicians, patients, app developers, and the public.

Results

Personal data collection practices

The analysis of apps files and codes identified 65 068 data collection operations; on average four for each app. This result provided the broad set of all information that the apps can potentially access and share with third parties. At the same time, analysis of apps traffic identified 3148 transmissions of user data across 616 (3.9%) different apps. The main types of data collected by mHealth apps include contact information, user location, and several device identifiers. Part of these identifiers (specifically, international mobile equipment identity (IMEI), a unique identifier used for fingerprinting mobile phones; media access control (MAC), a unique identifier of the network interface in the user's device; and international mobile subscriber identity (IMSI), a unique number that uniquely identifies every user of a cellular network) are unique and persistent (ie, they are immutable and cannot be changed or replaced) and can be used by third parties to track users across networks and applications. Supplementary appendix A provides further details about the collected data types.

Most of the mHealth apps included codes for collecting the MAC identifiers (67.0% (14 064) of apps) and app cookies (64.0% (13 434) of apps; fig 2)—that is, small text files used for customising web browsing

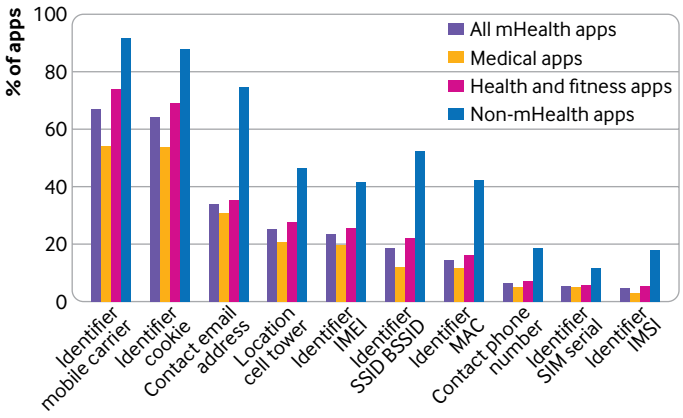

Fig 2 | Data collection operations in mobile health (mHealth) apps files and code. IMEI=international mobile equipment identity; SSID BSSID=service set identifier basic service set identifier; MAC=media access control; SIM=subscriber identity module; IMSI=international mobile subscriber identity

and app experience, but also for generating online user profiles. Other common types of data were the user's email address and current cell tower location (33.0% (6927) and 25.0% (5248) of apps, respectively). User data transmissions were observed in 3.9% (616) of mHealth apps, mostly for health and fitness apps (fig 3). This percentage is substantial and should be taken as a lower bound for the real data transmissions performed by the apps, because some transmissions might not be triggered in automated app testing. The most common transmissions were for contact (user's first or full name) and location (eg, zipcode; fig 3). When compared with baseline (non-mHealth) apps, mHealth apps, especially medical ones, were considerably less likely to collect personal data (fig 2).

Third parties that can access the personal data were also studied by distinguishing between collection on behalf of the first party (app's own entities and domains) and collection on behalf of third party services (eg, external adverts, analytics, and tracking providers). The results show a predominant role of third parties (fig 4); 54 155 of 61 920 data collection operations in the app codes (87.5%, fig 4) were related to third party services—that is, they originated from third party libraries embedded in the apps. The result might in part overestimate the actual role of these services, as some embedded libraries may never be used. The strong presence of third parties, however, was confirmed by the apps' traffic, where 1756 of 3148 detected transmissions of user data (55.8%, fig 5) were towards third party servers.

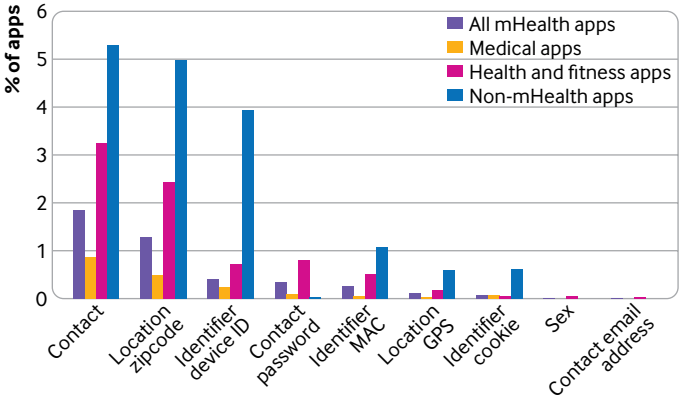

Fig 3 | Personal user data transmissions in mobile health (mHealth) app traffic. MAC=media access control; GPS=global positioning system

Third party data recipients

Overall, 665 unique third party entities were identified, of which a small list of prominent third parties (the top 50) were responsible for most data collection operations in app code, and data transmissions in app traffic (68.0% (2140), collectively).

Third party presence—in general, a strong integration (in app code and files) and interaction (in app traffic) with third parties indicated an increased collection of user data by these services. This is crucial, as these entities might also share personal information with commercial partners or transfer the information as a business asset.

To quantify the third parties in the app code, the number of third party libraries for each app was measured across the different app categories. Although 63.0% (13 224) of mHealth apps embedded at least one third party service, this proportion was substantially lower than for non-mHealth apps (table 2). In particular, only 6.0% (1260) of mHealth apps included six or more third party libraries compared with 43.0% (3641) of non-mHealth apps. Although medical and health and fitness categories showed similar trends, health and fitness apps integrated slightly more third party libraries. This difference could explain why data collection operations were less common in medical apps (fig 2).

Table 2 also reports the fractions of communications with third party services in the app traffic, focusing on advert and tracking services (other third-party services (eg, social, widgets) have negligible presence in the intercepted traffic). mHealth apps tended to have fewer interactions with advert and tracking services than non-mHealth apps. For example, advert related traffic was observed for only 5.3% (1103) of mHealth apps compared with 18.0% (1526) of non-mHealth apps. Supplementary appendix C shows the top 10 mHealth apps for presence of adverts, along with popular health and fitness apps.

Most common third parties—third party libraries Google Ads (adverts) and Google Analytics (analytics) were detected in mHealth apps code and files in 45.3% (3659) of medical apps and almost 50.0% (6453) of health and fitness apps (fig 6). Results were mainly consistent across the two mHealth app categories, although mHealth apps incorporated fewer Facebook widgets. Similarly, compared with non-mHealth apps, mHealth apps adopted SquareApp payment and Amazon services less often. The most common advert and tracking services contacted by the apps were Google ads (domains googlesyndication.com

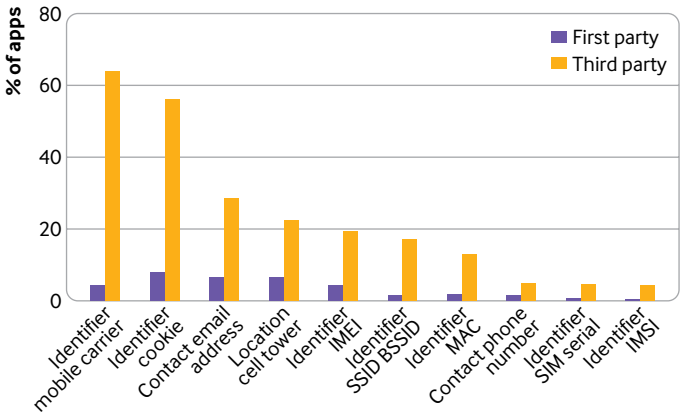

Fig 4 | Personal data recipients in mobile health (mHealth) app files and code. IMEI=international mobile equipment identity; SSID BSSID= service set identifier basic service set identifier; MAC=media access control; SIM=subscriber identity module; IMSI=international mobile subscriber identity

and doubleclick.net, which indicate the use of Google AdSense or Google Ad Manager for loading and managing adverts) and trackers (domain google-analytics.com) (fig 7).

Third party data collection in app code—a substantial fraction (34.0% (7137)) of the data collection operations in the app code were associated with Google services, and there was also a significant presence of Facebook (14.0% (2939) of apps embedded Facebook cookies), Flurry analytics (6.3% (1322) of apps), and PayPal payment service (table 3). The services most included in the app resources (eg, Google and Facebook libraries) were also prevalent in the data collection operations identified in the app code. Contact data were mainly shared with analytics services (eg, Google's crashalytics.com), whereas the location and device ID transmissions were mainly towards adverts (eg, Liftoff app marketing) and smartphone notification services (eg, Pushwoosh).

Privacy conduct issues

Privacy information disclosure—the mHealth apps were assessed for their privacy policies to check if the developers inform users about the app's data collection practices. Of the 20 991 mHealth apps, 5903 (almost 28.1%) provided no valid privacy policy text. Between the two mHealth categories, medical apps complied less with the privacy policy requirement—only 67.4% (5439) of medical apps provided privacy policies compared with 74.7% (9648) of health and fitness apps. A positive correlation was also found between an app's popularity (that is, number of installs) and the presence of a privacy policy (table 4). Around 94.4% (556) of the most popular mHealth apps (≥1 million downloads) included a privacy policy on Google Play.

Non-compliance with privacy policies—to determine whether user data transmissions complied with apps' privacy policies, each data transmission was classed as complying if the associated data collection practice was disclosed in the privacy policy, violating if the app had a privacy policy but the practice was not disclosed, and no privacy policy if the app lacked a privacy policy. Both the violating and no privacy policy cases are potentially illegal owing to breaches of privacy regulations such as the GDPR, which requires informed and unambiguous consent.[37] Overall, 55.0% (437) and 38.0% (894) of user data transmissions in medical and health and fitness apps, respectively, complied with the respective apps' privacy policies (table 5). The proportion of violations (>24.0%, 756) was consistent across the two app categories. A larger proportion of apps in the health and fitness category had no privacy policy—36.0% (847) compared with 17.0% (135) for the medical category. The apps tended to either fully comply with the privacy policy or not to comply at all. Overall, 34.0% (7136) of apps showed full compliance and 49.0% (10 286) showed no compliance either unavailable because a privacy policy was not present (21.0%, 4408) or all the user data transmissions violated the privacy policy (28.1%, 5903). Appendix D provides examples of compliant and non-compliant app behaviours for popular mHealth apps.

Insecure transmission of user data—as much as 23.0% (724) of transmissions took place on unencrypted HTTP traffic, with unencrypted transmissions being particularly common for sensitive data such as contact password and GPS location. Supplementary appendix E provides a detailed breakdown of insecure data transmission by user data type.

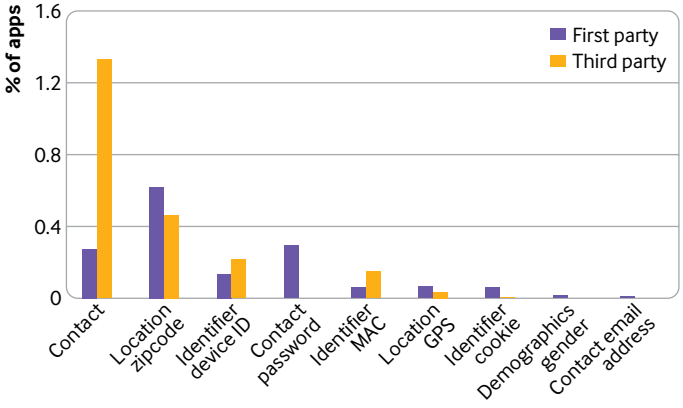

Fig 5 | First party and third party personal data transmission in mobile health (mHealth) app traffic. MAC=media access control; GPS=global positioning system

User complaints in app reviews

The main complaints raised by mHealth app users were extracted from negative app reviews (ratings with two stars). Supplementary appendix F lists 41 keywords mapped to six complaint categories that were searched through the review texts. For example, the keyword "crash" was mapped to the complaint category "bugs," whereas the keyword "private" was mapped to "privacy." A scan of the 235 210 negative reviews yielded a set of 288 238 user complaints, of which 58 349 referred to medical apps and 229 889 to health and fitness apps.

When those apps targeted by adverts, trackers, and privacy complaints were investigated further,

Table 2 | Number of third party libraries found in app code and percentage network traffic related to advert and tracker services in mobile health (mHealth) apps

	No (%) of apps			
	mHealth (n=20 991)	Medical (n=8074)	Health and fitness (n=12 917)	non-mHealth (n=8468)
No of embedded third party libraries				
0	7928 (37.8)	3558 (44.1)	4370 (33.8)	508 (6.0)
1	4618 (22.0)	1857 (23.0)	2713 (21.0)	423 (5.0)
2	2729 (13.0)	969 (12.0)	1679 (13.0)	847 (10.0)
3	1889 (9.0)	565 (7.0)	1292 (10.0)	1101 (13.0)
4	1469 (7.0)	404 (5.0)	1033 (8.0)	1016 (12.0)
5	1250 (6.0)	323 (4.0)	1033 (8.0)	931 (11.0)
≥6	1250 (6.0)	404 (5.0)	775 (6.0)	3641 (43.0)
Adverts in network traffic (% of requests)				
0.0	19 888 (94.7)	7696 (95.3)	12 087 (93.6)	6942 (82.0)
0.0-1.9	183 (0.9)	116 (1.4)	111 (0.9)	431 (5.1)
2.0-4.9	181 (0.9)	58 (0.7)	143 (1.1)	382 (4.5)
5.0-9.9	206 (1.0)	44 (0.5)	189 (1.5)	116 (1.4)
10.0-19.9	165 (0.8)	24 (0.3)	143 (1.1)	332 (3.9)
>=20.0	368 (1.8)	136 (1.7)	255 (2.0)	265 (3.1)
Trackers in network traffic (% of requests)				
0.0	19 075 (90.9)	7395 (91.6)	11 534 (89.3)	6759 (79.8)
0.0-1.9	161 (0.8)	58 (0.7)	113 (0.9)	340 (4.0)
2.0-4.9	426 (2.0)	117 (1.4)	324 (2.5)	398 (4.7)
5.0-9.9	381 (1.9)	107 (2.0)	263 (2.0)	373 (4.4)
10.0-19.9	401 (1.9)	165 (2.0)	274 (2.1)	232 (2.7)
≥20	545 (2.6)	232 (2.9)	409 (3.2)	366 (4.3)

a correlation was observed between the presence of the complaints and the actual behaviour of the app. Specifically, apps associated with complaints about adverts or trackers embedded more third party libraries, which suggests an increasing penetration of adverts and trackers. When reviews included direct

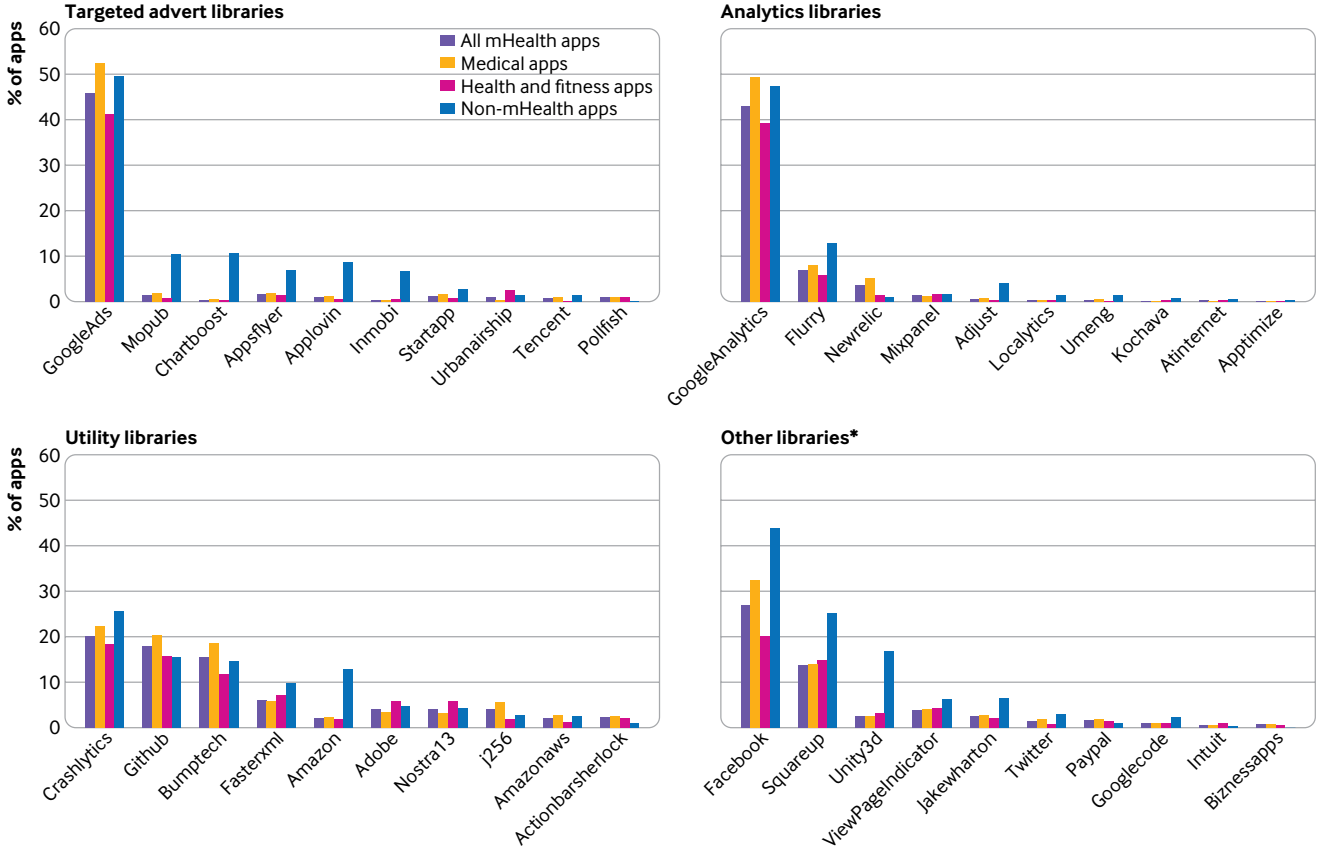

Fig 6 | Third party libraries in mobile health (mHealth) app categories and non-mHealth apps. *For example, social networks, banking, games

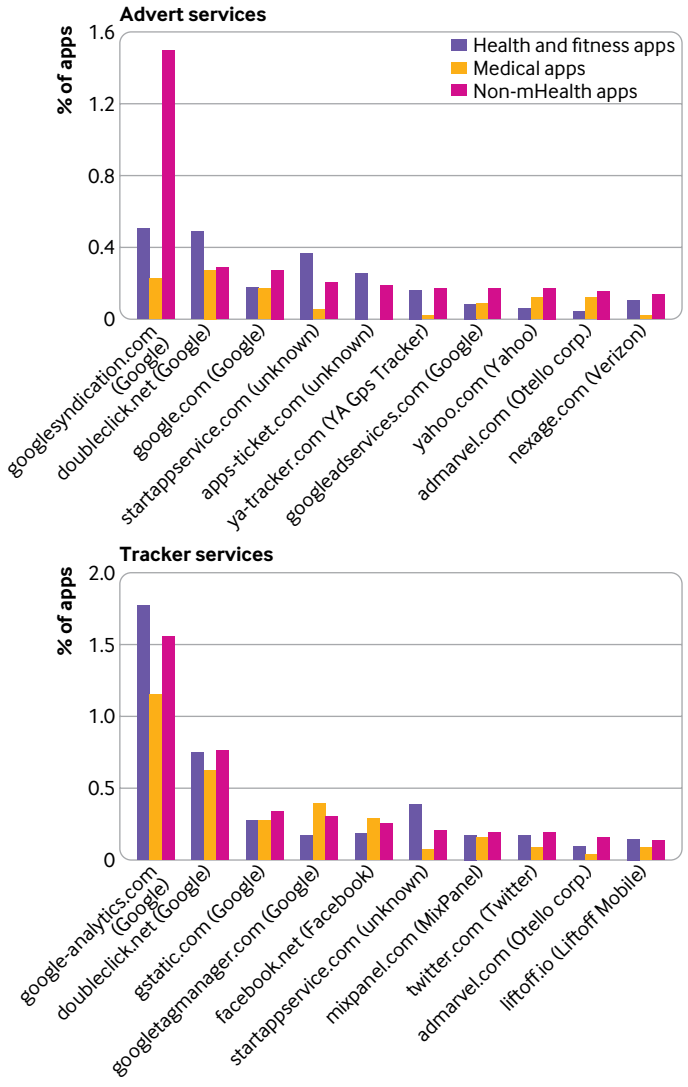

Fig 7 | Top 15 advert and tracker domains in mobile health (mHealth) and non-mHealth apps

complaints about privacy, the apps had more personal data collection operations incorporated in their code (supplementary appendix G provides further details).

Discussion

Our analysis, performed on a set of 20 991 mHealth apps, showed that most of the apps (88.0%, 18 472) could access and potentially share personal data. The transmission of user information in the app traffic was detected for 3.9% (616) of apps; however, the transmission obtained in automated app testing was a lower bound of the real data sharing by the apps. We also observed that, compared with baseline non-mHealth apps, the mHealth apps included fewer data collection operations in their code, transmitted fewer user data, and showed a reduced penetration of third party services. Health and fitness apps were generally more likely to collect and share user information than medical apps, and integration of adverts and tracking services was also more pronounced (fig 6 and fig 7).

Among the data that mHealth apps could collect, we found an important presence of persistent device identifiers and user contact information. The persistent device identifiers allowed individuals to be tracked over time and across different services, whereas the contact information directly affected an individual's privacy.

The role of third parties was predominant—more than 87.0% (54 155) of data collection practices were carried out on behalf of external services. Notably, 50 prominent services were responsible for roughly 70.0% (43 344) of the data collection operations in apps code and the data transmissions in apps traffic. In the analysed app set, Google owned services were the most common. This probably relates to the dominant position of Google's analytics and advert services and reflects the choice of Google Play store as the source of our app dataset. Android apps leverage support tools (eg, for reporting bugs) that directly report to Google, which might share additional information on devices. Hence, we would expect a slightly less pronounced role of Google for mHealth apps in the Apple store.

Although the retrieval and sharing of user information by mHealth apps were routine, data collection practices were far from transparent. Our comparative analysis of the privacy policies of the analysed mHealth apps and the actual transmissions of user information was of concern because 28.1% (5903) of the mHealth apps did not offer any privacy policy text, and at least 25% (15 480) of user data transmissions violated what was stated in the privacy policies. Another concern was the transmission of sensitive user information, such as users' fine grained geolocation (that is, GPS coordinates, 42% (26 006)) or password (75% (46 440)), using insecure communication channels. These findings are worrying given the recent reports on internet surveillance and unwanted commercialisation of user data.[8 26] Despite these issues being topical, our analysis of mHealth app reviews showed that app users seem to have a limited awareness of the privacy conduct of the apps.

Compared with user comments in the bugs category, user complaints related to privacy were less common (table 6). The reasons are, however, hard to untangle. We cannot confidently explain the limited number of 'privacy' complaints with the reduced user awareness of (or interest in) the privacy aspects, as the app reviews may not be the only nor the preferred destination for user concerns on privacy. Other channels existed, such as the contact us forms or contact details provided in the app privacy policy, or privacy regulators such as the Office of the Australian Information Commissioner.[38]

Strengths and limitations of this study

Strengths of our study included the sample size and the comparison between the behaviour of mHealth apps and that of non-mHealth (baseline) apps. We also determined the type of user information mHealth apps can retrieve and share, with our analysis building on both static app resources (application code and files) and dynamically generated app traffic.

doi: 10.1136/bmj.n1248 | *BMJ* 2021;373:n1248 | the bmj

Table 3 | Main third parties involved in user data collection practices from mobile health (mHealth) apps

Collected data	Main third parties (category) (% of mHealth apps); website			
Data collection operations in mHealth app code and files				
Identifier carrier	Google (analytics, adverts) (34.0%); google.com	Facebook (social media) (10.0%); facebook.com	Verizon (analytics) (6.3%); flurry.com	Amplitude (analytics) (3.1%); amplitude.com
Identifier cookie	Google (analytics, adverts) (20.9%); google.com	Facebook (social media) (14.0%); facebook.com	Apache (development aid) (10.7%); apache.org	PayPal (payment) (1.2%); braintreepayments.com
Contact email	Google (analytics, adverts) (21.8%); google.com	Google (analytics, adverts) (1.2%); androidquery.com	Apache (development aid) (1.3%); apache.org	Biznessapps (development aid) (1.2%); biznessapps.com
Location cell tower	Google (analytics, adverts) (11.7%); support.android	Google (analytics, adverts) (4.2%); appcompat.androidx	Facebook (social media) (1.9%); facebook.com	PayPal (payment) (2.1%); paypal.com
Identifier IMEI	New Relic (analytics) (3.8%); newrelic.com	Acra (utility) (2.2%); acra.org	Verizon (analytics) (2.1%); flurry.com	Google (analytics, adverts) (1.2%); fabric.io
Identifier SSID BSSID	Facebook (social media) (6.8%) facebook.com	Google (analytics, adverts) (2.1%); google.com	StartApp (ads) (0.9%); startapp.com	PayPal (payment) (0.8%); paypal.com
Identifier MAC	Leanium (adverts) (2.1%); learnium.com	PayPal (payment) (1.5%); paypal.com	Google (analytics, adverts) (1.1%); fabric.io	Pollfish (adverts) (1.1%); pollfish.com
Contact number	Learnium (adverts) (1.9%); learnium.com	Digits Financial (payment) (0.3%); digits.com	Mobi Mento (unknown) (0.3%); mobimento.com	PayPal (payment) (0.3%); paypal.com
Identifier SIM serial	PayPal (payment) (1.9%) paypal.com	Tencent (adverts) (0.4%); tencent.com	Swelen (adverts) (0.3%); swelen.com	Pushwoosh. (analytics) (0.2%); pushwoosh.com
Identifier IMSI	PayPal (payment) (1.7%); paypal.com	Ogury (adverts) (0.2%); presage.io	Anywhere Software (development aid) (0.2%); b4a.anywheresoftware	StartApp (adverts) (0.2%); startapp.com
User data transmission in mHealth app traffic				
Contact	Google (analytics, adverts) (2.1%) crashlytics.com	New Relic (adverts) (0.05%); newrelic.com	AgileMD (development aid) (0.04%); agilemd.com	Appioapp (analytics) (0.04%); appioapp.com
Location zipcode	Stack (adverts) (0.3%); bidmachine.io (0.07%)	Amazon (development aid) (0.2%); amazon-adsystem.com	Tapatalk (analytics) (0.1%); tapatalk.com	MobTech (analytics) (0.1%); mob.com
Identifier device ID	Pushwoosh (analytics) (0.2%); pushwoosh.com	PushBots (analytics) (0.02%); pushbots.com	InManage (development aid) (0.02%); inmanage.com	Insider (analytics) (0.01%); useinsider.com
Identifier MAC	Google (analytics) (0.1%); crashlytics.com	Axway (analytics) (0.02%); appcelerator.net	Alibaba (adverts) (0.01%); umeng.com	Jiguang-Aurora (adverts) (bf 0.01%); jpush.cn
Location GPS	Liftoff (adverts) (0.04%); liftoff.io	Kiip (adverts) (0.02%); kiip.me	Airnow Monet (adverts) (0.02%); airpush.com	Chukong Tech. (analytics) (0.01%); sdkbox.com
Contact password	Web Apps (unknown) (0.04%); fitnessitaly.com	Artexe (unknown) (0.01%); zerocoda.it	JVS Group (unknown) (0.01%); softcliniclive.com	AlleDaags (unknown) (0.01%); samenvoeden.nl

IMEI=international mobile equipment identity; SSID BSSID= service set identifier basic service set identifier; MAC=media access control; SIM=subscriber identity module; SIM= subscriber identity module; IMSI=international mobile subscriber identity; MAC=media access control; GPS=global positioning system.

To scale up the study and cope with a large number of mHealth apps, we leveraged automated analysis tools as well as modern machine learning techniques. Although the validity of the accuracy of these techniques was high (>96% for both the detection of user data transmissions and the disclosure of privacy practices), these techniques might still generate limited false positives. To deal with the scale of the app set, our live testing of mHealth apps relied heavily on extensive randomised interactions as opposed to hand crafted app usage patterns and profiles, with the drawback that some parts of the applications (eg, tabs, views, menus) might have not been triggered during testing. Owing to the number of available apps, we restricted our analyses to free apps. This restriction might have introduced a bias, because the business models of instore purchasable apps depend less on selling user data,[5] and therefore retrieve fewer user data, with a reduced presence of adverts and trackers. However, we believe that this should not have affected the generalisability of our findings, because up to 15.4% (3228) of mHealth apps found on Google Play could be purchased (table 1).

Comparison with previous studies

mHealth apps and associated privacy risks have received much attention from the research community. Huckvale et al investigated the privacy of 79 health and wellness mobile apps accredited by the UK's national health service[15] and found that most of the apps (78%, 62) that transmitted user information did not describe their data collection practices in the privacy policies. When the researchers assessed the privacy practices of 36 top ranked apps for smoking cessation and depression, they found that only a small fraction (12 of 29) disclosed the transmission of data to Facebook or Google in their privacy policies.[4] While these studies focused on consistency between the data collection practices and privacy policies of mHealth apps, the study by Grundy et al focused on the recipients of user information collected by 24 medical apps.[14] Their findings on the prevalence of analytics and advert services among user data recipients is in line with our results.

Our study analysed more than 20 000 mHealth apps on Google Play, 15 838 in detail, rather than the tens of apps assessed in previous studies.[4 12 14 15] The only other study to analyse a comparable range of mHealth apps was conducted in 2015.[18] That study, however, only categorised mHealth apps into classes of potential risk (low, medium, high risk of privacy leaks), while not providing any results on the type of user information collected, recipients of the information, and consistency of the app practices with the disclosed privacy policies.

Table 4 | Mobile health (mHealth) apps with privacy policy on Google Play store

Characteristics of apps	No (%) with privacy policy (n=15 088)	No (%) without privacy policy (n=5903)
Medical:	5439 (67.4)	2635 (32.6)
Geoblocked	730 (13.4)	208 (7.9)
Purchasable	701 (12.1)	887 (33.7)
Free	4008 (73.7)	1540 (58.4)
Health and fitness:	9648 (74.7)	3269 (25.3)
Geoblocked	745 (7.7)	189 (7.2)
Instore purchasable	910 (9.4)	728 (27.6)
Free	7993 (82.8)	2352 (89.3)
No of installs:		
‹ 100 (n=2929)	1713 (58.5)	1216 (41.5)
100-999 (n=4689)	3110 (66.3)	1579 (33.7)
1000-9 999 (n=5692)	4066 (71.4)	1626 (18.6)
≥10 000-99 999 (n=4835)	3752 (77.6)	1083 (22.4)
100 000-999 999 (n=2257)	1891 (83.8)	366 (16.2)
≥1 000 000 (n=589)	556 (94.4)	33 (5.6)

Our study presents a broad assessment of mHealth apps compared with previous studies. In previous studies, the analysis was generally restricted to the data transmitted by mHealth apps[14] or to the consistency of the apps with their privacy policies.[12 15] We analysed the privacy risks associated with mHealth apps by considering the information the apps transmit or can access through their code, the potential recipients of this information, and the correct disclosure of data sharing practices.

Considering the concentration of user data transmission towards dominant third party services, our findings on mHealth apps are aligned with recent large scale analyses of tracking and data sharing ecosystem in mobile apps.[39-41] An analysis of 959 426 apps found that most trackers embedded in the apps were linked to a small number of commercial entities, with Google the most prominent.[39] Similarly, traffic analysis of 14 599 Android apps found that despite owning just 3.9% (616) of all third party tracking services, Google was present in 50.8% (10 657) of the analysed apps.[40]

Recommendations

Our results show that the collection of personal user information is a pervasive practice in mHealth apps, and not always transparent and secure. Patients should be informed on the privacy practices of these apps and the associated privacy risks before installation and use. Clinicians should understand the main privacy aspects of mHealth apps in their specialist area, along with their key functionalities, and be able to articulate these to patients in lay language. This is important because of the scarcity of app privacy auditing tools and the substantial lack of information on the user data flows in the apps—neither Google Play store nor the Apple store currently provide such auditing functionalities.

Under these conditions, clinicians should resort to checking the permissions requested by the apps to access sensitive resources such as cameras, microphones, or locations; examine the app's privacy policy; or review the app's privacy behaviour. Previous studies suggest that privacy policies often remain unread because of their length and complicated and confusing language.[42] However, we noticed increasing research efforts towards using question answering systems to search for answers in long and verbose policy documents.[43 44] We suggest that such tools, which leverage artificial intelligence for querying privacy policies in natural language, can support clinicians in identifying relevant app privacy practices and explaining them to patients.

Besides the need for medical practitioners to familiarise themselves with the privacy aspects of mHealth apps, we believe that mobile app marketplaces, such as Google Play and the Apple store, should examine the privacy statements of apps thoroughly before the apps are available. Through a vetting process, mobile app marketplaces should ensure that a valid and meaningful privacy policy document is always provided, unlike the current situation, where we observed that the links to privacy policy pages accessible from Google Play were often broken or led to empty webpages.

Conclusions

For most of the 20 000 medical and health and fitness apps analysed, we found that most can collect and potentially share data with third parties, including

Table 5 | Consistency of data collection disclosure in privacy policy with user data transmissions in apps traffic. Values are numbers (percentages) unless stated otherwise

App category	User data transmissions	No privacy policy*	Complying†	Violating‡
All mHealth	3148 (100.0)	913 (29.0)	1479 (47.0)	756 (24.0)
Health and fitness	2353 (74.7)	847 (36.0)	894 (38.0)	613 (26.0)
Medical	795 (25.3)	135 (17.0)	437 (55.0)	223 (28.0)

*Personal data transmission not disclosed by app developers.
†Data transmission practice disclosed in app's privacy policy.
‡Practice not described in app's privacy policy.

Table 6 | Breakdown of user complaints found in reviews of mobile health (mHealth) apps. Values are numbers (percentages) unless stated otherwise

	All mHealth apps (288 238 complaints)		Medical apps (58 349 complaints)		Health and fitness apps (229 889 complaints)	
Complain category	Complaints	Apps	Complaints	Apps	Complaints	Apps
Bugs	201 240 (69.8)	2240 (10.7)	34 728 (59.5)	627 (7.8)	166 512 (72.4)	1613 (12.5)
Battery	7710 (2.7)	568 (2.7)	4784 (8.2)	120 (1.5)	2926 (1.3)	448 (3.5)
Mobile data	2058 (0.7)	427 (2.0)	169 (0.3)	70 (0.9)	1787 (0.8)	305 (2.4)
Privacy	3609 (1.3)	351 (1.7)	990 (1.7)	80 (0.9)	2619 (1.1)	271 (2.1)
Adverts	43 794 (15.2)	1128 (5.4)	10 702 (18.3)	262 (3.2)	33 092 (14.4)	866 (6.7)
Trackers	29 827 (10.3)	942 (4.5)	6976 (12.0)	138 (1.7)	22 851 (9.9)	804 (6.2)

advertising and tracking services. 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 being compliant with their declared privacy policies (if available at all).

Mobile apps are fast becoming sources of information and decision support tools for both clinicians and patients. Such privacy risks should be articulated to patients and could be made part of app usage consent. We believe the trade-off between the benefits and risks of mHealth apps should be considered for any technical and policy discussion surrounding the services provided by such apps.

Contributors: GT designed the study, led the data analysis, and wrote the first draft of the manuscript. MI secured funding, designed the study, led the data collection, and analysed the data. MI is the guarantor. KI collected the data and analysed the user reviews. MAK helped to design the study and acquired funding. SB designed the study and acquired funding. All the authors critically revised the manuscript drafts and approved the submission. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This work was funded by Optus Macquarie University Cyber Security Hub; the research was also supported by the National Health and Medical Research Council (NHMRC) grant APP1134919 (Centre for Research Excellence in Digital Health). GT and KI were supported by a postdoctoral fellowship from Macquarie University. Optus Macquarie University Cyber Security Hub and the NHMRC Centre of Research Excellence in Digital Health had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support from the Optus Macquarie University Cyber Security Hub and the National Health and Medical Research Council Centre of Research Excellence in Digital Health for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: Technical appendix, statistical code, and dataset available from the corresponding author at https://mhealthapps2020.github.io/.

The manuscript's guarantor (MI) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

Dissemination to participants and related patient and public communities: We will release all our dataset and analysis script for further research at https://mhealthapps2020.github.io/.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 AppBrain. Number of Android Apps on Google Play. 2021. https://www.appbrain.com/stats/number-of-android-apps.
2 Kay M. mhealth: New horizons for health through mobile technologies: second global survey on eHealth.World Health Organization 64.7, 2011. https://apps.who.int/iris/handle/10665/440607.
3 Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. *BMJ Open* 2017;7:e013518. doi:10.1136/bmjopen-2016-013518
4 Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Netw Open* 2019;2:e192542. doi:10.1001/jamanetworkopen.2019.2542
5 Sarah J. Iribarren, Kenrick Cato, Louise Falzon, and Patricia W Stone. What is the economic evidence for mhealth? a systematic review of economic evaluations of mhealth solutions. *PLoS One* 2017;12:e0170581. doi:10.1371/journal.pone.0170581
6 EU General Data Protection Regulation. https://gdpr-info.eu/
7 US Department of Health & Human Services. Health Information Privacy. https://www.hhs.gov/hipaa/.
8 US Food and Drug Administration. Guidance on the Postmarket Management of Cybersecurity in Medical Devices. 2016. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarket-management-cybersecurity-medical-devices/.
9 Consumer Council of Norway. Out of Control. 2020. https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf.
10 TechNode. Tencent, Xiaomi apps called out for illegal data collection. 2019. https://technode.com/2019/12/19/tencent-xiaomi-apps-called-out-for-illegal-data-collection/
11 Commission Nationale de l'Informatique et des Libertés. The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC. 2019. https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc
12 Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy policies of android diabetes apps and sharing of health information. *JAMA* 2016;315:1051-2. doi:10.1001/jama.2015.19426
13 Wang H, Liu Z, Liang J, et al. Beyond google play: A large-scale comparative study of Chinese android app markets. Proceedings of Internet Measurement Conference 2018. Association for Computing Machinery, 2018:293-307.
14 Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. *BMJ* 2019;364:l920. doi:10.1136/bmj.l920
15 Huckvale K, Prieto JT, Tilney M, Benghozi P-J, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med* 2015;13:214. doi:10.1186/s12916-015-0444-y
16 Google Play. Store website. https://play.google.com/store.
17 Play Console. Help—Distribute app releases to specific countries. https://support.google.com/googleplay/android-developer/answer/7550024?hl=en&ref_topic=7071529.
18 Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on ios and android. *JMIR Mhealth Uhealth* 2015;3:e8. doi:10.2196/mhealth.3672
19 Zion Market Research. Global mHealth Apps Market Will Reach USD 111.1 Billion By 2025. 2019. https://www.globenewswire.com/news-release/2019/01/24/1704860/0/en/Global-mHealth-Apps-Market-Will-Reach-USD-111-1-Billion-By-2025-Zion-Market-Research.html.
20 Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. *JMIR Mhealth Uhealth* 2016;4:e96. doi:10.2196/mhealth.6020

21 Statista. Number of mHealth apps available at Google Play from 1st quarter 2015 to 4th quarter 2020. 2020. https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/.
22 Statista. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 4th quarter 2020. 2020. https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/.
23 Hass T. Web crawler 101: what is a web crawler and how do crawlers work?. 2019. https://webfx.com/blog/internet/what-is-a-web-crawler/
24 Apktool. A tool for reverse engineering Android apk files. https://ibotpeaches.github.io/Apktool/
25 Ikram M, Kâafar MA. A first look at mobile ad-blocking apps [p 343-50]. 16th IEEE International Symposium on Network Computing and Applications; 2017 Oct 30-Nov 1.
26 Zimmeck S, Story P, Smullen D, et al. MAPS: Scaling privacy compliance analysis to a million apps. Privacy Enhancing Technologies Symposium (PETS 2019). *Sciendo* 2019;3:66-86.
27 Usable Privacy. Data and Tools. https://usableprivacy.org/data.
28 Ikram M, Vallina-Rodriguez N, Seneviratne S, Kaafar MA, Paxson V. An analysis of the privacy and security risks of android vpn permission-enabled apps [p 349-64]. Proceedings of 2016 Internet Measurement Conference. ACM, 2016.
29 Mitmproxy. Homepage. https://mitmproxy.org
30 Ren J, Lindorfer M, Dubois DJ, Rao A, Choffnes D, Vallina-Rodriguez N. Bug fixes, improvements, ... and privacy leaks: A longitudinal study of pii leaks across android app versions. NDSS, 2018.
31 Easylist. https://easylist.to/easylist/easylist.txt
32 Easyprivacy. General tracking systems. https://easylist.to/easylist/easyprivacy.txt
33 Ren J, Rao A, Lindorfer M, Legout A, Choffnes D. ReCon: Revealing and controlling PII leaks in mobile network traffic [p 361-74]. MobiSys 2016 - Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, 2016.
34 Choffnes D. Controlled experiments code and data. 2018. https://web.archive.org/web/20180925081031/https://recon.meddle.mobi/codeanddata.html
35 Ball J, Schneier B, Greenwald G. NSA and GCHQ target Tor network that protects anonymity of web users. http://www.theguardian.com/world/2013/oct/04/nsa-gchq-attack-tor-network-encryption. *Guardian* 2013.
36 NPR. Congress Overturns Internet Privacy Regulation. 2017. https://www.npr.org/2017/03/28/521831393/congress-overturns-internet-privacy-regulation
37 GDPR.eu. What are the GDPR consent requirements? https://gdpr.eu/gdpr-consent-requirements
38 Office of the Australian Information Commissioner. Homepage. https://www.oaic.gov.au/
39 Binns R, Lyngs U, Van Kleek M, Zhao J, Libert T, Shadbolt N. Third party tracking in the mobile ecosystem. *WebSci* 2018.
40 Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, Sundaresan S, Allman M, Kreibich C, Gill P. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. Network and Distributed System Security Symposium (*NDSS)*, 2018.
41 Vallina-Rodriguez N, Sundaresan S, Razaghpanah A, Nithyanand R, Allman M, Kreibich C, Gill P. Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem. *arXiv* [Preprint] 2016;1609.07190.
42 Joel R. Reidenberg, Jaspreet Bhatia, Travis D Breaux, and Thomas B Norton. Ambiguity in privacy policies and the impact of regulation. *J Legal Stud* 2016;45(S2):S163-90. doi:10.1086/688669
43 Ahmad WU, Chi J, Tian Y, Chang K-W. Policyqa: A reading comprehension dataset for privacy policies. *arXiv* [Preprint] 2020;2010.02557.
44 Ravichander A, Black AW, Wilson S, Norton T, Sadeh N. Question answering for privacy policies: Combining computational and legal perspectives [p 4947-58]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing. Association for Computational Linguistics, November 2019.

Supplementary material: appendices A-G