Data sharing practices of medicines related apps and the mobile ecos stem: tra c, content, and network anal sis

Quinn Grundy,[1,2] Kellia Chiu,[2] Fabian Held,[2] Andrea Continella,[3] Lisa Bero,[2] Ralph Holz[4]

[1]Faculty of Nursing, University of Toronto, Suite 130, 155

ABSTRACT

OBJECTIVES

To investigate whether and how user data are shared by top rated medicines related mobile applications (apps) and to characterise privacy risks to app users, both clinicians and consumers.

DESIGN

Traffic, content, and network analysis.

SETTING

Top rated medicines related apps for the Android mobile platform available in the Medical store category of Google Play in the United Kingdom, United States, Canada, and Australia.

PARTICIPANTS

24 of 821 apps identified by an app store crawling program. Included apps pertained to medicines information, dispensing, administration, prescribing, or use, and were interactive.

INTERVENTIONS

Laboratory based traffic analysis of each app downloaded onto a smartphone, simulating real world use with four dummy scripts. The app's baseline traffic related to 28 different types of user data was observed. To identify privacy leaks, one source of user data was modified and deviations in the resulting traffic observed.

MAIN OUTCOME MEASURES

Identities and characterisation of entities directly receiving user data from sampled apps. Secondary content analysis of company websites and privacy policies identified data recipients' main activities; network analysis characterised their data sharing relations.

RESULTS

19/24 (79%) of sampled apps shared user data. 55 unique entities, owned by 46 parent companies, received or processed app user data, including developers and parent companies (first parties) and service providers (third parties). 18 (33%) provided infrastructure related services such as cloud services. 37 (67%) provided services related to the collection and analysis of user data, including analytics or advertising, suggesting heightened privacy risks. Network analysis revealed that first and third parties received a median of 3 (interquartile range 1-6, range 1-24) unique transmissions of user data. Third parties advertised the ability to share user data with 216 "fourth parties"; within this network (n=237), entities had access to a median of 3 (interquartile range 1-11, range 1-140) unique transmissions of user data. Several companies occupied central positions within the network with the ability to aggregate and re-identify user data.

CONCLUSIONS

Sharing of user data is routine, yet far from transparent. Clinicians should be conscious of privacy risks in their own use of apps and, when recommending apps, explain the potential for loss of privacy as part of informed consent. Privacy regulation should emphasise the accountabilities of those who control and process user data. Developers should disclose all data sharing practices and allow users to choose precisely what data are shared and with whom.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Developers of mobile applications (apps) routinely, and legally, share user data

Most health apps fail to provide privacy assurances or transparency around data sharing practices

User data collected from apps providing medicines information or support may be particularly attractive to cybercriminals or commercial data brokers

WHAT THIS STUDY ADDS

Medicines related apps, which collect sensitive and personal health data, share user data within the mobile ecosystem in much the same way as other types of apps

A small number of companies have the potential to aggregate and perhaps re-identify user data owing to their network position

Introduction

Journalists recently revealed that Australia's most popular medical appointment booking app

Mobile health apps are a booming market targeted at both patients and health professionals. These apps claim to o er tailored and cost e ective health promotion, but they pose unprecedented risk to consumers' privacy given their ability to collect user data, including sensitive information. Health app developers routinely, and legally, share consumer data with third parties in exchange for services that enhance the user's experience (eg, connecting to social media) or to monetise the app (eg, hosted advertisements). Little transparency exists around third party data sharing, and health apps routinely fail to provide privacy assurances, despite collecting and transmitting multiple forms of personal and identifying information.

Third parties may collate data on an individual from multiple sources. Threats to privacy are heightened when data are aggregated across multiple sources and consumers have no way to identify whether the apps or websites they use share their data with the same third party providers. Collated data are used to populate proprietary algorithms that promise to deliver "insights" into consumers. Thus, the sharing of user data ultimately has real world consequences in the form of highly targeted advertising or algorithmic decisions about insurance premiums, employability, nancial services, or suitability for housing. These decisions may be discriminatory or made on the basis of incomplete or inaccurate data, with little recourse for consumers.

Apps that provide medicines related information and services may be particularly likely to share or sell data, given that these apps collect sensitive, speci c medical information of high value to third parties. For example, drug information and clinical decision support apps that target health professionals are of particular interest to pharmaceutical companies, which can o er tailored advertising and glean insights into prescribing habits. Drug adherence apps targeting consumers can deliver a detailed account of a patient's health history and behaviours related to the use of medicines.

We investigated the nature of data transmission to third parties among top rated medicines related apps, including the type of consumer data and the number and identities of third parties, and we characterised the relations among third parties to whom consumer data are transmitted.

Methods

We carried out this study in two phases: the rst was a tra c analysis of the data sharing practices of the apps and the second was a content and network analysis to characterise third parties and their interrelations (box).

Sampling

We purposefully sampled medicines related apps that were considered prominent owing to being highly downloaded, rated in the top , or endorsed by credible organisations. During October to November , we triangulated two sampling strategies to identify apps. In the rst strategy we used a crawling program that interacted directly with the app store's application programming interface. This program systematically sampled the metadata for the top ranked free and paid apps from the Medical store category of the United Kingdom, United States, Australian, and Canadian Google Play stores on a weekly basis. In the second strategy we screened for recommended or endorsed apps on the website

activities, data sharing partnerships, and privacy practices related to user data into an open ended form in RedCap. Data were extracted between February and July ; one investigator extracted data before, and the other after, the General Data Protection Rules (GDPR) were implemented in the European Union in May , which meant that some developers disclosed additional data sharing partnerships in their privacy policies. Any discrepancies were resolved through consensus or consolidation and by taking the more recent information as accurate.

Data analysis

We classi ed entities receiving user data into three categories: rst parties, when the app transmitted user data to the developer or parent company (users are considered second parties); third parties, when the app directly transmitted user data to external entities; and fourth parties, companies with which third parties reported the ability to further share user data. We calculated descriptive statistics in Excel (Microsoft) for all app and company characteristics. Using NVivo (QSR International), we coded unstructured data inductively, and iteratively categorised each company based on its main activities and self reported business models.

Network analysis

We combined data on apps and their associated rst, third, and fourth parties into two networks. Network analysis was conducted using R, and the igraph (. .) library for network analysis and tidygraph (. .) for visualisation. The rst network represented apps and entities that directly received data (rst and third parties), as identi ed by our tra c and privacy policy analysis. We use descriptive statistics to describe the network's data sharing potentiale app trewoicsnetwo0.5 (tw 8.r)10ented pps a749ee .5 (thirsn)10tion83 eacin10 (ei(, ar
10 (om 0 -1.2A)4.47CCAN 0 -1.2wn)10as 0p-1.2 h)10* [(tri(tim)5(gi)s10 10 c) -10.2n)'10 (65 473 5w) 65 (10 (s h)]TJ 0.tatistr 6 T*[(0d(r)]TJ 8(g 0 (1)2 0 (e2)1899(bmpa(
0d ear310 (R)-19.llecii(edc)1 r10 (on 9ti ed by anae)5 (g)ii 141.7;

.10.1136/.1 . 20 r . 2019;364 20 bmj

providers or analysis providers. Infrastructure related entities provided services such as cloud computing, networks, servers, internet, and data storage. Analysis entities provided services related to the collection, collation, analysis, and commercialisation of user data in some capacity.

Recipients of user data

Through tra c and privacy policy analysis, we identi ed unique entities that received or processed user data, which included app developers, their parent companies, and third parties. We classi ed app developers and their parent companies as " rst parties"; these entities have access to user data through app or company ownership, or both. Although rst parties collected user data to deliver and improve the app experience, some of these companies also described commercialising these data through advertising or selling deidenti ed and aggregated data or analyses to pharmaceutical companies, health insurers, or health services.

Developers engaged a range of third parties who directly received user data and provided services, ranging from error reporting to in-app advertising to processing customer service tickets. Most of these services were provided on a "freemium" basis, meaning that basic services are free to developers, but that higher levels of use or additional features are charged.

Third parties typically reserved the right to collect deidenti ed and aggregated data from app users for their own commercial purposes and to share these

A systems view of privacy

While certain data sources are clearly sensitive, personal, or identifying (eg, date of birth, drug list), others may seem irrelevant from a privacy perspective (eg, device name, Android ID). When combined, however, such information can be used to uniquely identify a user, even if not by name. Thus, we conducted a network analysis to understand how user data might be aggregated. We grouped the entities identi ed in the tra c analysis into "families" based on shared ownership, presuming that data as an asset was shared among acquiring, subsidiary, and a liated companies as was explicitly stated in most privacy policies. For example, the family "Alphabet," named for the parent company, is comprised of Google.com, Google Analytics, Crashlytics, and AdMob by Google.

Third party sharing

Supplementary gure displays the results of the network analysis containing apps, and families of rst and third parties that receive user data and are owned by the same parent company. The size of the entity indicates the volume of user data it sends or receives. We di erentiated among apps (orange), companies whose main purpose in receiving data was for analysis, including tracking, advertising, or other analytics (grey), and companies whose main purpose in receiving data was infrastructure related, including data storage, content delivery networks, and cloud services (blue).

From the sampled apps, rst and third parties received a median of (interquartile range - , range -) unique transmissions of user data, de ned as sharing of a unique type of data (eg, Android ID, birthdate, location) with a rst or third party. Amazon.com and Alphabet (the parent company of Google) received the highest volume of user data (both received n=), followed by Microsoft (n=). First and third parties received a median of (interquartile range - ; range -) di erent types of user data from the sampled apps. Amazon.com and Microsoft, two cloud service providers, received the greatest variety of user data (and types, respectively), followed by the app developers Talking Medicines (n=), Ada Health (n=), and MedAdvisor International (n=).

Fourth party sharing

Supplementary gure displays the results of a network analysis conducted to understand the hypothetical data sharing that might occur within the mobile ecosystem at the discretion of app developers, owners, or third parties. Analysis of the websites and privacy policies of third parties revealed additional possibilities for sharing app users' data, described as "integrations" or monetisation practices related to data (eg, Facebook disclosed sharing end user data with data brokers for targeted advertising). Integrations allowed developers

10.1136/ 2019;364 bmj

sharing partnerships with Nielsen, comScore, Kanta, and RN SSI Group for the purpose of "advertising and ad measurement purposes, using their own cookies or similar technologies." These partners "can collect or receive non-personally identi able information about your browser or device when you use Google sites and apps." Table exempli es the risks to privacy as a result of data aggregation within the fourth party network.

Discussion

Our analysis of the data sharing practices of top rated medicines related apps suggests that sharing of user data is routine, yet far from transparent. Many types

Table 5 | Top 10 companies receiving user data by number of apps

Company	Sector	No of apps receiving user data directly	No of apps able to receive user data indirectly	No of di[illegible]erent

been updated, or might have changed their data sharing practices. We purposefully sampled apps to include widely downloaded ones that were likely to collect and share user data (ie, requested "dangerous" permissions and had some degree of user interactivity). It is not, however, known how the data sharing practices of these apps compare with those of mobile health apps in general. A strength of this approach was in-depth use of the app using simulated user input, including logging in and interacting with the app while it was running. The use of the Agrigento tool allowed detection of privacy leaks that were obfuscated by encoding or encryption, for example. This sample is not representative of medicines related apps as a population; however, this approach benefited from focusing on the medicines related apps likely to be used by clinicians and consumers. Because all apps were available to the public and many had multiple functionalities and target users, we could not clearly classify apps as targeted at consumers or health professionals and randomised the simulated user profiles irrespective of target user group. Thus, it is not known whether or how patterns in user data collection and sharing differ among target user groups, which is an important question for future research. Our analysis was restricted to Android apps, thus it is not known whether the iOS versions of these apps or medicines related apps developed exclusively for iPhone differ in data sharing practices. Future work might explore the role of Alphabet (the parent company of Google) within a data sharing network of iOS apps to see whether its dominance is associated with the type of operating system. Our characterisation of the main activities and data sharing relations of entities is based on developers' self reported practices at the time of analysis and represents our interpretation of these materials. Data were, however, extracted in duplicate and discussed to ensure interpretation was robust.

Comparison with other studies

Our findings are consistent with recent large scale, crowd sourced analyses of app sharing of user data. An

5 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
6 Grindrod K, Boersema J, Waked K, Smith V, Yang J, Gebotys C. Locking it down: The privacy and security of mobile medication apps. *Can Pharm J* 2016;150:60-6. doi:10.1177/1715163516680226
7 Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C,