Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis

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 College St, Toronto, ON, Canada, M5T 1P8
[2]School of Pharmacy, Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia
[3]Department of Computer Science, University of California, Santa Barbara, CA, USA
[4]School of Computer Science, The University of Sydney, Sydney, NSW, Australia
Correspondence to: Q Grundy
quinn.grundy@utoronto.ca
(or @quinngrundy on Twitter
ORCID 0000-0002-7640-8614)

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

Cite this as: *BMJ* 2019;364:l920
http://dx.doi.org/10.1136/bmj.l920

Accepted: 25 February 2019

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, HealthEngine, routinely shared 100s of users' private medical information to personal injury law firms as part of a referral partnership contract.[1] Although the company claimed this was only done with users' consent, these practices were not included in the privacy policy but in a separate "collection notice," and there was no opportunity for users to opt-out if they wished to use the application (app).[1]

Mobile health apps are a booming market targeted at both patients and health professionals.[2] These apps claim to offer tailored and cost effective 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).[3 4] 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.[5-9]

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.[3] 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, financial 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.[10 11]

Apps that provide medicines related information and services may be particularly likely to share or sell data, given that these apps collect sensitive, specific medical information of high value to third parties.[12] For example, drug information and clinical decision support apps that target health professionals are of particular interest to pharmaceutical companies, which can offer tailored advertising and glean insights into prescribing habits.[13] Drug adherence apps targeting consumers can deliver a detailed account of a patient's health history and behaviours related to the use of medicines.[14]

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 first was a traffic 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 1).

Sampling

We purposefully sampled medicines related apps that were considered prominent owing to being highly downloaded, rated in the top 100, or endorsed by credible organisations. During 17 October to 17 November 2017, we triangulated two sampling strategies to identify apps. In the first 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 100 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 of an Australian medicines related not-for-profit

Box 1 Description of methods

- Differential traffic analysis
- Aim: to intercept and analyse data sent by apps to destinations on the internet
- Data sources: 24 apps downloaded to a Google Pixel 1 running Android 7.1
- Tools: Agrigento framework (https://github.com/ucsb-seclab/agrigento), a set of programs that allows monitoring of data transmission from app to network without interfering with the app program
- Procedures:
  - Simulation of user interaction by adoption of a dummy user profile and exploration of all features of the app
  - App run 14 times to establish a baseline of its data sharing behaviour
  - Alteration of one source of user information, such as device ID or location, and app run for 15th time
  - Observation for any deviations in network traffic compared with baseline behaviour, defined as a privacy leak
  - 15th run repeated for each of 28 prespecified sources of sensitive user information, altering one source for each run
- Analysis:
  - Privacy leaks inferred when sensitive information was sent to a remote server, outside of the app
  - Companies receiving sensitive user data identified by their IP addresses using the WHOIS, Shodan, and GeoIP databases

Content and network analysis

- Aims: to describe the characteristics of companies receiving sensitive user data and their data sharing relations from a systems perspective
- Data sources: Crunchbase profiles, developers' websites, company social media profiles, news media articles, app privacy policies, and terms and conditions
- Tools: author generated data extraction form in RedCap, analysis in R (3.5.2) using tidygraph (1.1.1)
- Procedures:
  - Two investigators, working independently, extracted data into the RedCap form
  - One investigator collected data before, and one after, implementation of the General Data Protection Rules (GDPR)
  - Collated extracted data, resolved errors, and took more recent information in case of discrepancy
  - Documented additional data sharing relations found in app privacy policies
- Analysis:
  - Descriptive analysis of company characteristics
  - Quantitative, descriptive analysis of data sharing among apps and third parties identified in the traffic analysis
  - Simulation of the potential distribution of user data among apps (presuming one person used all the apps in the sample), third parties identified in the traffic analysis, and "fourth parties" that can integrate with third parties
  - Calculated the number of data sources an entity could access directly from an app, or indirectly through a data sharing partnership with an intermediary

organisation, a curated health app library, a published systematic review, and personal networks of practising pharmacists.

One investigator screened 821 apps for any app names that were potentially related to medicines (ie, managing drugs, adherence, medicines or prescribing information) and excluded apps with irrelevant names (eg, "Pregnancy Calendar," "Gray's Anatomy–Atlas," "Easy stop Smoking," "Breathing Zone") (fig 1). Two investigators then independently screened 67 app store descriptions according to the following inclusion criteria:

- Pertains to medicines, such as managing drugs, adherence, medicines or prescribing information
- Available for the Android mobile platform in Google Play to an Australian consumer
- Requests at least one "dangerous" permission, as defined by Google Play,[15] or claims to collect or share user data
- Has some degree of interactivity with the user, defined as requiring user input.

We excluded apps if they were available exclusively to customers of a single company (pharmacy, insurance plan, or electronic health record), were targeted at or restricted to use in a single country (ie, a formulary app for UK health professionals employed by the National Health Service), were prohibitively expensive (>$100; >£76; >€88), or were no longer available during the analysis period.

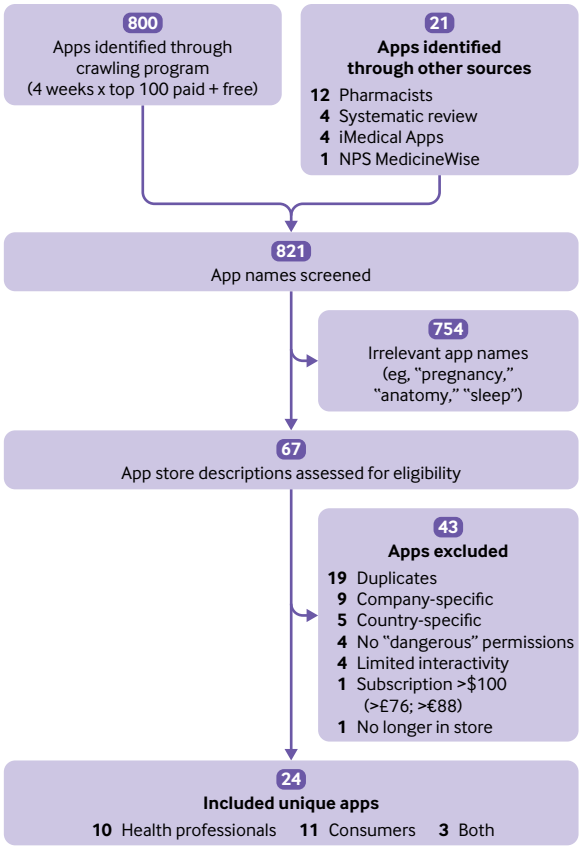

Fig 1 | Sampling flow diagram for prominent medicines related apps

Data collection

Traffic analysis

The methods of the traffic analysis are described in detail elsewhere.[16] For this analysis we made use of Agrigento, a tool for detecting obfuscated privacy leaks such as encoding or encryption in Android apps. In a laboratory setting, between November and December 2017, we downloaded each app onto a Google Pixel 1 smartphone running Android 7.1. We purchased subscriptions when required (in the form of in-app purchases).

Between December 2017 and January 2018 we simulated real world, in-depth use of the app using four dummy scripted user profiles (one doctor, one pharmacist, and two consumers; see supplementary file), including logging in and interacting with the app while it was running, which involved manually clicking on all buttons, adjusting all settings, and inputting information from the dummy profile when applicable. As all apps were available to the public, we randomised the dummy user profiles irrespective of the app's target user group.

Using one randomly assigned dummy scripted user profile for each app, we ran the app 14 times to observe its "normal" network traffic related to 28 different prespecified types of user data, such as Android ID, birthday, email, precise location, or time zone. Fourteen executions of the app were required to establish a baseline and to minimise the occurrence of false positives.[16] Then we modified one aspect of the user's profile (eg, location) and ran the app a 15th time to evaluate any change in the network traffic. This differential analysis allowed the detection of an incidence of user data sharing by observing any deviations in network traffic. Change in traffic during the 15th run indicated that the modified aspect of the user's profile was communicated by the app to the external network, meaning that user data were shared with a third party. We repeated the 15th run for each of the 28 prespecified types of user information, altering one type of data for each run.

The results of the traffic analysis included a list of domain names and respective IP addresses receiving user data and the specific types of user data they received. We identified the recipients of user data by integrating Agrigento with Shodan, a search engine for servers, to obtain geographical information for IP addresses. To reveal the identity of the entities involved, we used the public WHOIS service, a database of domain registrations. Leveraging these tools, we were able to obtain information about the hosts that receive data from the apps, such as location and owner of the remote server.

Content analysis

For each of the entities receiving user data in the traffic analysis, two investigators independently examined their Crunchbase profile, company website, and linked documents such as privacy policies, terms and conditions, or investor prospectus. The investigators extracted data related to the company's mission, main

activities, data sharing partnerships, and privacy practices related to user data into an open ended form in RedCap.[17] Data were extracted between 1 February 2018 and 15 July 2018; one investigator extracted data before, and the other after, the General Data Protection Rules (GDPR) were implemented in the European Union in May 2018, which meant that some developers disclosed additional data sharing partnerships in their privacy policies.[18] Any discrepancies were resolved through consensus or consolidation and by taking the more recent information as accurate.

Data analysis

We classified entities receiving user data into three categories: first 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 2016 (Microsoft) for all app and company characteristics. Using NVivo 11 (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 first, third, and fourth parties into two networks. Network analysis was conducted using R, and the igraph (1.0.1) library for network analysis and tidygraph (1.1.1) for visualisation.[19 20] The first network represented apps and entities that directly received data (first and third parties), as identified by our traffic and privacy policy analysis. We use descriptive statistics to describe the network's data sharing potential.

The second network represents the potential sharing of user data within the mobile ecosystem, including to fourth parties. To simplify the representation, we grouped apps, their developers, and parent companies into "families" based on shared ownership, and we removed ties to third parties that only provided infrastructure services as they did not report further data sharing partnerships with fourth parties. We report third and fourth parties' direct and indirect access to app users' data and summarise the scope of data potentially available to third and fourth parties through direct and indirect channels. This simulation assumes that the same person uses all apps in our sample and it shows how her or his data get distributed and multiplied across the network, identifying the most active distributors of data and the companies that occupy favourable positions in the network, enabling each to gather and aggregate user data from multiple sources.

Patient and public involvement

We undertook this research from the perspective of an Australian app user and in partnership with the Australian Communications Consumer Action Network (ACCAN), the peak body for consumer representation in the telecommunications sector. In continuation of an existing partnership,[21] we jointly applied for funding from the Sydney Policy Lab, a competition designed to support and deepen policy partnerships. A representative from ACCAN was involved in preparing the funding application; designing the study protocol, including identifying outcomes of interest; team meetings related to data collection and analysis; preparing dissemination materials targeted at consumers; and designing a dissemination strategy to consumers and regulators.

Results

Overall, 24 apps were included in the study (table 1). Although most (20/24, 83%) appeared free to download, 30% (6/20) of the "free" apps" offered in-app purchases and 30% (6/20) contained advertising as identified in the Google Play store. Of the for-profit companies (n=19), 13 had a Crunchbase profile (68%).

Data sharing practices

As per developer self report in the Google Play store, apps requested on average 4 (range 0-10) "dangerous" permissions—that is, data or resources that involve the user's private information or stored data or can affect the operation of other apps.[15] Most commonly, apps requested permission to read or write to the device's storage (19/24, 79%), view wi-fi connections (11/24, 46%), read the list of accounts on the device (7/24, 29%), read phone status and identity, including the phone number of the device, current cellular network information, and when the user is engaged in a call (7/24, 29%), and access approximate (6/24, 25%) or precise location (6/24, 25%).

In our traffic analysis, most apps transmitted user data outside of the app (17/24, 71%). Of the 28 different types of prespecified user data, apps most commonly shared a user's device name, operating system version, browsing behaviour, and email address (table 2). Out of 104 detected transmissions, aggregated by type of user data for each app, 98 (94%) were encrypted and six (6%) occurred in clear text. Out of 24 sampled apps, three (13%) leaked at least one type of user data in clear text, whereas the remainder 14 (58%) only transmitted encrypted user data (over HTTPS) or did not transmit user data in the traffic analysis (7/24, 29%). After implementation of the GDPR, developers disclosed additional data sharing relations within privacy policies, including for two additional apps that had not transmitted any user data during the traffic analysis. Thus, a total of 19/24 (79%) sampled apps shared user data (see supplementary table 2).

Table 3 displays the data sharing practices of the apps (see supplementary table 2 for overview of data sharing practices) detected in the traffic analysis and screening of privacy policies. We categorised first and third parties receiving user data as infrastructure

doi: 10.1136/bmj.l920 | *BMJ* 2019;364:l920 | the bmj

Table 1 | App characteristics

Characteristic	No (%)
Category*:	
Consumer medicines information	13 (54)
Clinician drug reference	12 (50)
Drug record	12 (50)
Drug adherence and reminders	8 (33)
Health information/symptom checker	5 (21)
Message health professional	5 (21)
Dose calculator	4 (17)
Pill identifier	4 (17)
Ordering prescription refills	3 (13)
Drug coverage/pricing	3 (13)
No of downloads†:	
500-1000	3 (13)
1000-5000	3 (13)
5000-10 000	4 (17)
10 000-50 000	1 (4)
50 000-100 000	2 (8)
100 000-500 000	6 (25)
500 000-1 000 000	1 (4)
1 000 000-5 000 000	3 (13)
5 000 000-10 000 000	1 (4)
Cost incurred to download:	
No	20 (83)
Yes	4 (17)
Contains advertising:	
No	18 (75)
Yes	6 (25)
Offers in-app purchases:	
No	18 (75)
Yes	6 (25)
Has a privacy policy:	
No	2 (8)
Yes	22 (92)
Type of developer:	
Privately held company	15 (63)
Publicly traded company or subsidiary	4 (17)
Individual	3 (13)
Not-for-profit organisation	2 (8)
Location:	
North America	14 (58)
Australia/New Zealand	7 (29)
Europe	2 (8)
China	1 (4)
Clinician involvement:	10 (38)
Founder	8 (33)
Peer reviewer	4 (17)

*Apps often had multiple functionalities therefore percentages do not add to 100%.
†As reported in Google Play store at time of sampling (November 2017).

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 traffic and privacy policy analysis, we identified 55 unique entities that received or processed user data, which included app developers, their parent companies, and third parties. We classified app developers and their parent companies as "first parties"; these entities have access to user data through app or company ownership, or both. Although first parties collected user data to deliver and improve the app experience, some of these companies also described commercialising these data through advertising or selling deidentified 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 deidentified and aggregated data from app users for their own commercial purposes and to share these data among their commercial partners or to transfer data as a business asset in the event of a sale. For example, Flurry analytics, offered by Yahoo! helps developers to track new users, active users, sessions, and the performance of the app, and offers this service free of charge. In exchange, developers grant Flurry "the right for any purpose, to collect, retain, use, and publish in an aggregate manner . . . characteristics and activities of end users of your applications."[22] In our sample, Flurry collected Android ID, device name, and operating system version from one app; however, its privacy policy states that it may also collect data about users, including users' activity on other sites and apps, from their parent company Verizon Communications, advertisers, publicly available sources, and other companies. These aggregated and pseudonymous (eg, identified by Android ID) data are used to match and serve targeted advertising and to associate the user's activity across services and devices, and these data might be shared with business affiliates.[22]

We categorised 18 entities (18/55, 33%) as infrastructure providers, which included cloud services (Amazon Web Services, Microsoft Azure), content delivery networks (Amazon CloudFront, CloudFlare), managed cloud providers (Bulletproof, Rackspace, Tier 3), database platforms (MongoDB Cloud Services), and data storage centres (Google). Developers relied on the services of infrastructure related third parties to securely store or process user data, thus the risks to privacy are lower. However, sharing with infrastructure related third parties represents additional attack surfaces in terms of cybersecurity. Several companies providing cloud services also offered a full suite of services to developers that included data analytics or app optimisation, which would involve accessing, aggregating, and analysing app user data. The privacy policies of these entities, however, stated this would occur within the context of a relationship with the developer-as-client and thus likely does not involve commercialising app user data for third party purposes.

We categorised 37 entities (37/55, 67%) as analysis providers, which involved the collection, collation, analysis, and commercialisation of user data in some capacity. Table 4 characterises these analysis providers based on their main business activities.

Table 2 | Types and frequency of user data shared with third parties in traffic analysis

User data type	Explanation	No (%) of apps sharing*
Device name	Name of device (eg, Google Pixel)	15 (63)
OS version	Version of device's Android operating system	10 (42)
Browsing	App related activity performed by user (eg, view pharmacies, search for medicines)	9 (38)
Email†‡	User's email address	9 (38)
Android ID†‡	Unique ID to each Android device (ie, used to identify devices for market downloads)	8 (33)
Drugs list‡	List of drugs taken by user	6 (25)
Name/Last name†‡	User's name and/or last name	5 (21)
Time zone	Time zone in which device is located (eg, GMT+11)	5 (21)
Connection type	Cellular data or wi-fi	4 (17)
Medical conditions‡	Users' medical conditions (eg, diabetes, depression)	4 (17)
Birthday‡	User's date of birth	3 (13)
Device ID†‡	Unique 15 digit International Mobile Equipment Identity code of device	3 (13)
Sex	User's sex	3 (13)
Carrier	Mobile network operator, provider of network communications services (eg, AT&T)	2 (8)
Country	Country in which device is located (eg, Australia)	2 (8)
Coarse grain location‡	Non-precise location. Usually city in which device is located (eg, Sydney)	2 (8)
Drug instructions	Instructions related to user's drugs (eg, orally, with food)	2 (8)
Drug schedule	Times for drug administration (eg, 8 pm, in the morning)	2 (8)
Personal conditions‡	Users' personal conditions (eg, smoker, pregnant)	2 (8)
Personal factors‡	Includes user's anthropometric measurements or vital signs (eg, height, weight, blood pressure)	2 (8)
Symptoms‡	User's symptoms (eg, headache, nausea)	2 (8)
Doctor's name‡	Name of the user's doctor	1 (4)
Doses‡	Dose of user's drug (eg, 100 mg aspirin per day).	1 (4)
Feelings	User's current feelings (eg, happy, sad, anxious)	1 (4)
Pharmacy name‡	Information about user's favourite pharmacies (eg, name, location)	1 (4)

*Total number is 24; percentages do not add to 100% as apps could share multiple types of user data.
†Unique identifier.
‡May be considered personal data under the General Data Protection Rules—that is, "any information relating to an identified or identifiable natural person."[18]

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 55 entities identified in the traffic analysis into 46 "families" based on shared ownership, presuming that data as an asset was shared among acquiring, subsidiary, and affiliated companies as was explicitly stated in most privacy policies.[23] 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 figure 1 displays the results of the network analysis containing apps, and families of first 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 differentiated 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, first and third parties received a median of 3 (interquartile range 1-6, range 1-24) unique transmissions of user data, defined as sharing of a unique type of data (eg, Android ID, birthdate, location) with a first or third party. Amazon.com and Alphabet (the parent company of Google) received the highest volume of user data (both received n=24), followed by Microsoft (n=14). First and third parties received a median of 3 (interquartile range 1-5; range 1-18) different types of user data from the sampled apps. Amazon.com and Microsoft, two cloud service providers, received the greatest variety of user data (18 and 14 types, respectively), followed by the app developers Talking Medicines (n=10), Ada Health (n=9), and MedAdvisor International (n=8).

Fourth party sharing

Supplementary figure 2 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 to access and export data through linked accounts (eg, linking a third party analytics and advertising service); however, privacy policies typically stipulated that once data were sent to the integration partner, the data were subject to the partner's terms and conditions.

App developers typically engage third party companies to collect and analyse user data (derived from use of the app) for app analytics or advertising purposes. The privacy policies of third parties,

Table 3 | Data sharing practices of apps

No of installs* and apps	No of different types of user data shared†	No of unique transmissions (type/entity)‡	No of unique recipients§	No (%) of infrastructure recipients	No (%) of analysis recipients
500-1000:					
Dental Prescriber	0	0	0	0 (0)	0 (0)
Medsmart Meds & Pill Reminder App	14	25	4	1 (25)	3 (75)
myPharmacyLink	5	5	2	2 (100)	0 (0)
1000-5000:					
DrugDoses	0	0	0	0 (0)	0 (0)
MediTracker	4	6	3	1 (33)	2 (67)
MyMeds	5	8	3	1 (33)	2 (67)
5000-10 000:					
CredibleMeds	1	2	2	1 (50)	1 (50)
Med Helper Pro Pill Reminder	0	0	1	0 (0)	1 (100)
Nurse's Pocket Drug Guide 2015	0	0	3	0 (0)	3 (100)
Pedi Safe Medications	0	0	0	0 (0)	0 (0)
10 000-50 000:					
MIMS For Android	3	6	2	1 (50)	1 (50)
50 000-100 000:					
ListMeds-Free	0	0	0	0 (0)	0 (0)
MedicineWise	5	9	5	1 (20)	4 (80)
100 000-500 000:					
Dosecast-Medication Reminder	9	16	3	1 (33)	2 (67)
Lexicomp	3	6	3	1 (33)	2 (67)
MedAdvisor	8	20	3	2 (67)	1 (33)
My PillBox(Meds&Pill Reminder)	0	0	0	0 (0)	0 (0)
Nurse's Drug Handbook	4	9	5	2 (40)	3 (60)
Pill Identifier and Drug list	5	10	4	1 (25)	3 (75)
500 000-1 000 000:					
UpToDate for Android	5	11	3	0 (0)	3 (100)
1 000 000-5 000 000:					
Ada-Your Health Companion	15	27	13	5 (39)	8 (62)
Drugs.com	5	5	2	1 (50)	1 (50)
Epocrates Plus	8	14	3	1 (33)	2 (67)
5 000 000-10 000 000:					
Medscape	7	21	8	3 (38)	5 (63)

*As reported in Google Play store at time of sampling (November 2017).
†As detected in traffic analysis of 28 possible types.
‡As detected in traffic analysis and defined as sharing of unique type of data with an external entity—for example, app shares Device Name and OS Version with Crashlytics, resulting in two unique transmissions.
§Identified in traffic and privacy policy analysis.

however, define a relationship with the app developer and disclose how the developer's data (as a customer of the third party) will be treated. App users are informed that the collection and sharing of their data are defined by the developer's and not by the third party's privacy policy, and thus are referred to the app developer in the event of a privacy complaint.

Supplementary figure 2 displays the network including fourth parties. All the companies in the fourth party network receive user data for the purposes of analysis, including user behaviour analytics, error tracking, and advertising. We classified entities in the fourth party network by sector, based on their keywords in Crunchbase, to understand how health related app data might travel and to what end.

The fourth party network included 237 entities including 17 app families (apps, developers, and their parent companies in orange) (17/237, 7%), 18 third parties (18/237, 8%), and 216 fourth parties (216/237, 91%); 14 third parties were also identified as fourth parties (14/237, 6%) meaning that these third parties identified in the traffic analysis could also receive data from other third parties identified in the traffic analysis. Supplementary figure 2 shows that most third and fourth parties in the network (blue) could be broadly characterised as software and technology companies (120/220, 55%), whereas 33% (72/220) were explicitly digital advertising companies (grey), 8% (17/220) were owned by private equity and venture capital firms (yellow), 7 (3%) were major telecommunications corporations (dark grey), and 1 (1%) was a consumer credit reporting agency (purple). Only three entities could be characterised predominantly as belonging to the health sector (1%) (brown). Entities in the fourth party network potentially had access to a median of 3 (interquartile range 1-11, range 1-140) unique transmissions of user data from the sampled apps.

The fourth parties that are positioned in the network to receive the highest volume and most varied user data are multinational technology companies, including Alphabet, Facebook, and Oracle, and the data sharing partners of these companies (table 5). For example, Alphabet is the parent company of Google, which owns the third parties Crashlytics, Google Analytics, and AdMob By Google identified in our analysis. In its privacy policy, Google reports data

Table 4 | Categorisation of first and third parties (n=37) performing data analytics

Main activity of parties	No (%)	Description*	Examples	Example domain names†
First parties:				
Freelance app development	3 (8)	Design, develop, and maintain apps for third party clients to specification; services might include app usage analytics, ad campaign setup, and reporting; app store optimisation or customer support	Atmosphere Apps (USBMIS); Mobixed	secure.usbmis.com; www.mobixed.com
Clinical decision support	7 (19)	Ranging from not-for-profit companies to corporations, these companies provide evidence based drug information and clinical decision supports on digital platforms, including websites and apps; some are available through individual or institutional subscriptions; those that are free to users generate revenue through hosted advertising and sponsored content	Epocrates (AthenaHealth); Medscape (WebMD); UpToDate; Lexi-Comp; MIMS Australia; AZCERT	services.epocrates.com; api.medscape.com; www.uptodate.com; update.lexi.com; iris.mimsandroid.com.au; crediblemeds.org
Consumer health management	6 (16)	Consumer-facing apps that support drug adherence, health management, and care coordination; free for consumers, these companies generate revenue from pharmaceutical companies, health insurers, or health services by licensing the app (on a per member basis), sponsorship, or selling data commodities	Ada Health; MedAdvisor; Talking Medicines; MyMeds; Montuno Software; Precedence Health Care	prod-mh-22.ada.com; mobile.medadvisor.com.au; talkingmedicines.azurewebsites.net; app.my-meds.com; ppserver.montunosoftware.com; cdm.net.au
Third parties:				
Analytics	5 (14)	Freemium services; in exchange, companies retain the right to collect, aggregate, and commercialise deidentified end user data; companies provide services to app developers, including error and bug reporting, and analysis of user numbers, characteristics, and behaviours; some also offer the ability to understand users' behaviours across devices and platforms and integrate with advertising data to target marketing activities	Crashlytics; Sentry; Google Analytics; Flurry; Amplitude†	settings.crashlytics.com; ssl.google-analytics.com; data.flurry.com
User engagement	6 (16)	Freemium services; in exchange, companies retain the right to collect, aggregate, and commercialise de-identified end user data; these software integrations allow developers to analyse how users navigate an app, features users find most engaging and provide push notifications to increase user engagement	One Signal; Apptimize; Urban Airship; Braze; Mixpanel; Customer.io†	onesignal.com; brahe.apptimize.com; combine.urbanairship.com; dev.appboy.com; api.mixpanel.com
Advertising	7 (19)	Includes services that provide advertisement attribution to tie each user to the ads they interact with; buying and selling of ad space; ad serving and ad management; and analytics that enable ad targeting and personalisation	Audience Network by Facebook†; AdMob by Google†; TUNE; Adjust; 24/7 Real Media; JanRain; AppsFlyer	169316.engine.mobileapptracking.com; app.adjust.com; oasc17.247realmedia.com; nps.au.jainraincapture.com; t.appsflyer.com
Social media	1 (3)	Integration with social media platforms, allowing apps to share users' data with social media or to import social media data into the app; this could include a Facebook login, status updates related to the app, sharing content via social media, or finding a list of contacts who have also installed the app; this integration also allows for cross-platform advertising	Facebook Graph API	graph.facebook.com
Customer support	1 (3)	Paid services based on level of use; a software product that allows for tracking, prioritising, and solving user support issues including live chat and messaging and AI-powered help tools	Zendesk†	
Government	1 (3)	Several application programming interfaces are available through the National Library of Medicine related to public drug information sources	National Library of Medicine	rximage.nlm.nih.gov

*Description based on content analysis of entities' websites and linked documents such as privacy policies, terms and conditions, and investor prospectuses.
†When there was no corresponding domain name, the developers self reported data sharing with the third party in the app's privacy policy.

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."[24] These partners "can collect or receive non-personally identifiable information about your browser or device when you use Google sites and apps."[24] Table 6 exemplifies 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 of user data are unique and identifying, or potentially identifiable when aggregated. A few apps shared sensitive data such as a user's drug list and location that could potentially be transmitted among a mobile ecosystem of companies seeking to commercialise these data.

Strengths and limitations of this study

This traffic analysis was conducted at a single time point, performed on a small sample of popular apps, and is limited in terms of scalability. Thus the apps analysed might no longer be available, could have

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 different pieces of user data accessible
Alphabet	Technology	10	7	140
Facebook	Technology	4	1	50
Oracle	Technology	0	17	92
Vista Equity Partners	Private equity	0	14	87
Nielsen	Marketing	0	12	59
comScore	Marketing	0	11	58
Providence Equity Partners	Private equity	0	10	53
Kanta	Technology	0	10	53
RN SSI Group	Marketing	0	10	53
Segment	Marketing	0	6	53

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.[16] 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

Table 6 | Risks to privacy owing to data aggregation within fourth party network

User action in app	Data transmission		3rd party recipient	4th party profile (Alphabet)
	Category	Content		
Searches UpToDate for "rosacea"	Profile nickname	Joy	Crashlytics (owned by Alphabet)	Pseudonym: Joy (1234567890) Device*: Google Pixel running Android 7.1 "Nougat" Phone No*: +61 555 555 555 Last seen*: 31 January 2019 4.55 pm Apps used: UpToDate, Pill Identifier and Drug List, Medsmart Meds & Pill Reminder App, Starbucks, Glow Period Tracker, Uber, Runtastic, eHarmony, Facebook, Whatsapp, CommBank Mobile carrier*: Vodafone Australia City‡: Sydney Location*: Camperdown Sex: Female Age‡: 30-45 Drugs: Jurnista, Mobic, Topamax, Crestor, Lexapro Hobbies‡: coffee, running, dating Health conditions: fertility, chronic pain, joint pain, epilepsy, migraines, high cholesterol, depression
	Android ID (unique)	1234567890		
	Operating System	Android 7.1 "Nougat"		
	Device	Google Pixel 1		
Looks up patient's "pain" pill in Pill Identifier and Drug List	Browsing	Search "red", "round" tablet; browse Jurnista images†; browse hydromorphone controlled release uses	Google Analytics (owned by Alphabet)	
	Last seen	1 hour ago		
	Operating System	Android 7.1 "Nougat"		
	Device	Google Pixel		
Sets reminder for own prescriptions in Medsmart Meds & Pill Reminder App	Operating System	Android 7.1 "Nougat"	Mixpanel (integrates with Google BigQuery, owned by Alphabet)	
	Device	Google Pixel 1		
	Mobile carrier	Vodafone Australia		
	Connection type	WiFi		
	Drug list	Drug list§: meloxicam (Mobic) 15 mg capsule daily; topiramate (Topamax) 50 mg tablet twice daily; rosuvastatin (Crestor) 10 mg tablet daily; escitalopram (Lexapro) 10 mg tablet daily		

*Information collected by Google from "apps, browsers, and devices you use to access Google services"
†App user may search for or input brand or generic names; Jurnista is brand name for hydromorphone hydrochloride, used for treatment of moderate to severe pain.
‡Information inferred by Google on basis of aggregated data from third party sources including "apps that use Google advertising services," "your activity on other sites and apps," and "trusted partners, including marketing partners" per Google's privacy policy.
§App user in this profile was prescribed meloxicam (Mobic tablets for relief of migraine associated pain), topiramate (Topamax for treatment of migraine headaches), rosuvastatin (Crestor to lower high cholesterol), and escitalopram (Lexapro for treatment of depression).

analysis of 959 426 apps in the Google Play store found a median of five third party trackers were embedded in each app's source code and that these were linked to a small number of dominant parent companies, such as Alphabet.[4] Analyses of data collected using the Lumen app found that 60% of 1732 monitored apps shared user data with at least one domain associated with advertising or tracking, or both, and 20% shared with at least five different services.[3] The top domains were Crashlytics, a Google owned error reporting service that also provides app testing and user analytics, and Facebook Graph API that allows app users to connect with their Facebook account, but also provides analytic services and cross platform advertisement delivery.[3] A second analysis of the Lumen app dataset identified 2121 advertising or tracking services, or both receiving user data from 14 599 apps on 11 000 users and characterised these according to their parent organisations.[23] Despite owning just 4% of all third party tracking services identified, Alphabet had a presence in more than 73% of apps in the dataset; Facebook and Verizon Communications were similarly identified as having achieved monopoly positions within the mobile ecosystem.[23]

Conclusions and policy implications

The collection and commercialisation of app users' data continues to be a legitimate business practice. The lack of transparency, inadequate efforts to secure users' consent, and dominance of companies who use these data for the purposes of marketing, suggests that this practice is not for the benefit of the consumer.[10] Furthermore, the presence of trackers for advertising and analytics, uses additional data and processing time and could increase the app's vulnerability to security breaches.[25] In their defence, developers often claim that no "personally identifiable" information is collected or shared. However, the network positions of several companies who control the infrastructure in which apps are developed, as well as the data analytics and advertising services, means that users can be easily and uniquely identified, if not by name. For example, the semi-persistent Android ID will uniquely identify a user within the Google universe, which has considerable scope and ability to aggregate highly diverse information about the user. Taking a systems view of the mobile ecosystem suggests that privacy regulation should emphasise the accountabilities of third parties, known as "data processors," in addition to first parties or "data controllers."[18] Currently, within the "big data" industry, users do not own or control their personal data[10 11]; at minimum, regulators should insist on full transparency, requiring sharing as opposed to privacy policies. The implementation of the GDPR in the European Union resulted in greater transparency around data sharing relationships among some developers in our sample. However, as big data features increasingly in all aspects of our lives, privacy will become an important social determinant of health, and regulators should reconsider whether sharing user data for purposes unrelated to the use of a health app, for example, is indeed a legitimate business practice. At minimum, users should be able to choose precisely which types of data can be accessed and used by apps (eg, email, location), and to have the option to opt-out for each type of data. More effective regulation, however, might focus instead on third parties engaged in commercialising user data or the companies that own and operate the smartphone platforms and app stores.[4]

Conclusion

Clinicians should be conscious about the choices they make in relation to their app use and, when recommending apps to consumers, explain the potential for loss of personal privacy as part of informed consent. Privacy regulators should consider that loss of privacy is not a fair cost for the use of digital health services.

We thank Chris Klochek for developing the app store crawling program and Tanya Karliychuk, grants officer at the Australian Communications Consumer Action Network, for advising on the study design, analysis, and dissemination strategy.

Contributors: QG acquired funding, designed the study, supervised and participated in data collection and content analysis, and wrote the first draft of the manuscript. KC participated in data collection and content analysis and critically revised manuscript drafts. FH participated in designing the study, conducted the network analysis, and critically revised manuscript drafts. AC conducted the traffic analysis and critically revised manuscript drafts. LB participated in designing the study and commented on the draft. RH designed the study, supervised the traffic analysis, and critically revised manuscript drafts. QG attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. QG and RH act as guarantors.

Funding: This work was funded by a grant from the Sydney Policy Lab at The University of Sydney. QG was supported by a postdoctoral fellowship from the Canadian Institutes of Health Research. The Sydney Policy Lab 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: this work was funded by the Sydney Policy Lab; 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: The full analysis is publicly available at: https://healthprivacy.info/.

Transparency: The lead author (QG) 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 planned have been explained.

1 McGrath P, Blumer C, Carter J. *Medical appointment booking app HealthEngine sharing clients' personal information with lawyers.* ABC News. 2018 June 25.
2 research2guidance. mHealth developer economics: How mHealth publishers are monetizing their apps. Berlin, Germany: research2 guidance; 2018.
3 Vallina-Rodriguez N, Sundaresan S, Razaghpanah A, et al. Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem. 1st Data and Algorithm Transparency Workshop; New York, NY, 2016.
4 Binns R, Lyngs U, Van Kleek M, Zhao J, Libert T, Shadbolt N. Third party tracking in the mobile ecosystem. Proceedings of the 10th ACM Conference on Web Science. 2018. p.23-31.

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, Andrews LB. Privacy policies of android diabetes apps and sharing of health information. *JAMA* 2016;315:1051-2. doi:10.1001/jama.2015.19426
8 Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: The alarming state of practice. *IEEE Access* 2018;6:9390-403. doi:10.1109/ACCESS.2018.2799522.
9 Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. *J Am Med Inform Assoc* 2015;22(e1):e28-33.
10 Ebeling M. *Healthcare and big data: Digital specters and phantom objects*. Palgrave Macmillan US, 2016. doi:10.1057/978-1-137-50221-6
11 Pasquale F. *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015. doi:10.4159/harvard.9780674736061
12 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
13 Wilson D. Drug ap comes free, ads included. The New York Times. 2011 July 29.
14 Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. *J Am Pharm Assoc* 2013;53:172-81. doi:10.1331/JAPhA.2013.12202
15 Android Developers. *System permissions*. Mountain View, CA: Google, Inc; May 7, 2018, https://developer.android.com/guide/topics/security/permissions.html#normal-dangerous.
16 Continella A, Fratantonio Y, Lindorfer M, et al. Obfuscation-resilient privacy leak detection for mobile apps through differential analysis. Proceedings 2017 Network and Distributed System Security Symposium.
17 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-81. doi:10.1016/j.jbi.2008.08.010
18 *European Union. General Data Protection Regulation*. Offical Journal of the European Union, 2018.
19 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2016.
20 Pedersen T. tidygraph: A tidy API for graph manipulation. R package version 1.1.1 2018.
21 Grundy Q, Parker L, Raven M, et al. Finding peace of mind: Navigating the marketplace of mental health apps. Sydney, Australia; 2017.
22 Yahoo. Developer Network. Flurry analytics terms of service Sunnyvale, CA: Oath Media; May 7, 2018. https://developer.yahoo.com/flurry/legal-privacy/terms-service/flurry-analytics-terms-service.html.
23 Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, et al. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. Proceedings 2018 Network and Distributed System Security Symposium.
24 Google Inc. Who are Google's partners? Mountain View, CA: Google, Inc.; 2018 https://policies.google.com/privacy/google-partners?hl=en.
25 Müthing J, Brüngel R, Friedrich CM. Server-focused security assessment of mobile health apps for popular mobile platforms. *J Med Internet Res* 2019;21:e9818.

Supplementary information: Dummy profiles of app users

Supplementary table 2: Summary of apps' data sharing practices

Supplementary figure 1: Network of data sharing among apps, first, and third parties

Supplementary figure 2: Network of data sharing among app families, third, and fourth parties