Availability and quality of mobile health app privacy policies

RECEIVED 22 December 2013
REVISED 7 July 2014
ACCEPTED 3 August 2014
PUBLISHED ONLINE FIRST 21 August 2014

Ali Sunyaev[1], Tobias Dehling[1], Patrick L Taylor[2], Kenneth D Mandl[3]

ABSTRACT

Mobile health (mHealth) customers shopping for applications (apps) should be aware of app privacy practices so they can make informed decisions about purchase and use. We sought to assess the availability, scope, and transparency of mHealth app privacy policies on iOS and Android. Over 35 000 mHealth apps are available for iOS and Android. Of the 600 most commonly used apps, only 183 (30.5%) had privacy policies. Average policy length was 1755 (SD 1301) words with a reading grade level of 16 (SD 2.9). Two thirds (66.1%) of privacy policies did not specifically address the app itself. Our findings show that currently mHealth developers often fail to provide app privacy policies. The privacy policies that are available do not make information privacy practices transparent to users, require college-level literacy, and are often not focused on the app itself. Further research is warranted to address why privacy policies are often absent, opaque, or irrelevant, and to find a remedy.

BRIEF COMMUNICATION

INTRODUCTION

Apple's iOS and Google's Android operating systems and associated application (app) stores, itunes.apple.com and play.google.com, are becoming the de facto global platforms for mobile health (mHealth).[1,2] Recently, both platforms additionally announced the roll out of their own apps fostering app interoperability and offering central storage for all mHealth apps and sensors of users' devices.[3,4] mHealth apps leverage a wide range of embedded technology in iOS and Android devices for collecting and storing personal data, including contacts and calendars, and patient-reported data as well as information collected with cameras and sensors, including location, acceleration, audio, or orientation.[5–7] Although patients value control of their personally identifiable data[8,9] and the Federal Trade Commission[10] recommends provision of privacy policies for mobile apps, little attention has been paid to the information security and privacy policies and practices of mHealth app vendors. Although both app stores retain the right to remove apps for infringements of privacy, neither has explicit policies addressing the information security and privacy of medical information. Users choose among an ecosystem of substitutable mHealth apps[11] and should have transparency as to which apps have privacy practices best aligned with their individual preferences. We sought to assess mHealth apps for the presence and scope of privacy policies, and what information they offer.

METHODS

We surveyed (figure 1) the most frequently rated and thus popular English language mHealth apps in the Apple iTunes Store and the Google Play Store. App stores organize their offerings in categories (eg, Books, Games, and News). We selected apps from the Medical and Health and Fitness categories offered in both stores in May 2013. The iOS app store lists all apps by category and offers the desired information in plain hypertext markup language (HTML), enabling us to automatically parse app information to extract data. On the other hand, the Android app store uses dynamically generated HTML pages so that the HTML texts displayed in the browser do not contain much useful information, which is dynamically loaded from an underlying database. Hence, we used a third-party open-source interface, the android-market-api (http://code.google.com/p/android-market-api), for retrieving app information.

Upon initial review, many apps were not available in English, did not have an English description, or were not health-related, despite being offered in the categories Medical or Health and Fitness (eg, apps offering wallpapers). In order to exclude such apps from further assessment, we tagged all app descriptions with descriptive terms. The tags characterize health-related app functionality, access to information, and handling of information. We manually tagged 200 apps (100 Health and Fitness, 100 Medical) establishing an initial tag corpus and employed string matching[12] to automatically tag the remaining apps. Apps not matched by at least four distinct tags were excluded from further assessment.

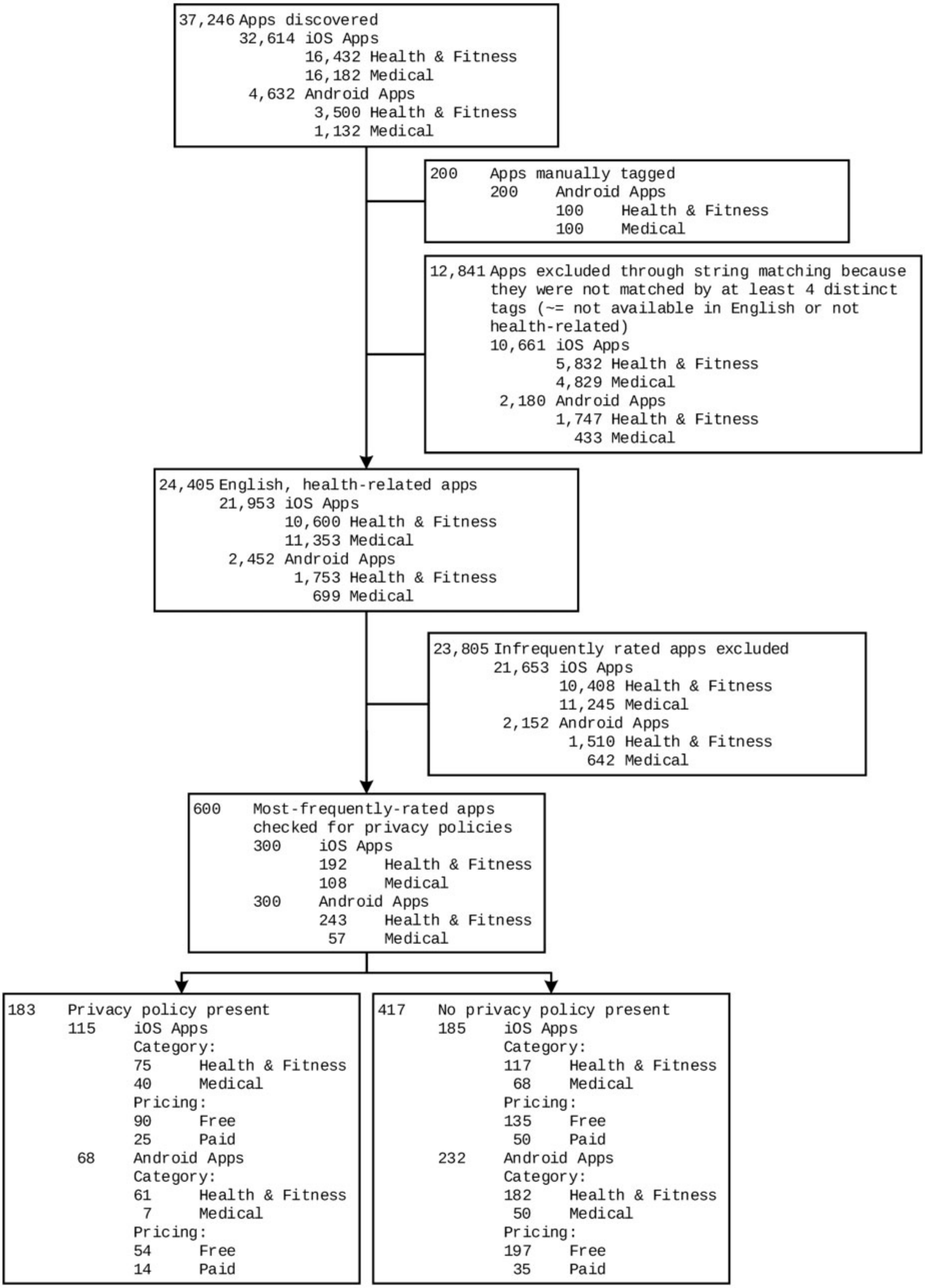
Discovery and evaluation of privacy policies

We used a three-step manual procedure for privacy policy discovery looking at typical locations for privacy policies. Privacy policies were abstracted from March 2013 to June 2013. First, we checked for a privacy policy on the app store web site for the particular app. Then we checked the web page maintained

Correspondence to Professor Ali Sunyaev, Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, Cologne 50923, Germany; sunyaev@wiso.uni-koeln.de

For numbered affiliations see end of article.

Figure 1: Flow diagram for app discovery and processing.



by the developer to advertise and introduce the company and its products. Finally, we reviewed the first 30 results of a Google search for the query '$APPNAME "privacy" "policy"'. Once a privacy policy was discovered, we omitted the remaining steps.

We surveyed the 300 most frequently rated apps in our sample for privacy policies in the iOS as well as the Android app store. We were interested in the most commonly used apps, a property best reflected by download count. However, since only Android (and not iOS) reports download count, we instead selected apps for privacy policy assessment based on their rating count. For Android apps, rating count and download count are strongly positively correlated (Spearman $r=0.89$, $p<0.001$), indicating that rating count is a good proxy for download count.

To identify differences in the availability of privacy policies, we used independence of proportions with the Pearson χ^2 test. Grade-level readability was calculated as the average of the Flesh Kincaid, Gunning Fog, and SMOG formulas.[13,14] Length was assessed as the number of words in the privacy policy. Two-sample Student t tests were used to compare privacy policy lengths. Privacy policy scope could be limited to the single app in question, apply to multiple apps, or pertain to a backend application supporting the app(s), other products and services offered by a developer, or seemingly unrelated topics. To assess the transparency of privacy policies focusing on apps or backend applications, we determined whether the privacy policies address: type of information collected (operational, behavioral, sensitive), rationale for collection (app operation, personalization, secondary use), sharing of information (service provision, social interaction, third party), and user controls (supervision, notification, correction).[15–17] Privacy policies rationalizing collection of personal information on the basis of 'personalization' indicated tailoring of app functionality based on collected user information. Similarly, privacy policies were categorized as addressing collection of 'sensitive' information if they referenced street address, finances, ideological orientation, location, government identifiers, or state of health. Privacy policies enabling users to supervise information-privacy-related aspects were assessed as addressing user controls regarding 'supervision'; this includes informing users about the limits of the privacy policy, about which app modules collect what information, or whether users are provided with access audits for shared information. Two researchers evaluated privacy policies along two axes—privacy policy scope and offered content. Reliability assessment with Janson's and Olsson's ι, a multivariate extension of Cohen's κ for multiple judges on the same scale,[18] led to an 'almost perfect'[19] agreement score of $\iota=0.94$. In the end, all differences were resolved through group discussion.

RESULTS

Initial search identified 32 614 mHealth apps in the iOS and 4632 mHealth apps in the Android app store. Tagging reduced the number of discovered apps to 21 953 iOS apps and 2452 Android apps that are available in English and offer some health-related functionality (figure 1).

Availability of privacy policies

Only 30.5% of apps had privacy policies. iOS apps were more likely to have privacy policies (38.3% vs 22.7%, χ^2 $p<0.001$; see figure 1). The χ^2 test revealed no influence of app category or app pricing on the availability of privacy policies. Correlation of privacy policy availability and app rating count is weak (iOS: Spearman $r=0.22$, $p<0.001$; Android: Spearman $r=0.31$, $p<0.001$).

Privacy policy characteristics

Privacy policies have an average length of 1755 (SD 1301) words and range from 65 to 6424 words and from 17 to 5333 words on iOS and Android, respectively. Android privacy policies are shorter (Student t, $p<0.001$) with an average length of 1353 (SD 1018) words in contrast to 1991 (SD 1393) words. Privacy policies have an average reading grade level (RGL) of 16 (SD 2.9) and two discovered privacy policies have an RGL below the recommended eighth grade level.[13,14] Privacy policy length and RGL have a weak positive correlation (Spearman $r=0.31$, $p<0.001$).

Table 1 shows the scope of the privacy policies. The six different scope categories are mutually exclusive and were determined according to the scope of obtained privacy policies. Aside from initial differences in naming, privacy policy scope assessments were unanimous. The findings showed that 66.1% of discovered privacy policies do not focus on the app, but a developer homepage, all services offered by a developer, or topics unrelated to the app.

We assessed the transparency of privacy policies that focus on a backend application, multiple apps, or a single app (table 2). Some aspects of each privacy policy content category most important to users[15–17] are addressed in over 85% of assessed privacy policies. All assessed privacy policies indicate whether information is shared with third parties. Whether sensitive information is collected is addressed in 74.2% of assessed privacy policies. Secondary use of information is

Table 1: Privacy policy scope for iOS and Android apps

Store	iOS, N (%)	Android, N (%)
Privacy policy scope		
Single app	4 (3.5)	10 (14.7)
Multiple apps	6 (5.2)	9 (13.2)
Backend application	21 (18.3)	12 (17.6)
Developer homepage	15 (13.0)	5 (7.4)
All developer services	55 (47.8)	27 (39.7)
No app-related scope	14 (12.2)	5 (7.4)

Table 2: Single, multiple, and backend application privacy policies addressing content categories important to users

Privacy policy content categories	Privacy policies, N (%)	Privacy policy content subcategories	Privacy policies, N (%)
Type of information collected	56 (90.3)	Operational	54 (87.1)
		Behavioral	56 (90.3)
		Sensitive	46 (74.2)
Rationale for collection	59 (95.2)	App operation	41 (66.1)
		Personalization	58 (93.5)
		Secondary use	48 (77.4)
Sharing of information	62 (100.0)	Service provision	57 (91.9)
		Social interaction	34 (54.8)
		Third party	62 (100.0)
User controls	54 (87.1)	Supervision	49 (79.0)
		Notification	37 (59.7)
		Correction	32 (51.6)

addressed in 77.4% of assessed privacy policies. Information regarding supervision of information access and use is offered in 79% of assessed privacy policies. Means for notifying users about changes to privacy policies or privacy practices are mentioned in 59.7% of assessed privacy policies.

DISCUSSION

Information privacy[20] is a highly charged concept, very subject to personal feelings, and its correct protection in the context of a purchase-sale bargain, a trade-off between sought-for personal benefits and real as well as hypothetical costs, is an open question heightened by great legal and cultural uncertainty, and lack of an organized industry policy. Privacy policies are often present as detached, legalistic documents that seem to be potentially fungible or borrowed from someone else because they are mostly incomprehensible, out-of-scope, and lacking transparency. There are no general international standards for the information a privacy policy should offer, for uses and disclosures it should permit, whether with consent or without it, or for the rights consent can waive. Public policies that do govern private information include the California Online Privacy Protection Act of 2003[21] which requires provision of privacy policies for all online services accessible by Californian residents, and the Federal Trade Commission encourages app developers to provide privacy policies as well as just-in-time disclosures requesting consent for information collection.[10] Extant guidance and regulation regarding privacy policies are, however, abstract and limited in scope, while corresponding IT offerings provide diverse functionality and are globally available.

In the domain of health information where many consumers are concerned about what happens to their private, sensitive data, our key finding is startling: apps are being highly rated and successfully sold although privacy policies are either absent, opaque, or irrelevant. There are several possible explanations, ranging from consumers' confidence in the general legal climate to protect them even in the absence of or despite app privacy policies, over consumers falling for the privacy paradox[20] and choosing short term benefits despite potential exposure to harm in the long term, complete misunderstanding of the extent to which such apps may compromise personal privacy, to an absence of real choice, which would be assisted by clear 'gold standards' against which consumers could compare app policies.

We assessed the privacy policies of the 300 most frequently rated apps in the iOS and the Android app store. Still, our results show that privacy policies have poor availability rates, correlation of app ratings and privacy policy availability is weak, privacy policy scope is lacking, high RGLs are required to understand privacy policies, and privacy practices are not made transparent in a comprehensive fashion. Although depending on our association of ratings with number of downloads, these results indicate that app developers seem to be competing without benefitting from protection against clear harm of failing to address information privacy or from availability and quality of privacy policies, which one might expect to be reflected in customer choice.

Many privacy policies did not focus on the app at all, and therefore were not informative for end users. On the one hand, consumers may be blissfully ignorant and more likely to use apps with unclear or difficult to find privacy policies. On the other hand, concerns about information privacy may inhibit physicians'[22] and patients'[23] information sharing, even for patients who are willing to share for altruistic purposes.[24]

An agreed upon community standard of not collecting personal data which is not necessary for the app's central function would go a long way toward eliminating issues. And the privacy policies should reflect use of best technical practices for designing privacy protection into mobile applications. Preventing undesirable breaches of privacy will be much more cost-efficient than remedying unwanted disclosures of private health information.

For information that does need to be collected and stored for future reference by the app, complete transparency about subsequent disclosures or sales in a standardized format, at the sixth grade reading level, should be expected. Because an overwhelming amount of text is unlikely to be read by users,[25] a bulleted, graphical, or tabular executive summary should be provided.

Assuming that privacy policies do fill an important niche in legal protection and consumer confidence, their relative absence points to an imperfection in the market, and deserves further research on the substantive ways the market fails and on whether failure is self-correcting or would benefit from a step that places collaboration above competition, such as creation of quality standards, self-regulation, or government regulation.

ACKNOWLEDGEMENTS

Computing resources were provided by the Regional Computing Center of the University of Cologne.

CONTRIBUTORS

KDM and AS conceived the project. Data acquisition and analysis were conducted by TD and AS. TD and AS performed the statistical analyses and implemented required custom software. All authors wrote the manuscript, and were responsible for the research concept and design as well as critical revision of the manuscript, and approved the final version.

FUNDING

This project was supported by Strategic Health IT Advanced Research Projects Award 90TR000101 from the Office of the National Coordinator of Health Information Technology.

COMPETING INTERESTS

None.

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

DATA SHARING

All data used for the analyses are available from AS or TD upon request.

REFERENCES

1. D'Heureuse N, Huici F, Arumaithurai M, *et al*. What's app?: a wide-scale measurement study of smart phone markets. *ACM SIGMOBILE Mob Comput Commun Rev*. 2012;16:16–27.
2. Istepanian RSH, Jovanov E, Zhang YT. Guest editorial introduction to the special section on m-health: beyond seamless mobility and global wireless health-care connectivity. *IEEE Trans Inf Technol Biomed*. 2004;8:405–414.
3. Apple. Health. 2014. http://www.apple.com/ios/ios8/health. Archived at: http://www.webcitation.org/6QtK0lqTv
4. Google. The Google Fit SDK. 2014. https://developers.google.com/fit. Archived at: http://www.webcitation.org/6QtJkTpQE
5. Lane ND, Miluzzo E, Lu H, *et al*. A survey of mobile phone sensing. *IEEE Commun Mag*. 2010;48:140–150.
6. Weiss GM, Lockhart JW. The impact of personalization on smartphone-based activity recognition. In: Proceedings of the Activity Context Representation Workshop. Toronto, Canada: 2012.
7. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? *JAMA*. 2013;310:2395–2396.
8. Pyper C, Amery J, Watson M, *et al*. Access to electronic health records in primary care—a survey of patients' views. *Med Sci Monit*. 2004;10:SR17–22.
9. Simon SR, Evans JS, Benjamin A, *et al*. Patients' attitudes toward electronic health information exchange: qualitative study. *J Med Internet Res*. 2009;11:e30.
10. Federal Trade Commission. Mobile privacy disclosures: building trust through transparency. Federal Trade Commission 2013. http://www.ftc.gov/os/2013/02/130201mobileprivacyreport.pdf (Accessed 9 Apr 2014).
11. Mandl KD, Kohane IS. No small change for the health information economy. *N Engl J Med*. 2009;360:1278–1281.
12. Faro S, Lecroq T. Twenty years of bit-parallelism in string matching. In: Holub J, Watson BW, Žďárek J, eds. *Festschrift for Bořivoj Melichar*. Prague, Czech Republic: Prague Stringology Club, 2012:72–101.
13. Ley P, Florio T. The use of readability formulas in health care. *Psychol Health Med*. 1996;1:7–28.
14. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. *Respir Care*. 2008;53:1310–1315.
15. Ackerman MS, Cranor LF, Reagle J. Privacy in e-Commerce: examining user scenarios and privacy preferences. In: Proceedings of the 1st ACM Conference on Electronic Commerce. Denver, CO, USA: ACM, 1999:1–8.
16. Antón AI, Earp JB, Young JD. How internet users' privacy concerns have evolved since 2002. *IEEE Secur Priv*. 2010;8:21–27.
17. Earp JB, Antón AI, Aiman-Smith L, *et al*. Examining internet privacy policies within the context of user privacy values. *IEEE Trans Eng Manag*. 2005;52:227–237.
18. Janson H, Olsson U. A measure of agreement for interval or nominal multivariate observations. *Educ Psychol Meas*. 2001;61:277–289.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

20. Smith HJ, Dinev T, Xu H. Information privacy research: an interdisciplinary review. *MIS Q*. 2011;35:989–1016.
21. California Online Privacy Protection Act of 2003. Bus. Prof. Code Sect. 22575–22579. 2004. http://www.leginfo.ca.gov/cgi-bin/displaycode?section=bpc&group=22001-23000&file=22575-22579. Archived at: http://www.webcitation.org/6Rv9DRvtC
22. Dünnebeil S, Sunyaev A, Blohm I, *et al*. Determinants of physicians' technology acceptance for e-health in ambulatory care. *Int J Med Inf*. 2012;81:746–760.
23. Agaku IT, Adisa AO, Ayo-Yusuf OA, *et al*. Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. *J Am Med Inform Assoc*. 2014;21:374–378.
24. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. *J Med Internet Res*. 2010;12:e14.
25. McDonald AM, Cranor LF. The cost of reading privacy policies. *J Law Policy Inf Soc*. 2008;4:540–565.

AUTHOR AFFILIATIONS

[1]Faculty of Management, Economics and Social Sciences, University of Cologne, Cologne, Germany

[2]Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, Harvard Law School, Boston, Massachusetts, USA

[3]Children's Hospital Informatics Program, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA