What do users think of the UK's three COVID-19 contact-tracing apps? A comparative analysis

Vahid Garousi [iD] ,[1,2] David Cutting[1]

ABSTRACT

Objectives Our goal was to gain insights into the user reviews of the three COVID-19 contact-tracing mobile apps, developed for the different regions of the UK: 'NHS COVID-19' for England and Wales, 'StopCOVID NI' for Northern Ireland and 'Protect Scotland' for Scotland. Our two research questions are (1) what are the users' experience and satisfaction levels with the three apps? and (2) what are the main issues (problems) that users have reported about the apps?

Methods We assess the popularity of the apps and end users' perceptions based on user reviews in app stores. We conduct three types of analysis (data mining, sentiment analysis and topic modelling) to derive insights from the combined set of 25 583 user reviews of the aforementioned three apps (submitted by users until the end of 2020).

Results Results show that end users have been generally dissatisfied with the apps under study, except the Scottish app. Some of the major issues that users have reported are high battery drainage and doubts on whether apps are really working.

Discussion Towards the end of 2020, the much-awaited COVID-19 vaccines started to be available, but still, analysing the users' feedback and technical issues of these apps, in retrospective, is valuable to learn the right lessons to be ready for similar circumstances in future.

Conclusion Our results show that more work is needed by the stakeholders behind the apps (eg, apps' software engineering teams, public-health experts and decision makers) to improve the software quality and, as a result, the public adoption of these apps. For example, they should be designed to be as simple as possible to operate (need for usability).

[1]School of Electronics, Electrical Engineering and Computer Science (EEECS), Queen's University Belfast, Belfast, UK
[2]–, Bahar Software Engineering Consulting, Belfast, UK

Correspondence to
Dr Vahid Garousi; vgarousi@gmail.com

Summary

What is already known?
► As of January 2021, more than 64 countries and regions have developed contact-tracing apps to limit the spread of COVID-19.
► Many articles in the academic literature and also the media have questioned the public adoption and quality of the COVID-19 contact-tracing apps.
► By the end of 2020, more than 25 583 user reviews were submitted for the UK's three COVID-19 contact-tracing mobile apps: 'NHS COVID-19' for England and Wales, 'StopCOVID NI' for Northern Ireland and 'Protect Scotland' for Scotland.

What does this paper add?
► In this paper, we derive empirical insights from the user reviews of the aforementioned three COVID-19 contact-tracing apps.
► Our two research questions are (1) what are the users' satisfaction levels with the three apps? and (2) what are the main issues (problems) that users have reported about the apps?

INTRODUCTION

As of January 2021, more than 64 countries and regions have developed contact-tracing apps to limit the spread of COVID-19.[1]

These apps use Bluetooth signals to log when smartphone owners (users) are close to each other; so if a user later tests positive for COVID-19, an alert can be sent to the other users that have recently been in close contact. Figure 1 shows several screenshots from an example app, depicting the typical features of these apps.

In the UK, three different apps have been developed and publicised by the regional and national governments for different constituent countries: the NHS COVID-19 app (for England and Wales), the Stop-COVID NI app for Northern Ireland and the Protect Scotland app for Scotland. Reports[1] indicate that the total cost of the NHS COVID-19 app alone is expected to top £35 million.

The apps have been promoted as a promising tool to help bring the COVID-19 outbreak under control. However, there are many articles in the academic literature[2] and also the media questioning the efficacy and public adoption of these apps. For example, a systematic review[3] of 15 studies in this area found that 'there is relatively limited evidence for the impact of contact-tracing apps'. A news search for 'what went wrong with UK contact-tracing apps' would return a few hundred hits.
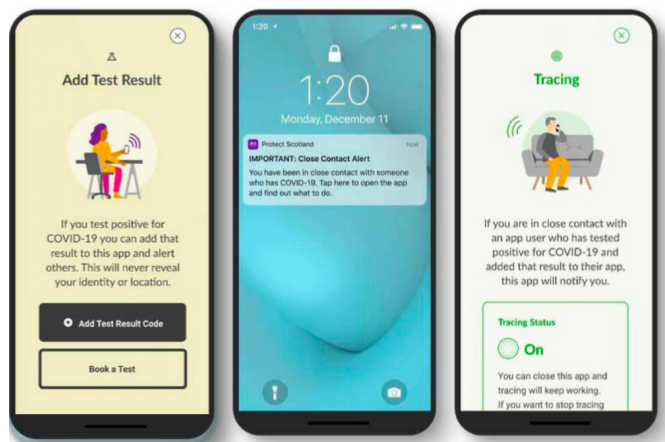
Figure 1 Screenshots of the Protect Scotland app's user interface method and data collection.

One cannot help but wonder about the reasons behind low adoption of the apps by the general public in the UK and many other countries. The issue is multi-faceted, complex and interdisciplinary, as it relates to fields such as public health, behavioural science,[4] epidemiology, information technology (IT) and software engineering. Since these apps are essentially software systems, we investigate them from a 'software in society' lens[5] in this paper. Software in society means the role of position of software systems (eg, mobile apps) in society, as they are used by billions of people in the society. Software systems should be of high quality and should be usable/useful for end users who are mostly non-technical people.

The software aspects of these apps are also quite diverse in themselves, for example, whether the app software would work as intended (eg, will it send the alerts to all the recent contacts?) and whether different apps developed by different countries will cooperate/integrate (when people travel between counties). A related news article reported that a large number of developers worldwide has reported a large number of defects in the NHS app (bit.ly/FlawsInNHSApp).

An interesting source of knowledge about the user experiences is through the availability of a large number of user reviews in the two major app stores: the Google Play Store for Android apps and the Apple App Store for iOS apps. A review often contains information about a user's experience with the app and opinion of it, feature requests or bug reports.[6] Many insights can be mined by analysing the user reviews of these apps to figure out what end users think of contact-tracing apps, and that is what we analyse and present in this paper.

The nature of our analysis is 'exploratory',[7] as we want to extract insights from the app reviews which could be useful for different stakeholders, for example, app developers, public-health experts, decision makers and the public. The two research questions (RQs) that we explore are (1) what are the users' satisfaction levels with the three UK apps? and (2) what are the main

issues (problems) that users have reported about the apps? While some studies[8] have shown that there may be some inherent negative bias in public app reviews (in app stores), many researchers[6] and practitioners are widely using app reviews to derive improvement recommendations on the apps.

User feedback has long been an important approach for understanding the success or failure of software systems, traditionally in the form of direct feedback or focus groups and more recently through social media (eg, tweets about a given app in Tweeter) and reviews submitted in app stores.[9] A systematic literature review[6] of the approaches used to mine user opinion from app store reviews identified a number of approaches used to analyse such reviews and some interesting findings such as the correlation between app rating and downloads.

Several related papers, similar to this work, have been published, for example, a recent paper[10] focused on sentiment analysis of user reviews of the Irish app. Another recent paper[11] analysed the user reviews of apps of a set of 16 countries (UK was not included). The paper presented thematic findings on what went wrong with the apps, for example, lack of citizen involvement, lack of understanding of the technological context of users and ambitious technical assumptions without cultural considerations.

As another related work, we have published online a recent comprehensive technical report[12] by analysing the review data of nine European apps from (1) England and Wales, (2) Scotland, (3) Northern Ireland, (4) Ireland, (5) Germany, (6) Switzerland, (7) France, (8) Finland and (9) Austria. In this current paper, our goal was to go in depth and focus on the three UK apps.

In addition to the academic (peer-reviewed) literature, in the grey literature (such as news articles and technical reports), there are plenty of articles on the software engineering aspects of contact-tracing apps. For example, an interesting related news article was entitled 'UK contact-tracing app launch shows flawed understanding of software development' (www.verdict.co.uk/contact-tracing-app-launch/). The article argued that 'In a pandemic, speed is critical. When it comes to developing high-quality software at speed, using open-source is essential, which other nations were quick to recognize. The article also criticised the approach taken by the UK healthcare authorities in developing their app from scratch: 'Countries such as Ireland, Germany, and Italy used open-source to build [develop] their own applications months ago. Sadly the UK did not follow suit, and wasted millions of pounds and hours of resources trying to build its own version'.

Another motivating factor for this study is the consulting engagement of the first author in relation to the NI's StopCOVID NI app, in the summer of 2020. Some of his activities included peer review and inspection of various software engineering artefacts of the app, for example, design diagrams, test plans and test suites; see page 13 of an online report by the NI's Health

and Social Care authority (https://covid-19.hscni.net/wp-content/uploads/2020/07/Expleo-StopCOVIDNI-Closure-Report-V1.0.pdf) In the project, a need was identified to review and mine insights from user reviews in order to be able to make improvements in the app.

In the rest of this paper, we first review our method and data collection approach, and then present the results of our analysis. We then conclude the paper with discussions and conclusions.

METHOD AND DATA COLLECTION

To retrieve the review data of the three apps, we used a commercial app–analytics tool named AppBot (https://appbot.co/). This is a widely used tool and, according to its website, is in use by companies such as Microsoft, Twitter, BMW, LinkedIn and the New York Times.

We conducted the data collection on 26 December 2020. Therefore, all the reviews data of the apps until that date were included in our dataset. We provide all the data extracted and analysed for this paper in an online repository (http://www.doi.org/10.5281/zenodo.4059087). We also think that some readers may be interested to explore the dataset and reviews by their own and possibly conduct further studies like ours. To help with those, we have recorded and provide a brief (10 min) video of live interaction with the dataset (to be analysed in this paper) using AppBot, which can be found online (youtube/qXZ_8ZTr8cc).

Table 1 lists the names, some key information, and descriptive statistics of both platform versions of the regional apps. Each app had received somewhere between 127 and 17 905 reviews, as of the data collection date.

The number of versions since first release are quite different. The NHS app had 16 updates, while the NI and Scotland apps each had 6 updates only. This could have a variety of root causes, for example, the NHS app team is more responsive to feedbacks and thus have updated it more often, or the app had more issues and thus had to be fixed more frequently.

Given the different scale of downloads for the three apps (table 1), we wondered about their correlations
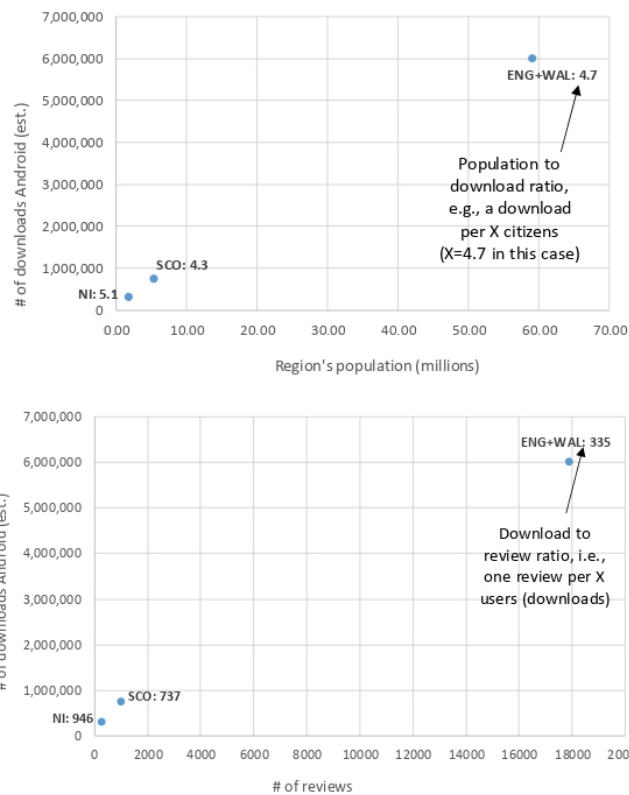

Figure 2 Region populations versus number of downloads (estimated) and reviews results. (Short terms: ENG: England; WAL: Wales; NI: Northern Ireland; SCO: Scotland)

with regions' population sizes. We visualised each region's population versus the number of downloads and reviews as XY plots in figure 2. We observed reasonable correlations between each pair of the metrics, that is, for a region with a larger population, as one would expect, there were more downloads and more reviews.

We show two scatter plots in figure 2 depicting the population to download, and download to reviews ratios for the three apps, that is, a download per 4.7 citizens on average for the NHS app. This metric is 4.3 and 5.1, respectively, for the Scotland and NI apps. One could analyse such slight differences based on the level of healthcare authorities' aggressiveness in publicity and/or social fabric of each region, but we do not investigate

Table 1	The three apps and their descriptive statistics					Reviews (so far)	
App	**Operating System (OS)**	**First release date**	**Downloads (n) (est.)**	**Versions since first release**		**Reviews (n)**	**Average stars**
NHS COVID-19	Android	13 August 2020	12 500 000+ (www.bbc.co.uk/news/technology-54326267)	16		17 905	2.8
	iOS					6021	2.7
Protect Scotland	Android	10 September 2020	1 270 000+ (prod.news.stv.tv/scotland/more-than-a-million-download-scotlands-tracing-app)	6		1018	3.8
	iOS					195	3.2
Stop COVID NI	Android	28 July 2020	375 000+ (play.google.com/store/apps/details?id=net.hscni.covidtracker)	6		317	2.8
	iOS					127	2.3

OS, operating system.

Figure 3 Distribution of review stars and review sentiment categories.

them further as we want to instead focus on technical (IT) aspects in this work.

RESULTS

We present the results in the next two subsections.

Users' experience and their satisfaction with the apps

Our first exploratory analysis is to assess the ratios of users who, as per their reviews, have been happy or unhappy with the apps.

To gauge satisfaction with an app, the built-in rubric of app stores is 'stars', a rating feature also used in many other online systems, such as Amazon. A user can choose between one and five stars, when shehe submits a review as well as optionally provide text. Another more sophisticated way to derive users' satisfaction with an app is to look at the positive/negative 'sentiment' score of their textual reviews. Sentiment analysis[13] refers to the use of natural language processing (NLP) to systematically quantify the affective state of a given text. Our chosen tool (AppBot) derives four possible types of sentiments for a given review text: positive, negative, neutral and mixed sentiments. 'Neutral' reviews lack strong sentiment, for example, 'I have used this app'. 'Mixed' reviews have conflicting sentiments (both positive and negative).

We show in figure 3 the distribution of reviews' sentiment categories and also the distribution of stars in the dataset. We show both a 100% stacked bar and a stacked bar of absolute values for the stars.

We can see from these charts and also the average stars of each app (table 1) that users are generally

dissatisfied with the apps under study, except the Scottish app. Based on the average stars metric, the Scottish app is the highest starred (3.8 and 3.2 out of 5.0). The NHS and NI apps have received 2.75 and 2.55 out of 5.0, respectively, on average.

One would wonder about the factors that have led to the Scottish app being ranked the highest in terms of stars. Reviewing a subset of its reviews reveals that the app is quite effective and easy to use; for example, one user said, 'Brilliant app. It collects zero personal data, no sign ups, no requirement to turn on location, nothing! All you have to do is turn on Bluetooth, that's it' (http://bit.ly/ScoAppPosReview).

We would expect that stars and the reviews' sentiments would be correlated, that is, if a user leaves one star for an app, she/he would most probably leave a negative (critical) review. We show in figure 3D a scatter plot of those two metrics, in which six data points correspond to both versions of the three apps under study. The Pearson correlation coefficient of the two measures is 0.97, thus showing a strong correlation.

Main issues that users have reported about the apps

As per table 1, the three apps have in total more than 25 000 reviews, and thus, manual analysis of such a large and diverse textual feedback was not an option. The AppBot tool provides features such as word clouds and 'topic modelling' to make sense of review texts, which we show in figure 4. Topic modelling is an NLP-based statistical semantic technique for discovering the abstract 'topics' that occur in a given textural dataset.
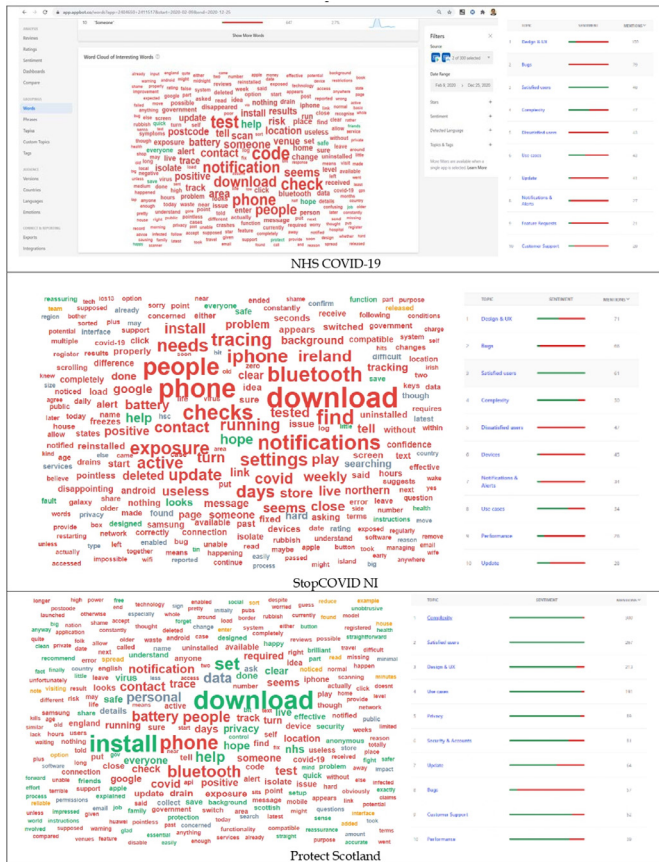
Figure 4 Word cloud, sentiment analysis and topic modelling of the app reviews.

We also include in figure 4 the AppBot tool's user interface as a glimpse into how it works.

Words in the word clouds of figure 4 are coloured according to the sentiment of the review in which they are contained: green for positive sentiments, grey for neutral, yellow for mixed and red for negative sentiments.

In the topic models of figure 4, we can see that certain issues have varying degrees of sentiment (positive or negative) in different apps. Topics are ordered by the number of 'mentions' (occurrence) in reviews. For example, with the NHS app, 'design & UX' was widely discussed in negative sentiments by the users. A valuable feature of AppBot is that it groups the topics under similar groupings, for example, topics such as 'bugs' and design & UX occur in almost all three topic models.

We review next a subset of the common problems reported for all three apps and then some of the issues reported for each.

Common problems reported for all three apps

One major issue reported by users is the lack of 'interoperability' between the apps; that is, if a user from one nation of the UK visits another, the app will not record the contact IDs in the new region, and in case of entering a positive COVID-19 result, the app

Lesson learnt/recommendations

There seems to be rather trivial usability issues with some of the apps (eg, the case of exposure notification errors). This raises the question of the inadequate usability testing of the apps and possibility of releasing them on a 'rush'.

will not notify those contacts. This issue has been reported in a large number of reviews, for example,

- ► 'Complete and utter waste of space. Only works if I come into contact with someone else using the same backstreet application, who has managed to get tested without being turned away, and inputs a code into their app. If I bump into someone from England, Wales, Ireland, or anywhere else for that matter with COVID-19 then this app does diddly squat - What's the point??' (https://app.appbot.co/apps/2437310-protect-scotland/reviews/1957885880/or directly in Google Play store: https://play.google.com/store/apps/details?id=gov.scot.covidtracker&hl=en_GB&reviewId=gp%3AAOqpTOHVDLw1vCASIapQQmiyem1xyhXCw4SQBcOPdRXy0v1YPz95_hcZ5CUz7kWFe8004v5TbARTEjYyuEJoGw)

Also a number of users, understandably, compared the features of the three apps, and complained about one of them not having the feature provided by another UK-based app, for example, a review of the Scottish app was

- ► 'Looks great, easy to use but oh how they missed out some useful features such as a [NHS] COVID-19 [app's] alert state notifier, scanning business QR Codes, etc. so user's data needn't be handed over in pubs, etc'. (https://app.appbot.co/apps/2437310-protect-scotland/reviews/1961498492)

One of the frequent words with negative sentiment for NHS and NI apps was 'notifications', which appeared in more than 2000 negative reviews, for example,

- ► 'Want to get people to uninstall it? Don't produce audible notifications you haven't been exposed this week at 6am on a Fri morning, waking people up'. (https://appbot.co/apps/2392818-stopcovid-ni/reviews/1950428641)
- ► 'I am getting a warning that exposure notifications may not work for the area I am in. As this is Northern Ireland I am unclear why it is saying this. The exposure log does not appear to have made any checks since early August. This does not give confidence that the app is working properly. I do hope the designers are reading these reviews as this appears to be a recurring issue'. (https://appbot.co/apps/2392818-stopcovid-ni/reviews/1950428665)

NHS COVID-19 app

As visualised in figure 4, one of the frequent words with negative sentiment for this app was 'code', which has appeared in 2604 of the 13803 (18.8%) negative reviews. This phrase referred either to a QR code which

Lesson learnt/recommendations

A large number of users reported issues related to mobile device 'fragmentation', that is, the app being incompatible with older phone models. To maximise the installation coverage of the apps, it was important to make the apps compatible with as many devices as possible.

is used in the app to identify physical locations, for example, restaurants, or the code (ID) of users who want to enter their positive/negative test results in the app. There were a great number of criticisms, and the following are some examples:

► *"Well, as a business we are directed to register for track and trace. Having registered for a QR code and subsequently printed said code. I thought, in good naval tradition, 'Lets give it a test before we put the poster up'. So download the app from Play Store. Scanned the code and a message pops up 'There is no app that can use this code'. Next move, open the application. What do we find!! Currently only for NHS Volunteer Responders, Isle of White and Newham residents'. What is the point of publicizing this if it does not have basic functionality? Measure twice cut once... Also there should be an option for no Star as it appropriate for this application!"* (https://appbot.co/apps/2411517-nhs-covid-19/reviews/1951839432) → poor alignment of publicity timing.

► *"QR location doesn't seem to work for me. Used a standard QR reader on my phone and it took me straight to venue but the QR reader in the app said QR code not recognized."* (https://appbot.co/apps/2411517-nhs-covid-19/reviews/1994520039) → poor testing of that software module.

► 'I move about a lot with my job. I can't change the post code to the area I'm in unless I uninstall the app' (https://appbot.co/apps/2411517-nhs-covid-19/reviews/1993233064) → the need for better software requirements engineering.

StopCOVID NI app

Many users reported having problems installing the app. Many of those complaints were about the incompatibility of the apps with certain (mostly older) phone models. For example, one review was 'Tried to download on an elderly relative's Samsung phone but the app isn't compatible. Nowhere can I find a list of compatible devices or Android versions. Sadly the app won't help the most vulnerable' (https://appbot.co/apps/2436851-stopcovid-ni/reviews/1950429102).

Lesson learnt/recommendations

Reviews reveal that not enough testing has been done on all possible types of QR codes.

Protect Scotland app

One of the common words with negative sentiments for the app is 'people', which appeared in 99 of the 1249 negative reviews, for example,

► 'Think about this... What's the point unless 100% of people have this app? I could be in a supermarket with 100 people. One person has COVID-19 in said Supermarket, but is the only one who does not have the app. That person inflects several people, but they won't know where they caught it – because that one person didn't have the app'. (https://appbot.co/apps/2437310-protect-scotland/reviews/1951729752): the user stresses the need for wide adoption of the app, which is a valid issue.

Just like for other apps, there were also multiple reviews about high battery usage and other issues related to when the phone's Bluetooth is on, for example,

► 'This app requires Bluetooth to be permanently on. A real battery killer, I also get bombarded by nearby devices that see my Bluetooth is on. Google location is accurate to around 1m, why is this not enough? Uninstalled until something better comes up'. (https://appbot.co/apps/2437310-protect-scotland/reviews/1951722760)

DISCUSSONS

We believe that our work in this paper makes useful contributions to the literature on this topic by presenting a comparative analysis what users think of the UK's three COVID-19 contact-tracing apps.

Our results provide various lesson learnt, recommendations and implications to different stakeholders of the apps (eg, software developers of the apps, and public-health experts managing the develop projects): (1) the end users are generally dissatisfied with the apps under study, except the Scottish app; this issue is perhaps the most clear and the most important message of our study, which should be investigated by stakeholders; (2) future studies could look into what factors has made the Scottish app different from others in the pool of apps under study; and that could be an RQ to be studied by researchers in future works; and (3) contact-tracing apps should be designed to be as simple as possible to operate (for usability), as we cannot expect layperson citizens to review the online frequently asked questions pages of the app to properly configure it, especially for a safety-critical health-related app.

CONCLUSIONS

The initial exploratory analysis of COVID-19 contact-tracing app reviews reported in this paper is only a starting point. As the COVID-19 pandemic has paralysed most of the life and businesses around the globe, contact-tracing apps, if developed and managed properly, may have the potential to help bring the COVID-19 outbreak under control. It is vital for governments and health authorities, including those in the UK, to develop and offer effective apps that all citizens can use.

Mining user reviews of contact-tracing apps seem like a useful analysis towards providing insights to various stakeholders, for example, app developers, public-health experts, decision makers and the public. Of course, such an analysis and software engineering aspects can only provide some pieces of the 'big picture'. Therefore, collaborations with other important disciplines, including public health and behavioural science,[4] shall be continued.

ORCID iD
Vahid Garousi http://orcid.org/0000-0001-6590-7576

REFERENCES

1 Rahman M. "The countries using Google and Apple's COVID-19 contact tracing API,". Available: https://www.xda-developers.com/google-apple-covid-19-contact-tracing-exposure-notifications-api-app-list-countries/ [Accessed Jan 2021].
2 Martuscelli C, Heikkilä M. "Scientists cast doubt on effectiveness of coronavirus contact-tracing apps,". Available: https://www.politico.eu/article/scientists-cast-doubt-on-the-effectiveness-of-contact-tracing-apps/ [Accessed Oct 2020].
3 Braithwaite I, Callender T, Bullock M, *et al*. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *Lancet Digit Health* 2020;2:e607–21.
4 The British Psychological Society. "*Behavioural science and success of the proposed UK digital contact tracing application for COVID-19*" *Technical report*, 2020.
5 Kazman R, Pasquale L. Software engineering in society. *IEEE Softw* 2019;37:7–9.
6 Genc-Nayebi N, Abran A. A systematic literature review: opinion mining studies from mobile app store user reviews. *J Syst Softw* 2017;125:207–19.
7 Runeson P, Höst M. Guidelines for conducting and reporting case study research in software engineering. *Empir Softw Eng* 2009;14:131–64.
8 Yin D, Mitra S, Zhang H. "Mechanisms of negativity bias: an empirical exploration of app reviews in Apple's app store". In: *International Conference on information systems*, 2012.
9 Morales-Ramirez I, Perini A, Guizzardi RSS. An ontology of online user feedback in software engineering. *Appl Ontol* 2015;10:297–330.
10 Rekanar K. "Sentiment analysis of user feedback on the HSE contact tracing app", 2020. Available: https://www.researchsquare.com/article/rs-96174/v1
11 Bano M, Zowghi D, Arora C. Requirements, politics, or individualism: what drives the success of COVID-19 Contact-tracing apps? *IEEE Softw* 2021;38:7–12.
12 Garousi V, Cutting D, Felderer M. "*Mining user reviews of COVID contact-tracing apps: an exploratory analysis of nine European apps*", 2020.
13 Liu B. Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 2012;5:1–167.