Data governance for the public acceptability of personalized COVID-19 advice: An experimental study in Hong Kong

Veronica Li[1], Masaru Yarime[1]*

[1]Division of Public Policy, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

**e-mail: yarime@ust.hk*

Abstract

As COVID-19 persists and mutates, governments will need to keep citizens updated with the latest information. During this time of high uncertainty, taking a personalised approach to COVID-19 advice could prove valuable for citizens to protect themselves based on their individual circumstances. Although efforts have been made to develop technologies that could make this approach viable, there is a lack of research focusing on the socio-political barriers that could lead to low public acceptance. Here, we present a survey experiment where we gauged the willingness of Hong Kong citizens to use a mobile application for personalised COVID-19 advice based on different data governance concerns and demographic characteristics, such as the sector of the developer or the method of data storage. We conclude that gender has a statistically significant effect on willingness, possibly due to women having greater concerns over the safety risks of sharing personal data than men. We also note that other concerns surrounding data security and access affect users' willingness to use a personalised advice application where they would need to share health and location data. Finally, we encourage further research on context-specific factors affecting the public acceptability of data tools for crisis management.

Keywords **–** COVID-19, personal data, data governance, citizen engagement, health advice

1 Introduction

More than one year after the World Health Organization's declaration of a pandemic, COVID-19 remains the most urgent crisis in the world. As of August 20, 2021, the disease has claimed more than 200 million confirmed cases and 4.3 million deaths (World Health Organization, 2021). At this time of crisis, it is imperative to keep citizens informed so that they could make prudent decisions to protect themselves and their communities.

Many data tools have been constructed to disseminate information on COVID-19 at a macroscopic level. Some of the most prolific tools include dashboards and thematic websites dedicated to tracking local COVID-19 cases and sharing anti-epidemic knowledge, respectively. This knowledge includes information on common symptoms, vaccination details, and the geographical distribution of local cases. While these websites are useful for understanding COVID-19 at a larger scale, they fail to help citizens to gauge the personal risk of COVID-19.

This paper seeks to explore how personal data could be utilised to generate tailored COVID-19 advice. In this study, we attempted an experiment to examine this issue in the case of Hong Kong. Although Hong Kong has mostly managed to prevent community infections for the time being (Department of Health, 2021), unlike other megacities where daily COVID-19 cases could reach the hundreds or even thousands (City of New York, 2021; Greater London Authority, 2021; Tokyo Metropolitan Government, 2021), the metropolis has struggled to eliminate the coronavirus. In fact, Hong Kong has already experienced four waves of COVID-19 cases within the past year (OT&P Healthcare, 2021). Given the rise of more infectious strains of the coronavirus emerging in various countries (Chow, 2021; Higgins, 2021; Reuters Staff, 2021), as well as the constant risk of new imported cases, the Hong Kong government must remain vigilant in protecting its citizens from COVID-19. It would be worthwhile to gauge the willingness of Hong Kong citizens to use a mobile application that could tailor COVID-19 advice to individual needs.

2 Literature Review

Helping individuals to understand the risks of COVID-19 in the contexts of their own lives is imperative in allowing them to make the best decisions to protect themselves and

their loved ones from the illness. Information provision tools could be designed to tailor advice to unique situations based on data on users' health conditions, household characteristics, frequently visited locations and more. This approach is often referred to as "personalised medicine" or "precision medicine" (Abrahams, 2008; Khoury et al., 2016), and it is often applied to long-term health concerns or conditions such as nutrition (Franco et al., 2018) or cancer survivor support (Lubberding et al., 2015). Less literature centres around precision public health for public health crises such as COVID-19.

Although this approach is more adaptable than a website or dashboard, it also requires far more sensitive data, raising issues of data privacy and security. Furthermore, a lack of public trust in the government could render such an intervention unpopular regardless of privacy protections, due to fears of government surveillance. This was evident when many Hong Kong citizens refused to use the LeaveHomeSafe contact tracing application for COVID-19 (Chau, 2021), even though the Hong Kong government has stated in the application's privacy policy that minimal personal data would be collected (OGCIO, 2020). As with all data tools dependent on the use of personal data, it is crucial to mediate the issues of functionality, trust, and data privacy for a personalised data tool to be feasible (Hand, 2018; Kostkova et al., 2016).

A few papers have emerged discussing the logistics of such tools; Prodhan & Fenton (2020) devised a Bayesian model to calculate an individual's probability of COVID-19 infection based on their risk factors, and Freeman et al. (2021) identified several design considerations to improve the communication of personalised COVID-19 risk. The Human Nature Lab at Yale University has even developed a functional mobile application to demonstrate the personalised advice approach; it refers to its users' demographic data, health data, and social networks to assess their personal risk of contracting COVID-19 or other respiratory illnesses (Belli, 2020). None of these papers address the social or political factors, such as stakeholder needs and public trust, that could affect whether an application for personalized advice would be publicly accepted. This paper aims to fill this research gap.

3 Methodology

In this study, we seek to identify the socio-political and data-related factors that would make a mobile application for personalised COVID-19 advice feasible and effective in Hong Kong. By drawing insights on citizens' concerns in the specific context of Hong Kong, we aim to identify generalisable policy strategies to increase public acceptability of personalised health advice apps. Other objectives are to discover how to effectively synthesise various types of data (e.g., environmental data, COVID-19 case data, personal data) for generating public knowledge, to identify potential users' data concerns and address them in the application's design, and to empower citizens to become active agents of knowledge creation and protect themselves during crises.

We conducted an anonymous online survey experiment to gauge the willingness of Hong Kong citizens to use a mobile application for personalised COVID-19 advice. Each participant was presented images of one of twelve variations of a mock-up design for the application. The participants were also instructed to pay special attention to the mock privacy policy of the application, which explicitly mentions the extent of demanded personal data, the data storage method, and the sector of the app developer as the independent variables. An example of such a privacy policy is displayed in Fig. 1.

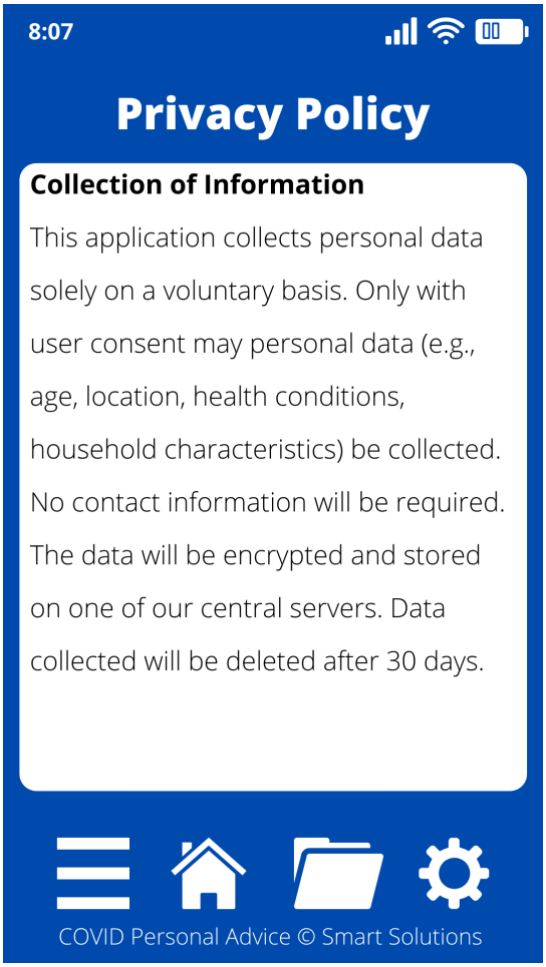

Fig. 1. Image of a mock "privacy policy" for a hypothetical mobile application for personalised COVID-19 advice.

The participants were then prompted to indicate how willing they would be to use the proposed application on a 5-point Likert scale, with 1 representing "completely willing" and 5 representing "completely unwilling". Finally, participants were given a text box to provide detailed feedback on the designs.

An ordered logistic regression (OLR) model was applied to the collected survey data to determine how the participants' willingness to use the application changed with its features. The OLS model took the form:

$$W = \alpha PersonalData + \beta Storage + \gamma Developer + \kappa \mathrm{X} + \epsilon$$

with W as willingness, α, β, γ, and κ as coefficients, X as the personal characteristics of the survey respondents (i.e., gender, age, education), and ϵ as an error term. In this study, we used the collection of contact information as a proxy for the collection of personal data. Using the model, we calculated the odds ratios (ORs) and the marginal effects (MEs) of the independent and control variables relative to the dependent variable of willingness. Since a lower integer is associated to a higher level of willingness, this means an OR less than 1 and negative MEs would be associated with an increase in willingness to use the application.

We used the Qualtrics panel service to randomly select about 800 participants aged 18 and above who were currently residing in Hong Kong. Qualtrics distributed our survey to other survey-taking companies on our behalf to reach the participants. We controlled the sample population to be representative of the Hong Kong population in terms of age, gender, and education (Census and Statistics Department, 2015; Census and Statistics Department, 2019). These three demographic characteristics were also the control variables for the OLR model. Data was collected between May 21 and June 3, 2021, and the data was processed in Stata.

4 Results

A total of 820 valid responses were collected between May 21 and June 3, 2021. Compared to the Hong Kong population, the sample had a slight overrepresentation of men and a significant underrepresentation of individuals aged 55 and over. The reason for the latter discrepancy could be due to the survey being in English. This could have made the study inaccessible to many of the older members of the survey panel, as most of the participants aged 55 and above had only completed their secondary education or lower.

The overwhelmingly popular response from the participants was that they would be somewhat willing to use the proposed application. This was the case across all independent and control variables, although "very willing" was the most frequent response among participants aged 25-34. However, based on the OLR model, gender is the only statistically significant variable relative to willingness with an odds ratio of 1.302 (see Table 1), suggesting that male respondents are likelier than female respondents to accept the proposed application.

Table 1. ORs of Independent and Control Variables Relative to Willingness to Use the Hypothetical Mobile Application

Variable	**Odds Ratios**	**95% Confidence Intervals**	
Developer sector	0.997	0.856	1.161
Demanded personal data	0.851	0.663	1.092
Data storage method	0.833	0.649	1.069
Age group	1.062	0.971	1.162
Gender	1.302**	1.014	1.672
Highest level of education	0.983	0.848	1.139
*p < .1, **p < .05, ***p < .01			

Other variables with p-values of about 0.2 or lower are the extent of demanded personal data, the storage method, and age. For the personal data variable, the OR and p-value of 0.833 and 0.206 respectively suggest a potential preference, though slight, for applications demanding contact information from the users (see Fig. 2). However, according to the participants' comments, some participants were deterred by the collection of contact information and other personal data. It could be worthwhile to continue studying these factors in more detail to uncover whether they are indeed important to the entire population.

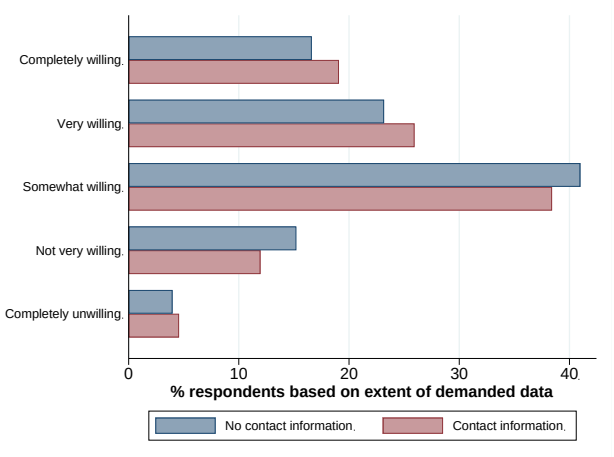

Fig. 2. Willingness to use a hypothetical mobile application for personalised COVID-19 advice based on the extent of collected contact information.

The participants' detailed comments provided further insights into what potential users may expect to be included or excluded from the application. For one, some participants wanted the application to be more comprehensive, such as by including information on COVID-19 as an illness, testing, and vaccines. For another, 31 participants commented on their concerns regarding data privacy and security, some of which were not very willing or completely unwilling to use the application due to these concerns. Two female participants, both of which indicated a low level of willingness to use the application, specifically mentioned that they would not want their location data to be collected. A handful of participants also commented that they would be uncomfortable with their data being shared with the government, the public, or "unrelated parties", suggesting that access to data should be restricted to increase the appeal of the application. The full list of factors identified in the comments can be found in Table 2.

Table 2. Factors Affecting Willingness to Use Mock-up Design

Factor	**Number of comments on the factor**
design	49
data	31
privacy	18
simplicity	17
information	16
functions	15
user-friendliness	13
language	8
clarity	5
government	5
safety	4
advice	3
creativity	3
other	3
entertainment	2
choice	2
cost	2
usefulness	2
motivation	1
trust	1
accessibility	1

5 Discussion

The findings of the study were unexpected, as they suggest that, aside from gender, the independent and control variables have little to no effect on the willingness of participants to use the proposed mobile application for personalised COVID-19 advice. The fact that women were generally less willing to use the application may suggest that women in Hong Kong are more hesitant than men to share their data. These results appear to corroborate conclusions from past studies, which state that women tend to exhibit higher levels of concern than men (Kehoe et al., 1997; Park, 2015). However, we do not yet fully understand why this is the case. One hypothesis is that women could associate higher levels of risk with certain types of data, such as location data, that could compromise their safety should it be leaked. Further research should be conducted to explore the potential connection between gender and data sharing, so that data policies could account for gender differences and address underlying issues.

As for the slight preference for versions of the app that collect contact information, this could possibly be due to the privacy policy of the application detailing why and how the data is being collected and used. The high level of transparency regarding data use may have reassured the participants. The preference may also be due to an expectation of increased effectiveness in generating helpful personalised advice based on the provision of contact information. However, this was not universal across the responses, as indicated by some of the participants' comments stating that they would prefer not to share personally identifiable information. The lack of statistical significance may also imply that there may not be a dependence of willingness on the extent of collected data. Rather, other data concerns could be of equal or higher priority, such as how secure the data protection system is or whether the data would be shared with unwanted parties. It may be worthwhile to refine the OLR model based on the participants' feedback.

It is also important to emphasise that collective preferences do not always align with individual preferences. Although most of the independent and control variables did not have statistically significant effects on the sample's overall willingness to use the proposed application, there were still individual statements highlighting how these variables would have an impact, as stated above. Addressing all these concerns will be difficult if not impossible, especially since there were conflicting comments with opposite preferences for the design and functionality of the application. However, if the most pressing issues of data privacy, data management, and design could be addressed, there is potential for an application like the one proposed to be created and used with widespread public acceptance.

Finally, we acknowledge several limitations of our study. For one, participants may not have paid close attention to the data management details of the privacy policies even if they were instructed to do so, leading to a seemingly indifferent response to variables they may in fact find important. For another, using the word "design" in the survey may have led to participants overemphasising the visual elements of the application, although the original intention of using the word "design" was to explain that participants would only be shown one variant of the mock application. Lastly, creating a Traditional Chinese version of the survey would make it more accessible for older

participants, allowing them to gain a better understanding of the study and share more relevant comments. Similar studies should take these limitations into account in the future.

6 Conclusion

As demonstrated in this paper, it is crucial to keep citizens updated during rapidly evolving crisis situations such as COVID-19 so that they could make informed decisions and protect themselves from harm. As COVID-19 persists, communities will need help adapting to new waves and mutations, and individuals will benefit from specific advice catered to their own needs. A personalised COVID-19 application would help, but its effectiveness depends on the amount of personal data it receives from its users. Therefore, such an application must be designed carefully to balance its functionality with societal concerns over data privacy and use.

In this study, we discovered that Hong Kong citizens were generally indifferent to the data storage method, the sector of the application developer, and whether contact information would be collected by the application, although individual responses may sometimes indicate otherwise. Instead, aesthetic appeal, data security, simplicity, comprehensiveness, and restrictions to access to personal data were among some of the key considerations of the citizens in deciding whether they were willing to use a mobile application for personalised COVID-19 advice. Moreover, we identified gender as a key factor in the acceptance of data tools that utilise personal data, with women being less willing to use the proposed application. We urge researchers to conduct similar studies in other policy contexts, as well as comparative analyses across different jurisdictions, so that more general conclusions could be drawn regarding the universal issue of using personal data for crisis management.

References

World Health Organization. (2021). *WHO Coronavirus Disease (COVID-19) Dashboard.* World Health Organization.

Department of Health. (2021). *Data in Coronavirus Disease (COVID-19)* DATA.GOV.HK.

City of New York. (2021). *COVID-19: Data.* NYC Health.

Greater London Authority. (2021). *Coronavirus (COVID-19) Cases and Vaccinations.* London Datastore.

Tokyo Metropolitan Government. (2021). *COVID-19 The Information website.* COVID-19 The Information website.

OT&P Healthcare. (2021). *COVID-19 Timeline of Events.* OT&P Healthcare.

Chow, D. (2021, Feb 4). Why a particular coronavirus mutation has scientists' attention. *NBC News.*

Higgins, T. (2021, Feb 7). UK coronavirus strain is doubling in the U.S. every 10 days, study finds. *CNBC.*

Reuters Staff. (2021, Feb 19). Japan finds new COVID-19 strain, while immigration centre reports infections. *Reuters*.

Abrahams, E. (2008). Right Drug—Right Patient—Right Time: Personalized Medicine Coalition. *Clinical and Translational Science, 1*(1), 11-12.

Khoury, M. J., Iademarco, M. F., & Riley, W. T. (2016). Precision Public Health for the Era of Precision Medicine. *American Journal of Preventive Medicine, 50*(3), 398-401.

Franco, R. Z., Fallaize, R., Hwang, F., & Lovegrove, J. A. (2018). Strategies for online personalised nutrition advice employed in the development of the eNutri web app. *Proceedings of the Nutrition Society, 78*(3), 407-417.

Lubberding, S., Uden-Kraan, C. F. v., Velde, E. A. T., Cuijpers, P., Leemans, C. R., & Leeuw, I. M. V. (2015). Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. *Journal of Clinical Nursing, 24*(9-10), 1367-1379.

Chau, C. (2021, Feb 19). Covid-19: Some Hongkongers shun gov't tracking app over privacy concerns as new rules rolled out at eateries. *Hong Kong Free Press*.

OGCIO. (2020). *Privacy Policy.* LeaveHomeSafe.

Hand, D. J. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data, 6*(3), 176-190.

Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R. A., Ross, E., Sasse, A., Sullivan, R., Chaytor, S., Stevenson, O., Velho, R., & Tooke, J. (2016). Who Owns the Data? Open Data for Healthcare. *Frontiers in Public Health, 4,* 7.

Prodhan, G., & Fenton, N. (2020). Extending the range of COVID-19 risk factors in a Bayesian network model for personalised risk assessment. *medRxiv*, 2020.10.20.20215814.

Freeman, A. L. J., Kerr, J., Recchia, G., Schneider, C. R., Lawrence, A. C. E., Finikarides, L., Luoni, G., Dryhurst, S., & Spiegelhalter, D. J. (2021). Communicating personalised risks from COVID-19: guidelines from an empirical study. *medRxiv*, 2020.10.05.20206961.

Belli, B. (2020, June 5). Yale app Hunala aims to be 'Waze for coronavirus'. *YaleNews*.

Census and Statistics Department. (2015). *Educational Characteristics of Hong Kong Population.* 2016 Population By-census.

Census and Statistics Department. (2019). *Population Estimates: Table 002: Population by Age Group and Sex.* Census and Statistics Department.

Kehoe, C., Pitkow, J. & Morton, K. (1997). *GVU's 8th WWW User Survey*. GVU's WWW User Surveys.

Park, Y. J. (2015). Do men and women differ in privacy? Gendered privacy and (in)equality in the Internet. *Computers in Human Behavior, 50*, 252-258.