Tensions and Mitigations: Understanding Concerns and Values around Smartphone Data Collection for Public Health Emergencies

Colin Watson, Open Lab, School of Computing, Newcastle University

Ridita Ali, Open Lab, School of Computing, Newcastle University

Jan David Smeddinck, Open Lab, School of Computing, Newcastle University

Smartphones increasingly serve as the source for, or to aggregate, a considerable amount of data that can be relevant in public health emergencies. Hence the sharing and utilisation of mobile health data, for example to help control the spread of communicable diseases, has become a relevant issue, with the COVID-19 pandemic adding a sudden urgency mirrored in debates around contact tracing apps. Building on exploratory work that indicated user perceptions and values around consent, and the notion that smartphones and mobile health data can be perceived as elements of self-embodiment, we present an online study comparing three scenarios of representative diseases undertaken during the first wave lockdown in the UK. Using a mixed-methods analysis of responses from 86 participants, we identify tensions and mitigations in user values and from those present the description of four characteristic user-groups that can inform considerations for design and development activities in this space.

1 INTRODUCTION

The growing interest in mobile health and wellbeing tracking has increased broader awareness and debate around the related collection, sharing and use of personal data. Smartphones with their native capacities and app extensions are central to this expansion of capabilities and services – generating, storing and facilitating data use directly from their own sensors, or mediating data from other sources such as wearable and stationary sensing devices. Human-computer interaction research has responded by undertaking studies exploring the potentials and pitfalls [11,45], including matters relating to collaborative efforts such as citizen science, as well as more generally efficacy, privacy and security [54,59,73,80,81].

Due to the COVID-19 virus situation, the use of mobile technology to help combat the pandemic has become a society-wide disease-control priority with digital technologies being developed in an urgent manner for purposes such as proximity tracing/tracking [37,41,48,76,95], symptom tracking [37,41], treatment management [6,99], lockdown/quarantine enforcement [16,95] and immunity certification [71]. However, community support and public engagement are required to achieve widespread population adoption of such technology. These underline the relevance of consent and data use considerations, as well as public debate about privacy, data ownership, and balancing personal versus societal benefits and disbenefits [26,32,39,70]. There is a timely relevance to understanding people's attitudes towards data access and consent for public health uses to inform the design of digital technologies that are mindful and sensitive to concerns. Accordingly, we discuss the research questions:

RQ1: How do people react to consented/unconsented mobile phone (or phone mediated) data collection for public health purposes?

RQ2: How do the reactions/values depend on scenarios with differing "public health emergency relevance" (subsequently referred to as *severity* where communicability, infection rates and mortality are relevant factors)?

RQ3: In what way do traits/positions around general values of privacy and civic engagement influence the reactions/values?

Building on a survey-based data collection with 86 participants, our analysis follows a three-stage mixed methods approach, initially undertaking separately a thematic qualitative and a quantitative analyses, before using results from the latter to reframe understandings of the qualitative data. The quantitative elements of the analysis follow the directional hypotheses that:

H1: Increasing *severity* correlates to increasing willingness / obligation to share personal data.

H2: Increasing *severity* – possibly pandemic – scenarios will lead to significantly increased acceptance that personal data could be taken from mobile phones and used contrary to the wishes of individuals.

H3: The greater the concerns people have about privacy are, the less willing they are to share personal data.

H4: The more civically engaged people are, the more willing they are to share personal data.

Our work was intended to mirror and then expand upon a prior small-scale proof-of-concept focus group study undertaken before the COVID-19 pandemic by Watson and Smeddinck [93]. That study indicated individuals may perceive smartphones metaphorically akin – and hence likely tied to psychophysiological manifestations as embodiment implies [31,82] – to organs or limbs. It also indicated they consider the benefits to society of unconsented access to personal data on mobile devices must be overwhelming, before such acquisition is considered acceptable. The analysis also pointed to the difficulty of gaining consent, a lack of knowledge about legal aspects, and a distrust about the state collecting data.

For consistency with the previous study, the work reported here employed three scenarios also based around increasing *severity* of public health emergency (from noncommunicable heart palpitations with no public health emergency impact, through moderately contagious meningitis, to an influenza pandemic). Given the timing of the study two months into the UK's initial lockdown, questions relating to the three scenarios were augmented by, rather than replaced by, supplementary questions seeking the 86 participants' views on data access and consent related to the COVID-19 virus pandemic situation. These were to capture additional different nuances beyond the influenza pandemic scenario.

Our study indicates concerns and considerations people have when they specifically think about sharing data from their smartphones to participate in public health emergency response activities, based on their experiences and thoughts. It highlights tensions between privacy, lawfulness, consent, mitigations, and sense of duty. These findings provide additional understanding on the nature of concerns about data access and consent, and identify implications for the design and application processes of public health technologies for supporting citizen-collaborative data collection and responses to the control of highly communicable diseases.

2   RELATED WORK

2.1 Privacy contextual factors

A better understanding of people's privacy norms and expectations can help support the development of appropriate and better-accepted mobile applications. Nissenbaum's "contextual integrity" [61] links adequate privacy protection for information gathering and dissemination of surveillance data with the norms of specific contexts, in terms of appropriateness and flow. Further, the social context provides a model for understanding what is appropriate for technologies that track, analyse and disseminate personal data [62]. Lin et al. [49] identified the importance of people's mental models of mobile privacy and that revealing misconceptions about privacy-related behaviours assists people to make better app-related trust decisions. It has also been noted [88] that sharing of personal data in co-present interactions (e.g.: a face-to-face encounter in a hospital) can be actively managed by the individual, but control features are used more when there is distributed interaction (e.g.: data sharing with a remote organisation). Martin and Shilton [53] have examined how two dominant theoretical models for privacy, individual privacy attitudes and situational/contextual privacy expectations, interact proposing that "experience provides the bridge between general privacy attitudes and nuanced contextual factors in privacy judgments". Martin and Shilton's study used a survey method based on context-specific scenarios to investigate the privacy expectations of mobile app users in the USA, and showed the need to consider both contextual and individual-level attitudes, concerns and beliefs in future research. Mobile device data privacy mental models and expectations in the context of public health emergencies have not been studied in the same way.

There are trade-offs between *perceived usefulness of data* and privacy concerns [42]; during health events with greater severity, wider societal usefulness will also be an influence. Our study explores this in the context of scenarios with increasing public health emergency relevance. Lin et al. [49] found that mobile users felt more comfortable when they knew in advance why data was being used, and that uncertainties about why data like location information was being used, considerably undermined trust. Provision of information in advance to reduce concern is particularly important in the context of public health emergencies [49] where there is an imperative for prompt collaboration between citizens and state authorities in order to respond in a timely manner [33].

2.2 Mobile health data collection

Digital health technologies use data and communications to gather, share and process information to improve the delivery of health care and patients' health, but they also entail concerns about robustness, data quality, patient safety, ease of use and accessibility, privacy and security and regulatory aspects [28,54,80,96]. Increased prevalence of smartphones combined with improvements in sensors, connectivity and cloud computing, has led to a quickly growing number of smartphone-based devices and applications for healthcare monitoring and management [52,91]. Apart from use for healthcare and personal wellbeing, these can also be important tools for public health systems to support "surveillance, alerting, situational awareness, emergency planning and response, resource assessment and management" preparedness [33] and operations during emerging epidemics or food contamination outbreaks [19,91] building on participatory sensing concepts [38].

Data sharing is routine but not transparent to mobile health and wellbeing app users [36]. Patients' perceptions of security and privacy can be related to their experiences with technologies, their own health, whether they are receiving treatment [66], varying sensitivities about the condition itself [14,50] and whether there will be commercial use [32]. For example, a UK study found three-quarters of participants were worried about the security risks introduced by information sharing associated with the introduction of an electronic health record system by the National Health Service (NHS), identifying the need for trustworthy security and privacy-preserving mechanisms together with intensive awareness and engagement initiatives [69].

Better communication between patients and those they trust most in clinical teams can reduce patient privacy concerns and assist them becoming more comfortable with the collection and sharing of their own information [56]. It is also recognised that informed consent cannot always be achieved when there is an imbalance in the relationship [21], as there is between patients and the state, especially when using digital methods where engaging with information, deliberating and making decisions about consent are recognised as more difficult [8,34].

Previously there has been significant work on the sharing of health data generally [94] and app health data [17] between health professionals and with researchers, including for public health research, but the COVID-19 pandemic has increased attention on data sharing for immediate collaborative responsive action, such as detection, protection and enforcement to combat the spread of communicable diseases, particularly regarding mobile technology for contact tracing. Whilst there is a growing realisation of the costs and benefits of personal data in pandemic situations [9,10], unlike our study, others have not examined such privacy concerns by individuals themselves in this public health emergency context where the data is for immediate use to identify arising pandemics.

2.3 Contact tracing and other public health emergency applications

Mobile health data has the potential to support disease surveillance, preparation and response [44]. Contact tracing had been found to be effective in contributing to the control of smallpox and SARS, but was less effective for some other contagious diseases [43]. Such findings led to a belief that isolation and contact tracing could potentially contribute to the control of COVID-19 [37]. Although our study focuses on generic device data extraction rather than via apps with full informed consent and reliable privacy protection, we illicit broader responses regarding contact tracing, which is important as intense public dialogue related to the COVID-19 pandemic is progressing. In addition, some countries mandate the installation of a contact tracing application (e.g.: India, Qatar, Turkey) [2]. Mobile data was used from an early point for COVID-19 in South Korea to facilitate surveillance, testing, contact tracing, and quarantine/self-isolation [95]. All forms of contact tracing require widespread community engagement, support [100] and trust [1,26]. Digital contact tracing apps were suggested as a method to disrupt transmission pathways in ways that manual tracing would not be able to achieve [26], requiring higher rates of uptake to be effective in reducing new cases and hospital admissions [77], but it is suspected adoption is being held back by insufficient engagement with citizens to reassure them about protecting their privacy [76]. Although contact tracing benefits are mostly realised by the wider population, an individual's personal health risk can reduce if they and all their close contacts participate [70].

It is recognised that technology considerations cannot be examined in isolation, and require people's trust in the systems being developed [77]. The WHO suggests communications about contact tracing should "emphasize solidarity, reciprocity, and the common good" but recognises the need to be open about how information is used, stored, accessed, how people's identity and data will be protected [100].

Perceived public health benefit, individual benefit, accuracy and privacy concern affect willingness to install COVID-19 contact tracing apps [27] and there are trade-offs between these, user agency and transparency [75]. Both principle types of implementation – with centralised or decentralised storage and processing – have their own underlying privacy risks [18,24,29,92,101]. Development and deployment of contact tracing apps during pandemics are both live research and an intervention to support the response [63] where ethical arguments have to balance the benefits and harms in the same way as other health interventions [26].

In this context, the suffering and even death of others as the result of COVID-19 might be used to justify some degree of privacy infringement [67,70]. But privacy infringements do not necessarily occur at a single moment in time: people might also be concerned about their privacy in the future – by harmful use of their data or loss of privacy protection that is never reinstated after the end of the pandemic [3,55,70], or as circumstances and their views on consent change over time [21,57]. Immediately prior to the lockdown starting, UK citizens had been surveyed about specific, and thus limited, app-based contact tracing, showing wide support, similar to findings for the same survey

conducted in Germany, France, Italy and the USA [39]. Most respondents thought the app should be installed automatically, with the main concerns being government surveillance, increased anxiety about the pandemic, and fear of phones being hacked. Other studies in progress are collecting data on people's values and concerns around sources and sharing of data in COVID-19 contact tracing apps [85]. Our study connects to this line of work but instead examines broader device data access and sharing within the context of COVID-19.

3   METHODOLOGY

We explore attitudes and values regarding the access of personal digital-device data for public health purposes, building upon a 2018 study undertaken by Watson and Smeddinck [93]. Making use of scenario-based research, which is commonly used in research around privacy expectations [53] that study had used verbatim transcription of facilitated discussions with a small number of participants, coding and thematic analysis to examine people's opinions towards extraction of personal data by the state in the event of three different health event scenarios. The study we report on in this work was likewise based on three situations but was designed to provide both a qualitative and a quantitative assessment and comparison between the scenarios, as well as to consider demographic and trait factors that might inform and influence people's concerns and attitudes, thus requiring a larger number of participants. Mixed methods [90] research has the ability to provide a more complete understanding [22], additional insights [5] and opportunities for considering multiple perspectives and triangulation [84] especially where there are significant uncertainties and complexities such as human factor and ergonomic problems in healthcare systems [15]. We explore extraction on a device-level in order to avoid focus on the particularities of a specific (version of an) app. Our study was undertaken after the point when the WHO had declared the COVID-19 virus situation a pandemic, approximately 6-7 weeks from when initial lockdown regulations had come into effect in the UK, influencing both the design, execution and analysis of the study.

3.1 Timing of the study

The UK government's scientific advisors had been using epidemiological models to examine the relative effects of non-pharmaceutical interventions and, in March 2020, ceased mass testing and contact tracing and gave advice that contact tracing should be adopted at a future point of the pandemic once case numbers reduce [60]. Our study was undertaken two months later in mid-May 2020 when there were over 200,000 confirmed cases, and had been almost 32,000 deaths in the UK [97], at the height of trust in the government's approach when 60% of people believed Britain would be "very" united after COVID-19 [65]. In the week prior to starting the data collection phase of our study, a UK contact tracing app began widely publicised trials in one geographical area. As it transpired, the timing of the study was more significant because our data collection took place while there was still broad agreement and social-cohesion regarding the efforts being taken to control the spread of the COVID-19 virus, but only two weeks prior to a recognised breakdown of trust in the UK government as a result of the actions of its most senior official [25].

3.2 Impact of COVID-19

We were unable to use in-person discussion group, ethnography or contextual inquiry methods to understand people's views on the three scenarios due to restrictions imposed in the UK when lockdown regulations came into effect on 26 March 2020, as well as restrictions by the sponsoring university on research with human participants. Therefore, to avoid any face-to-face contact and to prevent excluding participants who were clinically vulnerable, self-isolating or otherwise limited due to virus control measures or personal concerns, a remote survey method was selected. To support our mixed-methods analysis technique we created a single deployment including both questions suited for qualitative analysis and other questions suited to quantitative analysis. It was recognised that being online-only,

this would need to be reflected in the subsequent analysis when considering wider applicability of the results. In addition to these methodological impacts, the context of the pandemic also led to content changes in the survey itself to reflect the reality that participants would be sensitised to potential virus effects and controls, and related public debate about these.

3.3 Ethical considerations

The study sought information from people capable of providing fully informed consent and thus chose to exclude those under 18 years of age. Although personal health matters were avoided through the use of scenarios and the phrasing of the question items, we were aware some people might already be distressed or become anxious while taking part. Participant briefing information included a warning the survey touched on health considerations in a general sense which some might find upsetting. Ethical approval was granted through the sponsoring university's ethics review processes.

3.4 Repeatability and reproducibility

Given methodological restrictions and timing, it is important to comprehensively reason for and document the study design, configuration and analysis [7,46]. Replication may be particularly important for future work extending to a longitudinal study to examine the research questions through different phases of the COVID-19 virus pandemic as changes occur to the related lockdown restrictions and digitally-supported government interventions are deployed. Full study artifacts including the survey are made available (cf. Acknowledgements).

3.5 Survey design

Building on the previous foundational study and a piloting iteration, the questions for our online survey were structured around three medical event scenarios (heart palpitations [S1], meningitis [S2] and influenza pandemic [S3]), including both closed-ended Likert scale and open-ended responses. Heart palpitations are noncommunicable and would not usually be the cause of a wider public health emergency; meningitis and influenza are both contagious, but influenza normally kills 50% more people worldwide per annum than meningitis [30,68]; a pandemic is "an epidemic occurring worldwide, or over a very wide area, crossing international boundaries and usually affecting a large number of people" [40]. Due to advances in emergency response, meningitis generally no longer leads to pandemics; thus, influenza pandemic (S3) is the most significant of the scenarios from a public health perspective (highest *severity*). The scenarios described situations for some other person, named "Joe" for three reasons. Firstly, the previous study used this format, secondly to permit participants to contribute without revealing personal information about themselves, and thirdly to reduce potential distress in case of participants being currently afflicted by conditions.

As the survey's design progressed, the COVID-19 situation grew more serious and lockdown began, leading the researchers to consider whether the third scenario should be replaced by coronavirus, declared a pandemic by the WHO on 11 March 2020. The situation was a current event rather than an imaginable scenario, with so little medical consensus at the time combined with high levels of confusion reinforced by misinformation circulating in the general public, as well as people being affected in many different ways. We believed there would not be consistency to the public understanding of this affliction, to the point it could not facilitate a reasonable comparison, whereas influenza is a clearly established communicable disease, and therefore contrasting with the other scenarios would be more meaningful. Consequently, like in the previous study, COVID-19 was not mentioned in the questions on demographics/other factors or three scenarios. The research team were aware that the public, and as it transpired later the government too [79], were conflating influenza with COVID-19 and as such wanted to allow participants to respond in any way they wished to the third scenario (S3), influenza pandemic. However, given the context, it was considered necessary to extend the data collection beyond

the three scenarios with some supplementary questions, specifically related to COVID-19 to gather further data for the coinciding pandemic in addition to the standard questions in S3.

One survey question touches on forced extraction of personal data from mobile devices without consent; it should be noted this is not currently a power the UK's public health authorities have, but as the previous study identified, it is a potential activity permitted by UK anti-terror legislation. Demographic questions were included (age, gender, current work situation, UK country/region and whether living in an urban or rural area), a Likert scale question asked about *health concern*. Other traits/positions aspects included *civic engagement* [4] using 5 questions posed in the periodic Oxford Internet Survey (QP3 from OxIS 2019 [35]), a 14-item checklist [87] developed by the Scottish Council for Voluntary Organisations (SCVO) to measure the UK government-defined *foundation/essential digital skills* (EDS) [89], a 12-item inventory on *privacy concern* [72] (excluding factor 1a which related to a non-relevant concern about website providers), and a *computer anxiety assessment* (6 items) [47]. The trial survey was tested for phrasing, timing and any other issues with convenient subjects.

Randomising the order of scenarios was dismissed for three reasons: 1) learning effects are less important as this was not a trainable task, 2) we invited participants to reflect on the differences between scenarios allowing going back and forth and changing answers, 3) discussions in the foundational study had been undertaken in a fixed order of increasing *severity* of public emergency. Implemented in Google Forms in English only, Prolific was used for recruitment and screening, targeting a UK-only adult demographic. Participants were paid a fixed amount to compensate for their time, calculated at £10/hr.

3.6 Data collection, cleansing and approach to analysis

50 submissions were received within hours of deploying the study before it was noticed that a late change to the labelling of the scenario-based Likert scales meant that scale anchor value selections – while correctly displayed and selected – were not being stored, rendering the submissions unfit for full quantitative analysis. An additional 40 complete responses were captured after corrections. Four submissions were later removed due to containing random character sequences for some response fields. This left 86 participant records, of which 38 had fully recorded answers to all the Likert scale questions. We used a three-stage mixed-methods data analysis approach, to inform a grouped qualitative analysis and provide a holistic view of findings from which to make recommendations and conclusions. Qualitative analysis was initially carried out independently from the quantitative analysis. Lastly, quantitative outputs were used to section and scrutinise further the qualitative data.
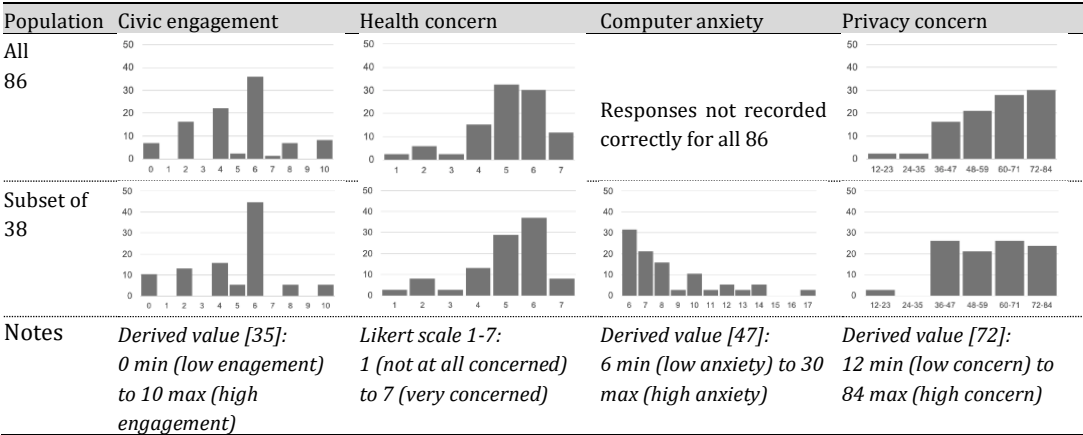
3.7 Participants

Demographic properties are shown in Table 1a including available UK-wide comparisons [23,51,64] with breakdowns showing comparable characteristics for both the subset (n=38) and the whole response set (n=86), indicating proper relatability for the third analysis stage building on combined data from both sets. Derived values were calculated from relevant answers; the scenario-independent quantitative scales for civic engagement, essential digital skills, computer anxiety and privacy concern are summarised in Table 1b.

Table 1a: Demographic comparisons of all 86 participants, subset of 38, and all UK [23,51,64].

Population	Proportion female	Age		Have full EDS	Live in a rural area
	%	M	SD	%	%
All 86	64	33	11	86	15
Subset of 38	63	34	13	84	11
All UK	51	40	–	78	17

Note: Full EDS means being able to do all 14 SCVO checklist questions for the UK's Essential Digital Skills framework.

Table 1b: Scenario-independent response comparisons by proportions (%) in stated value ranges.

Population	Civic engagement	Health concern	Computer anxiety	Privacy concern
All 86			Responses not recorded correctly for all 86	
Subset of 38				
Notes	*Derived value [35]: 0 min (low enagement) to 10 max (high engagement)*	*Likert scale 1-7: 1 (not at all concerned) to 7 (very concerned)*	*Derived value [47]: 6 min (low anxiety) to 30 max (high anxiety)*	*Derived value [72]: 12 min (low concern) to 84 max (high concern)*

The participants were from every UK country/region, were more predominantly female, younger and more likely to have full EDS than the general UK population. There was no matching UK-wide data to compare personal health concern, civic engagement, computer anxiety or privacy concern.

4 ANALYSIS AND FINDINGS

We report on all three stages of analysis independently below. Together they provide greater insights into the research questions, by contextualising Stage 1 qualitative findings relative to outputs from Stage 2 quantitative findings, as summarised in the discussion section.

4.1 Stage 1: Thematic Analysis

Analysis was undertaken on response data from all 86 submissions. Querying open comments, as well as reasons for/against public health agencies having data access, the validity of data, for the individual user and other people were extracted and analysed using thematic analysis [12]. This was performed using a social constructionist approach, where perceptions of public health risk impact and likelihood are not objective conditions, but are rather something developed by individuals and communities in wider societal and current event contexts [83]. Therefore, this is a method particularly suited to RQ1 and RQ2 relating to people's perceptions and views [20]. Stage 1 was separated from Stage 2 analysis, to develop data-derived and researcher-derived codes independently from early quantitative analysis findings. The research team member allocated to this task undertook familiarisation with the data by reading and re-reading the text responses making initial noticings in the form of notes and hand-sketched diagrams. There was no attempt to duplicate codes/themes reported in the previous study. It had been hoped to analyse responses between scenarios but the written responses were somewhat terse (mode character count ranging from 13 to 199 per question, excluding "is valid..." and "is there anything else.." questions), leading to the decision to consider whether additional insight could be gained once the per-scenario quantitative questions had been analysed in Stage 2. Codes were collated into candidate themes, checked against the data and the theme coherence reassessed and modified repeatedly. Quality assurance began by re-reading the coded data and checking this against the candidate themes. Further checks involved other team members reviewing the code book and updated candidate themes, discussing the relevance of the interim findings, and further re-analysis. Through this reflexive organic thematic analysis, an overarching theme of *lockdown may unlock the data* was conceptualised describing how coveted personal data should be protected and thus "locked up" (RQ1: opposed to access), but that wider societal need is perceived as a reason to "unlock" and share the data (RQ2: increasing public health risk reduces this

opposition) where "lockdown" acknowledged the situation at the time of data collection. The overarching theme comprises four themes *location, location, location* (T1); *mitigations matter* (T2); *questionable data* (T3); and *it's risky for me and others* (T4).

4.1.1 T1: Location, location, location

This theme is directly derived from the predominant number of references to the usefulness of location data, relative to other data types, mentioned by over 90% of participants (68 participants of 86 coded with 136 references to 'physical locations' and 57 coded with 91 references to 'nearby people') across responses to the three scenarios and supplementary COVID-19 responses. These related to two codes – one for *physical location* of the person themselves and the other for *nearby people*. The second was not mentioned at all by any participants for S1 (heart palpitations); for S2 and S3 both codes were mentioned more often than in S1, and approximately equally. In COVID-19 responses, these two codes were mentioned as many times, as in all 14 responses for the three scenarios i.e. much more often. This theme gathers ideas that location data can assist in people's diagnosis, treatment or recovery during a health event, with mentions especially prevalent in responses to the communicable conditions. Location data can reveal *"people they have been in contact with"* (P19) foregrounding proximity *"who do they get close to"* (P49), and timeliness *"recent location"* (P67) data, being aware how relevance can change. A few participants mention *"where people who are infected have been"* (P40), how *"close they have been to people with symptoms"* (P71) with some specific references to a *"Corona virus contact and tracing App"* (P38). There is concern location data may be used against the individual, for example during a pandemic situation:

> *"expose his movements/contacts with people contrary to lockdown rules"* (P70)
> *"to see where they are going when they are supposed to be on lockdown"* (P77)
> *"if there was a lockdown situation and he wasn't meant to be socialising"* (P78)
> *"[to check] if they have been for COVID19 testing"* (P73)

The insights (I) we derive in Stage 1 (I1-) from this theme's findings are:

I1-1 Location-based data are uppermost in people's minds when thinking about what data might be useful to share with health authorities to assist in combatting communicable diseases.

I1-2 In noncommunicable diseases, other people's location is not considered relevant, but location data may still be relevant when exercise/mobility may affect diagnosis, treatment or recovery.

Potential harms from sharing location data are addressed in T4 *it's risky for me and others*, but ways to prevent or reduce harms fall within T2 *mitigations matter.*

4.1.2 T2: Mitigations matter

During data familiarisation it was noted there was a common sentiment apparent in the data of the need to be convinced that any shared data would be used legally for the intended purpose, not be misused, and be protected. This was only mentioned in the responses to the 2 supplementary COVID-19 questions (by 35 out of 86 participants coded with 48 references comprising 11 sub-codes e.g.: 'general data protection measures', 'limited types of data accessed', 'optional consent or opt-in permission'). Another 29 (of 86) opposed to sharing (coded 'no [data access] is not acceptable') did not refer to any mitigations at all. T2 draws risk-reduction ideas together, revealing a wide-range of mitigations people thought were relevant. With a reluctance to share data but an acceptance that personal data should be shared for the greater good, therefore allowing it conditionally (e.g.: "no, but okay if..." or "okay but..."). Particular mitigations identified are anonymisation, notification or being informed, only extracted from a dedicated app, only for defined purpose, comply with legislation, limits to data sharing *"no further than the relevant NHS department"* (P39), short data retention *"during the coronavirus outbreak"* (P36), limits such that *"only vital information was collected"* (P70), ability to control data contribution and *"people should be able too see"* (P66) what data is accessed. A requirement for prior consent was mentioned by a quarter of participants:

> *"I think they should always ask for consent before accessing the data"* (P14)
> *"I think it is important to ask for consent and respect the answer"* (P40)

The additional Stage 1 insights we derive from theme T2's findings are:

I1-3   In situations where *severity* is greatest (i.e. COVID-19), a willingness to share comes with caveats that data must be used within expected legal bounds, only used for the intended purpose, not be misused and should be protected.

I1-4   Consent is a dynamic thing, that can change with time and situation.

4.1.3   T3: Questionable data

T3 incorporates participants' views that personal data on mobile phones is a poor source of information for use by authorities in health events. 78 of 86 participants were coded with 325 references spanning 12 codes in three sub-code groups: 'poor quality data' (e.g.: gaps, modified, unrepresentative), 'poor device or detector' (e.g.: faulty, inaccurate, not medical standard), 'poor information source for the condition' (not relevant or, relevant but nor for this situation). Of these, 15 participants were coded 'data relates to someone else'. All three scenarios included references to unsuitability of data, but there were almost no occurrences of these in the supplementary COVID-19 responses. S1 was dominated by mention of *data gaps, unrepresentative data* and especially *inaccurate, imprecise or unreliable*; where some participants also mentioned the type of data as activity or heart rate data. Across all three scenarios, there was a significant proportion of participants who thought phone data was not relevant to the particular health condition (heart palpitations, meningitis or influenza), and slightly more so for S2 and S3 than S1. Despite some *"[people] probably not understanding why the data is needed or how it will be used"* (P50), 60% mentioned available phone data *"isnt relevant"* (P01) for the particular condition or *"no relevant data on his phone that could help others"* (P79), although for communicable diseases there was widespread acknowledgement that location data was relevant health-related data. The necessity for higher medical standards is revealed through perceptions that data, sensors, apps and phones are *"not medical-grade"* (P49) with a preference that *"proper hospital equipment and monitoring should be used for actual diagnosis"* (P07).

Questions about data validity were also mentioned specifically about tracking how and where a disease is spreading: *"a lot of people do not track their movements on their phone"* (P31), *"it cannot track every single human he came into contact with"* (P43), *"it's unlikely that the device is going to sufficiently identify everyone he has been in contact with"* (P49), and with an assumption there is a dedicated app *"a lot of people with the virus wouldn't have the app on their phone so there wouldn't be enough data"* (P71). It was recognised that use of data for this purpose can also be undermined by poor data quality due to being weak evidence, gaps in data collection, data may have been modified, data is unrepresentative, or relates to someone else. The insights we derive from theme T3 are:

I1-5   Smartphone data could be useful personally if it is either very specific to the condition (e.g.: exercise routines or pulse rate data for heart palpitations), or to the wider public in a pandemic (e.g.: locations and contacts).

I1-6   There is widespread concern about data quality from personal devices.

I1-7   People are aware their phones store information, not only about themselves, but also about other people.

4.1.4   T4: It's risky for me and others

The fourth theme, *it's risky for me and others*, is drawn from participants' concerns about negative impacts that might occur, spanning 19 codes from the thematic analysis. All 86 participants were coded with 458 references to risks comprising 19 sub-codes; 205 references related to 'general loss of privacy' and the others to more specific codes e.g.: 'catch disease', 'discrimination (general),' 'discrimination (treatment)', 'embarrassment (sins and vices, immoral actions)'. These codes were not found in responses to the supplementary COVID-19 questions, but those questions were seeking different types of information. Across all the 19 codes, there was little variation between the scenarios. Apart from *general unspecified loss of privacy*, the risks coded most often were *embarrassment* followed by *misuse or illegal use [of data] by others*, *become subject to legal action*, and *[cause] problems for others*. Concerns highlighted issues that might occur as a result of data release: loss of privacy about personal hopes,

fears or plans, embarrassment (35 participants), discrimination (in treatment 3 participants, or because of who/what they are 7 participants), revelation of incriminating information or of illegal activity (19 participants 'become subject to legal action' and 7 'incrimination'), and in data transfer procedures: loss of access to or damage to their device, or even catching the disease from the device once it is returned to them. These are similar to the concerns raised under the theme *access is risky* for individuals in the previous study [93] but broadens these worries to others including their employer:

> *"he doesn't want to jeopardise someone else's privacy"* (P09)
> *"he feels he coud get himself or someone else in trouble"* (P39)
> *"he may also use his phone for work purposes and this could be a breach of data protection"* (P83)

Participants' thoughts also extend beyond the moment of the health event to future risks concerning misuse of the data, whether accidental or malicious such as being *"afraid personal information will be misused"* (P45) or of *"third parties accessing the data"* (P90) or bank details being revealed, which lead to data or financial loss. There is an acceptance that *"sadly not everybody is trust worthy"* (P19) with some mentioning a lack of trust in the state and its institutions *"not trusting the government, therefore not trusting the NHS"* (P82), but for some risk reduces if data *"go no further than the relevant NHS department"* (P39). The insights we take from T4's findings are:

I1-8 Personal risks from data sharing are similar for all types of disease with significant concerns about revealing illegal behaviours and activities.

I1-9 Sharing of personal data could have adverse impacts on people other than themselves.

4.1.5 Discussion of Stage 1

Given the widespread public discussion and debate about contact tracking and tracing at the time of the study's data collection [39], and how commonly the value of apps is derived from location-based services [42], it is not surprising that location data was strongly represented in participants' responses (I1-1). This location data was deemed less relevant to S1 (heart palpitations) and nearby people not relevant at all, where it seemed possible that other sensor data might be more useful (I1-2).

Increasing severity of scenario is also likely to coincide with decreasing co-presence of interaction [88], but even so across all three scenarios there is consistency in the perception that much data on a mobile phone is not relevant, apart from location data during pandemics situations (RQ2) when there is greater willingness to share (I1-3), and consistency in the risks participants believe exist (I1-8) requiring ongoing consent which can change over time (I1-4); this might be that given unrestricted phone access, all data is seen to be equally available. Apart from general privacy concerns, there were many concerns (RQ1) about embarrassment or risk of legal action due to what data on their phone might reveal across all scenarios (I1-8), but also notably concerns about risks to others too (I1-9) including data that might be about others rather than themselves (I1-7).

Comparing T1 to T4 with themes from the previous study we found less emphasis on bodily concerns/comparisons and more emphasis on location data and physical position, thereby not emphasising an intrusion into the body. Both location and body are relevant to RQ1, but public debate at the time of our data collection had possibly shifted the concern from "about my body" to "where is my body" along with a wider recognition "it's not just me" but that "other's data and health matter too". These findings suggest a need to communicate purposes, analysis aims and use clearly, requiring consent and ongoing control on the use of personal data, even during a pandemic, rather than being a one-time event recognising that data value and acceptable use change over time.

Unlike in the previous study, participants did not mention compromised smartphones (e.g.: hacked devices, backdoors). The scenarios were purposefully unspecific about how data might be shared/taken but it was notable that some participants pointed out that sharing location data from a dedicated app would be preferential to a phone-wide deep trawl for data, as this could be limited to those data relevant to the disease and context at the time:

"I think better off would be creating an application for the NHS that automatically updates information to their servers (only if you opt in to it, ie you might opt in to sharing heartbeat data with the NHS but not want to show your location)" (P26)

"There is no such platform for users to share only certain information with, for example NHS. In this case, if the user could only share his movements in a map in the last 7 days, and do so transparently, I believe many people would be willing to share this data, although a strong data protection policy should be in place" (P42)

This holds an understanding of the importance to identify which types of data are subject to processing when creating applications in this area – whether sensors are capturing location data or vital signs – participants clearly note and then express different values and concerns for public health events, as similarly identified by Lin et al. [49].

4.2 Stage 2: Quantitative Analysis

Analysis was undertaken on response data from the subset of participants (P53-P90, 38 submissions) which had not been affected by the mis-recording of anchor value Likert scale questions. Analysis was undertaken of responses to the 6 closed-ended questions for each of the three scenarios, 4 supplementary COVID-19 questions and scenario-independent questions including the derived values for civic engagement, digital skills, technology anxiety, and privacy concern. Between scenario differences were examined to address RQ2, H1 and H2, and responses to pandemic situations (S3/influenza and supplementary COVID-19 items) were examined to address RQ1, RQ3, H3 and H4.

4.2.1   Between scenario comparisons for each question

Response summaries for identical questions across S1-3 are shown in Table 2. Shapiro Tests identified responses to some questionnaire data to be not-normal so non-parametric methods were used.

Table 2: Quantitative data means (M) and standard deviations (SD) for S1-3.

Question items (Q1-6)	S1: M	*S1: SD*	S2: M	*S2: SD*	S3: M	*S3: SD*
Q1 Data on the phone **is valid for use in Joe**'s diagnosis, treatment and/or recovery	2.8	*1.3*	2.4	*1.3*	2.8	*1.4*
Q2 Data on the phone **will assist Joe**'s diagnosis, treatment and/or recovery	**2.9**	*1.2*	**2.3**	*1.3*	2.6	*1.3*
Q3 Data on the phone **will assist other people**'s treatment and/or recovery	**1.8**	*0.9*	**2.9**	*1.4*	**3.2**	*1.2*
Q4 If Joe in **unconscious** or unable to give fully informed consent for some other reason, the NHS should be able to **access data on the phone without consent**	2.4	*1.5*	2.3	*1.5*	2.3	*1.6*
Q5 If Joe is fully informed but **refuses permission**, the NHS should be able to **access data on the phone without consent**	**1.1**	*0.3*	**1.6**	*1.0*	**1.8**	*1.2*
Q6 It is Joe's **duty to share** all data on the phone in this scenario	**1.7**	*1.1*	**2.3**	*1.6*	**2.6**	*1.7*

Note: Data for 38 participants, out of all 86 asked "For this scenario, please indicate in how far you agree or disagree with each of the following statements" with responses selected from 1 (disagree) to 5 (agree).

These were utilised to examine reactions/values to differing *severity* between scenarios with Friedman Tests identifying significant differences on four items:

Q2: Whether data on the phone will assist Joe's own diagnosis, treatment and recovery (x^2=6.6, df=2, p<.05)

Q3: Whether data on the phone will help other people's treatment and recovery (x^2=25.6, df=2, p<.05).

Q5: If Joe is fully informed but refuses permission, […] access data without consent (x^2=21.3, df=2, p<.05).

Q6: It is Joe's duty to share data on the phone (x^2=14.4, df=2, p<.05).

Bonferroni corrected post-hoc Wilcoxon Signed-rank Tests (Wilcoxon used due to being well established and more conservative) showed sig. differences for all four questions between S1 and S2, and S1 and S3 (p<.05), with at least medium effect size, but not between S2 and S3 (p>.05):

Q2: Between S1 and S2 (Z=2.81, p<.05, r=.32).

Q3: Between S1 and S2 (Z=-3.71, p<.05, r=-.43) and S1 and S3 (Z=-4.58, p<.05, r=-.53).

Q5: Between S1 and S2 (Z=-3.31, p<.05, r=-.38) and between S1 and S3 (Z=-3.59, p<.05, r=-.41).

Q6: Between S1 and S2 (Z=-2.77, p<.05, r=-.32) and between S1 and S3 (Z =-3.17, p<.05, r=-.36).

We examined any Pearson's correlations between the six scenario questions with the health event independent questions, other than EDS, described in Table 3.

Table 3: Definitions, means (M) and standard deviations (SD) for *health concern* and four derived variables.

Variable	Source	Description	Possible Range	M	SD
Civic engagement	[35]	5 questions, scored as 5 plus sum of 'yes'-answers minus sum of 'no' answers	0-10 (higher is more engaged)	4.8	2.5
Essential Digital Skills (EDS)	[87]	14 life-skill questions either "I can do this" or "I can't do this"	"Full" (all 14), "Some" or "None"	–	–
Health concern	See note	Single question 7-point scale 1 (not at all concerned) to 7 (very concerned)	1-7 (higher is more concerned)	5.0	1.5
Computer anxiety	[47]	5-point scale 1 (disagree) to 5 (agree) for 6 questions (two reverse scored), summed	6-30 (higher is more anxious)	8.4	2.8
Privacy concern	[72]	7-point scale 1 (not at all concerned) to 7 (very concerned) for 12 questions, summed	12-84 (higher is more concerned)	58.1	15.4

Note: Data for 38 participants, out of all 86 asked. The health concern question was "Would you describe yourself as a person who is concerned about their health?".

The correlation between privacy concern (pc) or civic engagement (ce) and unconsented data access have been calculated along with their effect sizes [r: correlation and d: effect size]. In both S1 and S3 when either privacy concern or civic engagement increases, the inclination towards agreeing to unconsented data access against the person's decision decreases (S1 (ce): r=-.4, d=1.2, S2 (ce): r=-.3, d=1.2, S3 (ce): r=-.3, d=1.2) and (S1 (pc): r=-.005, d=5.1, S2 (pc): r=-.06, d=5.2, S3 (pc): r=-.1, d=5.3) respectively. Together, these findings appear in-line with Lin et al. [49] who demonstrated users' personal feelings and trust decisions are affected by their expectations and the reason for using sensitive app resources.

For Stage 2 (I2-) quantitative findings based on correlation and significant effects:

Regarding usefulness of mobile phone data:

I2-1   Participants agreed more that data helps Joe in S1.

I2-2   Participants agreed more that data helps others in S2 and S3.

Regarding differences between scenarios relating to RQ2:

I2-3   Participants felt there is a greater duty to share data if there is a communicable disease, and even more so in a pandemic, but overall even in a pandemic people neither agree nor disagree that there is a duty to share.

I2-4   Participants agreed somewhat that data might be useful and valid for the patient themself in the case of non-communicable diseases (possible e.g. with relevant data like heart rate monitoring

or exercise tracking) or in the case of a pandemic (bigger problem, and widespread knowledge of location data and contact tracing), but agreed slightly more than for communicable diseases the data may be useful for other people's benefit.

I2-5   Participants disagree with the possibility of unconsented access when a person has refused, especially so in the case of non-communicable diseases (S1).

Regarding consent:

I2-6   As privacy concern increases participants felt it was increasingly important the choice not to share data should be upheld in S1 and S3.

Regarding response variability:

I2-7   Four (Q2 to Q6 above) of the six scenario questions have greater differences in responses.

I2-8   Responses to S2 and S3 are more similar to each other and less similar to S1 (least *severity*), with S3 most different to S1.

4.2.2   Between questions comparisons for each scenario

Pearson correlations between pairs of questions within each scenario as well as between scenarios showed large effect sizes (mean d>=.8 after Cohen) in the following cases: In S1, *agreeing that data on the phone might help Joe's diagnosis, treatment or recovery* is positively correlated (d=2.0) with *agreeing that the NHS should be able to access his data even if he has refused*. This also shows a positive correlation (r) of medium effect, in S2 (r=.5, d=.6) and S3 (r=.6, d=.7). There is a positive correlation across all three scenarios between agreeing the *data might help others* and that *data may be accessed even if consent was not given* (d_{S1-3}=1.1, 1.1, 1.2).

Pairwise Pearson correlations between scenario questions with the health event independent scales (Table 3) and separately between the supplementary COVID-19 questions showed negligible or small effect sizes. The only derived insight regarding correlations between questions is:

I2-9   If the *data are perceived as possibly useful*, participants believed *access without consent* is more reasonable.

4.2.3   Reactions/values to COVID-19 virus

The supplementary COVID-19 questions were not a separate scenario, but complement the information gathered. As the summary shown in Table 4 indicates, the participants agree somewhat with all four statements, but there is reasonably large variance of opinion and the phrasing of the items is positively inclined.

Table 4: Quantitative data means (M) and standard deviations (SD) for the 4 COVID-19 supplementary questions.

Variable (see supplementary material identified in Acknowledgements)	M	SD
I am now more concerned about disease prevention	**3.5**	*1.0*
I would now be **more willing to share** private information for the purpose of controlling the spread of highly contagious diseases	3.2	*1.4*
I think that **information** that can be **gathered** from private smartphones **can be helpful** in controlling the spread of highly contagious diseases	3.4	*1.3*
I am **worried that information** gathered from private digital devices for the purpose of controlling the spread of contagious diseases **may be misused**	**3.6**	*1.3*

Note: Data for 38 participants, out of all 86 asked "Please indicate in how far you agree or disagree with each of the following statements" with responses selected from 1 (disagree) to 5 (agree).

4.2.4   Clustering in pandemic-related question responses

The lack of strong correlations in **pairwise** comparisons led us to consider whether there might be **groups** of properties or groups of participants that would provide greater insight to answer RQ3. Data driven methods can be used to generate archetypal user descriptions [58,78,86,98]. Manual participant

clustering analysis of qualitative data can use techniques such as affinity diagrams, card sorting and expert panels but this is less suited to larger numbers of data subjects, whereas semi-automated analysis of qualitative data depends on the ability to compare textual data [13]. In our study, we also have quantitative data which can guide the identification of groups with structural differences. To take the study's analysis from a question-by-question exploration to groups of participants, the most commonly used clustering analysis method [78] k-means unsupervised clustering was utilised to partition the 38 participants into k-groups, where each group's members are as similar to each other, and as dissimilar as possible to those in other groups. As clustering with too many variables can be problematic, the aspects identified with most variance from the Stage 2 analysis were selected to exclude less relevant variables. The aspects used for clustering comprised the previously identified four most dissimilar S3 responses (I2-7) and responses to all four COVID-19 questions. The precision of segmentation was considered of lesser importance than what might be derived from our subsequent Stage 3 analysis. In other words, the research team expected some degree of defining feature overlap [74,98] and therefore were less interested in the uniqueness of partition achieved by cluster analysis but rather to use the clusters as lenses to further explore and gain insights from the qualitative data. The Hartigan-Wong algorithm with 25 random starts for k-means was explored against the eight quantitative values for the 38 participants. The means square differences were calculated for k up to 10 (Figure 1a) with this suggesting using a k value of 2 or 3.
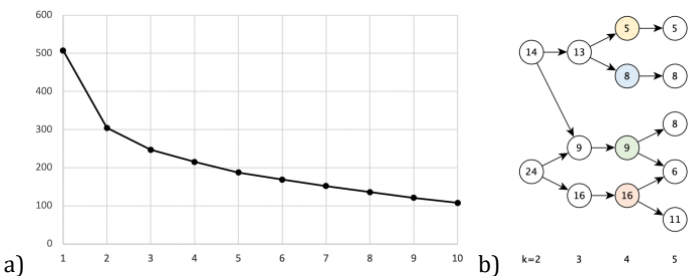
Figure 1: a) Elbow chart showing within k-means cluster sum of squares versus number of clusters (k) for k=2 to k=10.
b) Illustration of how k-means cluster size and total membership of 38 change for k=2 to k=5.

While k=2 provides good separation, we wanted to generate a reasonable number of clusters for the next stage of analysis. It was observed that by incrementing k from 3 to 4, two clusters remained with identical participants (Figure 1b). As differences decrease only slightly and almost constantly thereafter, k=4 was chosen, resulting in groups A-D. Data for each of the 8 aspects used for the cluster segmentation were calculated from the participants assigned to each group (Table 5).

Table 5: Means (M) and standard deviations (SD) for the cluster-defining variables by Group (A-D) and All.

Section	Variable (see supplementary material)	A (n=5) M	A SD	B (n=9) M	B SD	C (n=8) M	C SD	D (n=16) M	D SD	All (n=38) M	All SD
S3	Data helps Joe	**4.2**	0.8	2.8	1.1	3.3	1.3	**1.8**	0.9	2.6	1.3
	Data helps others	**4.6**	0.5	2.8	0.7	4.3	0.9	2.5	1.0	3.2	1.2
	Access without consent	**4.4**	0.5	**1.7**	0.7	1.6	0.7	1.1	0.3	**1.8**	1.3
	Duty to share	**4.6**	0.5	**1.7**	0.7	4.8	0.5	1.4	0.9	2.6	1.7

Section	Variable		A (n=5)		B (n=9)		C (n=8)		D (n=16)		All (n=38)	
	(see supplementary material)		M	SD	M	SD	M	SD	M	SD	M	SD
COVID	More concerned about disease prevention		**4.2**	0.4	3.9	0.6	3.8	0.9	3.0	1.0	3.5	1.0
	More willing to share		**4.0**	0.7	3.9	0.6	**4.6**	0.5	**1.9**	0.9	3.2	1.4
	Information can be helpful		**4.0**	1.2	**4.0**	0.7	**4.5**	0.8	2.3	1.0	3.4	1.3
	Worried about misuse		2.8	1.5	2.9	1.1	3.6	1.4	**4.2**	1.2	3.6	1.4

Note: Data for 38 participants, out of all 86. Means <=2 or >=4 are in bold.

Data for the health event independent quantitative aspects of each group were calculated (Table 6).

Table 6: Means (M) and standard deviations (SD) for *health concern*, the four derived variables, age and gender for Groups A-D and All.

Variable	Range	A (n=5)		B (n=9)		C (n=8)		D (n=16)		All (n=38)	
		M	SD	M	SD	M	SD	M	SD	M	SD
Civic engagement	0-10	**5.6**	1.7	4.9	3.1	4.3	2.3	4.8	2.6	4.8	2.5
Essential Digital Skills	% "Full"	40		89		88		94		84	
Health concern	1-7	4.6	1.1	5.1	0.8	4.5	2.4	5.3	1.3	5.0	1.5
Computer anxiety	6-30	8.2	2.3	8.1	2.6	8.3	2.7	8.6	3.2	8.4	2.8
Privacy concern	12-84	51	16	**65**	13	51	18	60	14	**58**	15
Age	–	36	17	34	16	34	16	33	9	34	12
Proportion female	%	**80**		**44**		50		75		63	

Note: Data for 38 participants, out of all 86. More notable values are emboldened.

Group A is the most *civically-engaged* group (M=5.6, SD=1.7) and as one of the groups that tends to agree there is a *duty to share*, they also agree the *NHS should be able to access data on the phone without consent* (M=4.4, SD=.5). They are the group now *more willing to share* (M=4.0, SD=.7) and also agree that *data will be helpful for the individual* (M=4.2, SD=.8) and *others* (M=4.6, SD=.5) for influenza and for the purposes of helping to control the spread of contagious diseases more generally (M=4.0, SD=1.2).

Group B slightly agree that data is useful for others in an influenza pandemic (M=2.8, SD=.7) but agree much more so for coronavirus (M=4.0, SD=.7). *Group C* rather disagree the *NHS should be able to access data on the phone without consent* (M=1.7, SD=.7) although being one of the two groups that tends to agree there is a *duty to share*. They rather agree that *data is useful for others* in an *influenza pandemic* (M=4.3, SD=.9) but more so for coronavirus (M=4.5, SD=.8). Group C is the group most *willing to share* (M=4.6, SD=.5).

Group D are the most worried that *personal data may be misused* (M=4.2, SD=1.2), as well as being the second highest *privacy concern* group (after group B). This group slightly *disagree* such data would be *helpful in controlling the spread of a highly contagious disease* (M=2.3, SD=1.3) whereas across the other three groups people tend to agree it would be helpful (M=4.2, SD=.9). Group D also believe data isn't useful in the influenza case study, either to the patient Joe or others - and this is stronger disagreement than any other group. Group D also think data is less useful for COVID-19 than any other group mirroring Keith et al.'s observations [42] on perceived usefulness of data in mobile apps.

Unlike group D, groups A, B and C all agree information gathered from private smartphones can be helpful in this situation, whereas those in Group D are more even-minded, neither agreeing nor disagreeing. The group-wise comparisons suggest the following implications:

I2-10 Privacy concern does not appear strongly related to willingness to share.

I2-11 Civic engagement does not appear strongly related to willingness to share.

I2-12 The most civically engaged group are those least concerned about the NHS gaining access to personal data without consent against an individual's wishes.

I2-13 Belief in a duty to share does not appear strongly related to concerns about consent.

I2-14 Those with greater concerns about data misuse are also those least willing to share.
Based on clusters, these implications should be considered most formative and are clear candidates for future work.

4.2.5   Discussion of Stage 2

Regarding a requirement for consent (RQ1), participants believed (I2-9) access without consent is more reasonable if the data might possibly be useful for a public health situation.

The contagiousness of the event does affect participants' reactions and values (RQ2). Although participants neither (I2-3) agree nor disagree there is a duty to share, they felt there is a greater duty to share if there is a communicable disease and more so in a pandemic (H1 confirmed). They somewhat (I2-4) agreed data would be useful and valid for the patient themself in the case of a non-communicable disease and in a pandemic situation, and agreed slightly more the data may be useful for other people's benefit for communicable disease situations. Participants tended (I2-5) to be rather against any unconsented access to data when the person has refused, but more so in the case of non-communicable diseases (positive evidence for H2).

As privacy concern increases (I2-6), participants believe more that declining to share data should be respected, and data should not be taken regardless (RQ3), but only (I2-6) in the case of non-contagious disease (S1) or a pandemic (S3).

Considering concerns people have about privacy and willingness to share data these do not (I2-10) appear to be strongly related (H3 not confirmed). Similarly, civic engagement does not (I2-11) appear strongly related to willingness to share (H4 not confirmed).

The use of clustering exploration analysis has demonstrated how multiple variables are required to properly describe the groups and that individual aspects cannot be used to infer a comprehensive sense of group values and concerns. Given these participant groups had been derived through purely quantitative methods, we wanted to look back to the qualitative data associated with the groups to consider whether this might provide further insight.

4.3   Stage 3: Use of quantitatively-derived user groups to re-examine qualitative data

The quantitative analysis in Stage 2 had identified and characterised four groups of participants. In Stage 3 we reconsider the qualitative data associated with each group to see if this gains more nuanced characteristics.

4.3.1   Reaction to personal data access without consent

We examined responses by group members to the COVID-19 related open-ended question "What would you think if the NHS were able to obtain data from everyone's phones in a certain geographical area of the UK, without requesting consent?". This question had uniquely been coded earlier in the Stage 1 analysis against seven feelings of increasing concern due to the types of responses received such as *"I would feel violated"* (P32), *"I would not like this"* (P40) and *"i think this should be allowed"* (P66). These were coded with the interpretations from "unconcerned", through "reluctant acceptance", "unhappy", "concerned", "disagree" and "perturbed", to "feel violated" and, for the purposes of comparison, given a value of 1 to 7 respectively. Figure 2 shows these for each participant (if any code applicable) by group.
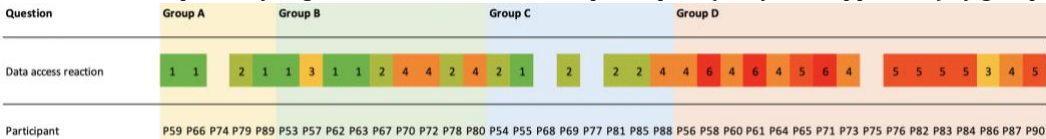
Figure 2: Relative concern ranking (1 unconcerned-7 feel violated) to unconsented access by the UK's NHS to personal data.

This visual comparison of researcher-assigned codes illustrates greater concerns for those in groups B and D, than A and C, showing consistency with the quantitative analysis results that groups B and D did not feel there is a duty to share and have greater *privacy concerns* (Table 6). There are noticeably higher values of concern on this descriptive scale for those in group D, which the quantitative analysis identified as being the group most worried about *data misuse* and the *least willing to share data* (Table 5).

4.3.2   Relation of clusters to themes

We then looked at the four groups of participants relative to the Stage 1 theme allocation, which had been undertaken in advance and separately to Stage 2, conceptualising four themes through researcher-assigned codes:

T1: Location, location, location (2 codes *nearby people* and *physical locations* [the further codes *nearby devices* and *nearby objects* were not assigned to any of the 38 quantitative participants]).

T2: Mitigations matter (11 sub-codes of *COVID19 data access mitigations*).

T3: Questionable data (4 sub-codes of *validity of phone data: No*).

T4: It's risky for me and others (16 sub-codes of *risks from data access*).

Code counts were not considered a useful measure because matching phrases were tagged in the data rather than individual words, and thus counts depend very much on verbosity and the sentence structure used by individuals. Instead, code coverage by participant was used to make a comparison by participant by group. The resultant analysis is presented visually in Figure 3 where each row of numbers represents code coverage (%). Greyscale-shading, from light grey (0%) to dark grey (40%), is used to highlight extremes and similar numbers. Blank values indicate relevant codes were not assigned in any responses.
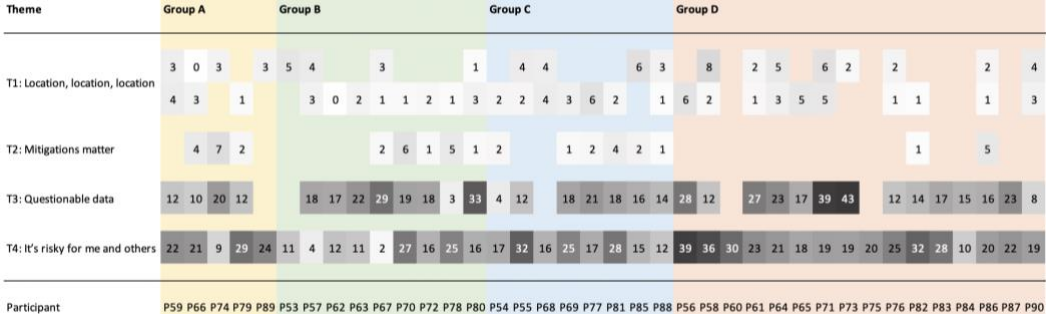
Theme	Group A	Group B	Group C	Group D
T1: Location, location, location (row 1)	3 0 3 · 3	5 4 · 3 · 1 · 4 4	6 3	8 · 2 5 · 6 2 · 2 · 2 · 4
T1: Location, location, location (row 2)	4 3 · 1	3 0 2 1 1 2 1 3	2 2 4 3 6 2 · 1 6 2	1 3 5 5 · 1 1 · 1 · 3
T2: Mitigations matter	4 7 2	2 6 1 5 1 2	1 2 4 2 1	1 · 5
T3: Questionable data	12 10 20 12	18 17 22 29 19 18 3 33 4	12 18 21 18 16 14 28 12	27 23 17 39 43 · 12 14 17 15 16 23 8
T4: It's risky for me and others	22 21 9 29 24	11 4 12 11 2 27 16 25 16	17 32 16 25 17 28 15 12	39 36 30 23 21 18 19 19 20 25 32 28 10 20 22 19
Participant	P59 P66 P74 P79 P89	P53 P57 P62 P63 P67 P70 P72 P78 P80	P54 P55 P68 P69 P77 P81 P85 P88	P56 P58 P60 P61 P64 P65 P71 P73 P75 P76 P82 P83 P84 P86 P87 P90

Figure 3: Stage 1 analysis code coverage (%) by theme related code categories (two for T1) for each participant by group.

For T1, group D is distinctive being the only group to include participants (5 out of 16) who did not mention any form of location data in their responses. Those in group D mention mitigations (T2) less than those in other groups. In terms of percentage coverage, all groups mention questionability of personal data (T3 e.g.: quality, relevance), these percentage numbers are generally lower (indicated by lighter grey shading) for those in group A. In terms of risks (T4), there are fears across all groups but those in group D mention risk more (higher number percentages/darker grey shading). Those not mentioning mitigations seem to be those who are most against any sharing of data, and perhaps would not countenance it; location data is mentioned less by these participants too.

Thus, group A is more different to the other three groups for T3, and group D is also different to the others for T1, T2 and T4, which is consistent with the sets of groups identified during the quantitative analysis (Table 6).

4.3.3 Archetypal user-groups

The examination of qualitative data for the four quantitative-derived clusters revealed some consistencies, permitting greater insight into their values and concerns. In Figure 4 we present a diagrammatic overview illustrating the overlaps between the cluster segmentation variables and *civic engagement*, as well as *privacy concern* (RQ3, H3 and H4), using I2-10 to I2-14 and supported by the findings from Stage 3:
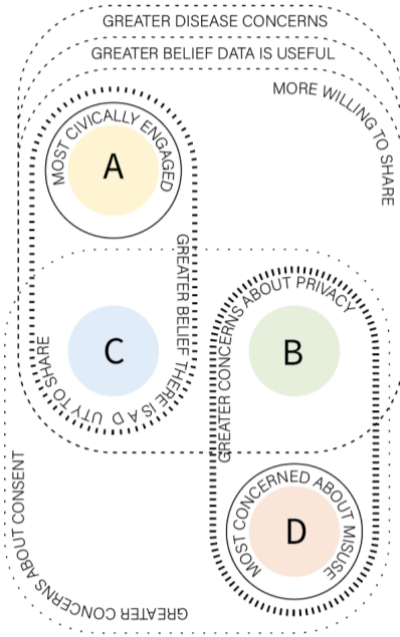
Figure 4: Venn diagram showing simple overview of the broad tendencies of the four groups.

Given we found some alignment between the Stage 1 and Stage 2 analyses, we continued our Stage 3 analysis by examining participants' responses to the two open-ended COVID-19 questions, to find whether we could gather additional meaning to the group segmentations concerning data collection for public health purposes (RQ1). Supporting more general findings elsewhere [53], our resultant groups could be used to help define characteristic concerns and methods how to speak to those, to assist the design of systems that may be designed for sharing or collection of personal data during public health emergencies, to contribute to a better understanding of people's concerns, to take them seriously and to address them.

Using grouped responses, together with findings from the previous stages of our analysis, we developed key phrases, summaries and selected participants' quotations for each cluster, to be indicative of their participants' values and concerns rather than to define the groups completely. We have also included a short statement of how design processes might speak to their concerns and preferences (addressing their mental models [49]).

Group A: Civically engaged and concerned about disease prevention

Group A are more civically engaged and are concerned about contagious diseases believing during public health emergencies there is a duty to share data, consent is less of a concern, and data might help both the individual and the wider public, for both influenza and coronavirus.

"I would be fine with that" (P59)

"done properly just to check out who had been near who for contact tracing that might be okay" (P79)

For design processes: Raise aspects of duty and highlight the contributions someone is making to society. Explain how data is analysed for any particular health purposes.

Group B: Sensitive to particular health-care applications and most concerned about privacy

Group B are the most concerned with privacy, do not believe there is any duty to share data, and hold a strong view that individual consent matters, even though they believe the data could be helpful depending on the situation and are generally more willing to share their own data.

"Should be an opt-out, but if the date was annonymous and provide recent location to help alert people that they maybe infected, then I approve" (P67)
"It would be invading someones privacy, it should be optional" (P80)

For design processes: Focus on methods that are privacy preserving and clearly communicate processes.

Group C: Willing to share and believe data will be helpful, but consent is important

Group C believe there is a duty to share data but strongly think that consent should not be overruled, even though such data might help the wider public, for both influenza and coronavirus public health emergencies, they also have some concerns about its misuse.

"They should have to gain consent first" (P54)
"I think permission should be sought" (P69)

For design processes: Make obtaining consent a key step, and also keeping this updated dynamically rather than a one-time decision, as it is often presented.

Group D: Privacy law matters, the data is not useful and very wary of misuse

Group D have concerns about privacy and highlight the importance of legal safeguards rather than extra mitigations. They worry a lot about data misuse, being less willing to share data now and do not think the data would be helpful in any case.

"i would think that is a breach of data protection laws" (P56)
"Breach of our rights as UK citizens" (P90)

For design processes: Apart from explaining the privacy-preserving aspects also state explicitly how this is staying within the frame of the law. Accept that not all users may agree to share data, consider strictly local processing or offering benefits without requiring processing of each users' data.

The reanalysis of the qualitative data using the clusters as lenses suggest the following Stage 3 (I3-) implications:

I3-1   There is a core of people who are yet to be convinced about data being used for public health purposes, for whom the risks outweigh the any form of participation and who are much less likely to share their data despite any amount of mitigations.

I3-2   Participants of all group affiliations feel there are many risks involved with sharing data.

I3-3   Privacy risk mitigations are more important to those who would consider sharing their device data.

I3-4   People cannot easily be defined with simplistic independent characteristics as there are interactions between privacy, lawfulness, consent, mitigations, and sense of duty which arise in different ways.

4.3.4   Discussion of Stage 3

The participants were very much aware of the risks involved in sharing their data to collaborate with the state combating public health emergencies (I3-2), and those who are more willing to share want to know appropriate mitigations are in place to reduce these (I3-3). But there is a significant group of people who are much less likely to cooperate with the state in this way (I3-1) regardless of what measures are put in place to counter the risks they perceive. People cannot be addressed by their stance

on isolated parameters (I3-4), and therefore considering multiple aspects holistically may be required to develop ways to empower them to make informed choices that support their concerns and values.

Although our 38 participants are not a representative sample of the UK population, particularly since they were recruited through an online system to undertake the survey, the findings indicate there is a core of people who are yet to be convinced about data being used for public health purposes, for whom the risks outweigh the any form of participation and who are much less likely to share their data despite any amount of mitigations. These groups, their summary properties and relative sizes are not meant to represent the wider population but are proposed as design tools to help grasp the tensions around the subject forming part of the analysis technique.

5   DISCUSSION

Having utilised the Stage 2 quantitative analysis to build on the analysis of the qualitative findings, our overall findings highlight tensions between privacy, lawfulness, consent, mitigations, and sense of duty. The design of digital interventions to undertake or facilitate sharing of personal data during public health emergencies needs to consider how such technology is perceived by citizens and how shared personal information is used (supporting Nissenbaum [62]) so that they are well-informed to make their own choices.

5.1 Research questions

5.1.1 RQ1: How do people react to consented/unconsented mobile phone (or phone mediated) data collection for public health purposes?

Our combination of qualitative and quantitative analysis found participants wanted to understand the purposes and uses of personal data collection and had concerns about what might be revealed (I1-8 and I1-9). Although some participants felt there was a duty to share such data, many do so reluctantly and with concerns about its quality (I1-6) and whether it would be useful (I1-5), privacy (I1-3) and possible future misuse (T4). These concerns mirrored findings from the previous foundational study [93] which had used the focus group method but with a much smaller number of participants. In the unlikely case of forced extraction of data against the wishes and consent of the individual, a wide range of reactions existed (see Section 4.3.1), although some participants felt (I2-9) access without consent is more reasonable if the data might possibly be useful for a public health situation, but very few thought it should be allowed without any other mitigating controls (I1-4). Participants also expressed an interest in ongoing, rather than one-time, control over consent recognising that data value and acceptable use change over time. There is a core of people who are yet to be convinced about data being used for public health purposes no matter what the imperative, for whom the risks outweigh the any form of participation and who are much less likely to share their data despite any amount of mitigations (I3-1). Public debate at the time of our data collection had possibly shifted the concern from "about my body" to "where is my body" (I1-1 and I1-2) along with a wider recognition that mobile data can relate to or belong to others (I1-7 and I1-9).

5.1.2 RQ2 How do the reactions/values depend on scenarios with differing severity?

Participants neither agreed nor disagreed there is a duty to share mobile data (I1-3 and I2-3), and such sense of duty was lower still for the case of non-communicable disease. Similarly, participants tended to rather disagree that data on a smartphone would assist other people's treatment and/or recovery, although this increased gradually to slightly agree for a pandemic situation like influenza (I1-5 and I2-4). Participants believed requiring consent is most important for non-contagious disease situations (I1-3), but even for the most contagious conditions they disagreed with access contrary to an individual's choice (I2-5). Overall, we interpret our findings to indicate a dynamic towards being more accepting of data collection and processing with increasing "contagiousness", but it must be noted that acceptance does not reach high levels, in particular regarding access without clear consent. Mitigating consent is therefore crucial.

5.1.3   RQ3 In what way do traits/positions around general values of privacy and civic engagement influence the reactions/values?

The more privacy concerned (I2-6) participants had greater concerns with data being accessed contrary to an individual's choice, for a non-contagious disease or a pandemic relating to a contagious disease, but not for communicable diseases in general. From cluster analysis, the most civically engaged group were the least concerned (I2-12) about the NHS gaining access to personal data without consent against an individual's wishes, and those with greater concerns about data misuse (I2-14) are also those least willing to share.

5.2   Directional hypotheses

We presented evidence (I2-3) that increasing contagiousness correlates to increasing willingness / obligation to share personal data (H1) and that the more contagious, possibly pandemic, scenarios can lead to significantly increased acceptance (I2-5) that personal data could be taken and used contrary to the wishes of individuals (H2).

Countering initial expectations, we did not confirm clear relationships that the greater the concerns people have about privacy are, the less willing they are to share personal data (H3). Nor did we confirm that the more civically engaged people are, the more willing they are to share personal data (H4). Additionally, our cluster analysis suggested that neither privacy concern (I2-10) nor civic engagement (I2-11) are unilaterally related to willingness to share, and a belief there is a duty to share (I2-13) does not appear strongly related to concerns about consent. However, our clustering technique suggests such associations may well not be clear conditionals depending on the interplay of just one or two variables.

5.3   Grouped qualitative analysis

Quantitative clustering techniques have been used to create user-groups and personas [58,78,86,98]. Our work expands on this to a) apply these techniques in the particular situation of attitudes and values to the access of personal digital-device data for public health purposes, and to b) progress the analysis further and use the generated groups to re-explore the qualitative data to define and describe the user-groups. Through our use of escalating scenario *severity*, in terms of public health significance, we identified which variables were most appropriate to include in the clustering analysis. The clusters were then used as lenses to re-examine the qualitative findings and draw additional insights. We propose that the cumulative effect of this three-stage analysis provides a method for deeper analysis of the collected data providing differentiated understandings when seeking group-based perspectives.

5.4   Potential negative impacts of the research

Informed cooperation can be a benefit both to individuals and wider society public health emergencies. With ongoing state-imposed restrictions during the pandemic and increased deployments of digital technologies, there are risks that insights from our research are used against the interests of the domestic population to support increased surveillance through greater acceptance of sharing health data from mobile devices during public health emergencies. Our intentions are not to aid coercion, but rather to contribute how to design for informing people so they can better make their own choices, and that digital technologies are designed in a way that acknowledges and supports the range of human concerns and values. Understanding tensions and mitigations around both consented and unconsented data collection and access is relevant given how, for example, different national contact tracing programmes are taking different approaches in this space that affect large populations at scale.

5.5 Limitations

We acknowledge that the analysis method used is exploratory and that our findings relate to a relatively small sample size. Our findings should therefore be considered to be formative and indicative. Our participants were not a representative sample of the UK population, and we recognise the particular political and social circumstances at the time of the survey cannot be reproduced and will be different in other nations and times. Demographic questions were asked at the start of the survey which could lead to the presence of stereotype biases. The deduced tensions and values, as well as user clusters, are yet to be validated with stakeholders and through active application in design and development processes.

5.6 Future work

Given the ebb and flow of lockdown restrictions and changing public debate, repeating the same survey using the same methods can provide insight into people's values and concerns, in the United Kingdom, and with citizens from other jurisdictions. Our analysis approach could be developed and opportunities to validate the technique considered, possibly through a statistical approach such as examining within-cluster variability of parameters as an indicator of generalisability. Other future work can focus on exploring the application of the suggested implications in the interventional context of application development and deployments. For example, regarding T3 *questionable data*, (see Section 4.1.3), practical approaches towards deriving clear ways of describing / visualising data processing pathways to improve understanding (e.g.: the concept of Bluetooth beacons) remain to be explored. In terms of the user groups as a potential tool, further work might put these into action in further development and then re-identify users from these groups and see whether the features derived from these groups, actually appeal to these groups.

6 CONCLUSION

The development and deployment at scale of devices and applications that collect and/or process considerable health-relevant data sourced from consumer technology, for citizens to work with state authorities responding to public health emergencies, remains both an ongoing trend of increasing relevance, as well as a current hot-topic in public debate, especially regarding contact tracing applications as a side-effect of the ongoing COVID-19 pandemic. With our three-stage analysis of a comparative scenario-based online study we contribute to the understanding of values, tensions and possible mitigation strategies with a focus on stances on data collection, privacy and brokering consent in this space.

Our thematic analysis contributes considerations around *location*, *mitigating consent*, *data reliability* and *privacy risks for oneself and others* based on the participant responses. Our quantitative analysis highlights how different health-care application scenarios have a measurable impact on the willingness to share data and how the complexity of considerations in the space likely prevent straight-forward mapping from personality traits regarding *privacy consciousness* or *civic engagement* to a *willingness to share data*. However, we hope that the quantitatively identified user clusters that have been qualitatively related to being groups with expressed values around being A) *civically engaged and concerned about disease prevention,* being B) *sensitive to particular health-care applications and most concerned about privacy*, being C) *willing to share and believing data will be helpful, but that consent is important*, and finding that D) *privacy law matters, the data is not useful and being very wary of misuse* respectively can serve both to inform further research, as well as to provide creative guidance on relevant considerations for ongoing and upcoming design and development processes in this area.

Data and and figures for this research are openly available at https://doi.org/10.25405/data.ncl.c.5530326

REFERENCES

[1]   Johannes Abeler, Matthias Bäcker, Ulf Buermeyer, and Hannah Zillessen. 2020. COVID-19 Contact Tracing and Data Protection Can Go Together. *JMIR mHealth and uHealth* 8, 4: e19359. https://doi.org/10.2196/19359

[2]   Ada Lovelace Institute. 2020. COVID-19 digital contact tracing tracker. Retrieved December 22, 2020 from https://www.adalovelaceinstitute.org/project/covid-19-digital-contact-tracing-tracker/

[3]   Ada Lovelace Institute. 2020. *COVID-19 Rapid Evidence Review: Exit through the App Store?* Ada Lovelace Institute. Retrieved September 16, 2020 from https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19-exit-through-the-app-store/

[4]   Richard P. Adler and Judy Goggin. 2005. What Do We Mean By "Civic Engagement"? *Journal of Transformative Education* 3, 3: 236–253. https://doi.org/10.1177/1541344605276792

[5]   Pär J Ågerfalk. 2013. Embracing diversity through mixed methods research. *European Journal of Information Systems* 22, 3: 251–256. https://doi.org/10.1057/ejis.2013.6

[6]   S. Attipoe-Dorcoo, R. Delgado, A. Gupta, J. Bennet, N.E. Oriol, and S.H. Jain. 2020. Mobile health clinic model in the COVID-19 pandemic: Lessons learned and opportunities for policy changes and innovation. *International Journal for Equity in Health* 19, 1. https://doi.org/10.1186/s12939-020-01175-7

[7]   Vaibhav Bajpai, Mirja Kühlewind, Jörg Ott, J. Schoenwaelder, Anna Sperotto, and Brian Trammell. 2017. Challenges with Reproducibility. 1–4. https://doi.org/10.1145/3097766.3097767

[8]   Masooda Bashir, Carol Hayes, April D. Lambert, and Jay P. Kesan. 2015. Online Privacy and Informed Consent: The Dilemma of Information Asymmetry. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (ASIST '15).

[9]   Alexander Bernier and Bartha Knoppers. 2020. Pandemics, privacy, and public health research. *Canadian Journal of Public Health* 111. https://doi.org/10.17269/s41997-020-00368-5

[10]   Chantal Bernier, Liane Fong, and Timothy M Banks. 2015. Pandemics in a Connected World: Integrating Privacy with Public Health Surveillance. *University of New Brunswick Law Journal* 66: 117–136.

[11]   Ann Blandford. 2019. HCI for health and wellbeing: Challenges and opportunities. *International Journal of Human-Computer Studies* 131: 41–51. https://doi.org/10.1016/j.ijhcs.2019.06.007

[12]   Braun, Virginia and Clarke, Victoria. 2013. *Successful Qualitative Research*. Sage Publications Ltd. Retrieved November 3, 2018 from https://uk.sagepub.com/en-gb/eur/successful-qualitative-research/book233059

[13]   Jon Brickey, Steven Walczak, and Tony Burgess. 2010. A Comparative Analysis of Persona Clustering Methods. In *AMCIS*. Retrieved from https://aisel.aisnet.org/amcis2010/217/

[14]   K. Caine and R. Hanania. 2013. Patients want granular privacy control over health information in electronic medical records. *Journal of the American Medical Informatics Association* 20, 1: 7–15. https://doi.org/10.1136/amiajnl-2012-001023

[15]   Pascale Carayon, Sarah Kianfar, Yaqiong Li, Anping Xie, Bashar Alyousef, and Abigail Wooldridge. 2015. A systematic review of mixed methods research on human factors and ergonomics in health care. *Applied ergonomics* 51: 291–321. https://doi.org/10.1016/j.apergo.2015.06.001

[16]   José M. Cecilia, Juan-Carlos Cano, Enrique Hernández-Orallo, Carlos T. Calafate, and Pietro Manzoni. 2020. Mobile crowdsensing approaches to address the COVID-19 pandemic in Spain. *IET Smart Cities* 2, 2: 58–63. https://doi.org/10.1049/iet-smc.2020.0037

[17]   Juliana Chen, Adrian Bauman, and Margaret Allman-Farinelli. 2016. A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research. *Telemedicine and e-Health* 22, 8: 655–665. https://doi.org/10.1089/tmj.2015.0159

[18]   Hyunghoon Cho, Daphne Ippolito, and William Yu Yun. 2020. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. *arXiv.org*. Retrieved from https://arxiv.org/abs/2003.11511

[19]   Andrew Clarke and Robert Steele. 2015. Smartphone-based public health information systems: Anonymity, privacy and intervention. *Journal of the Association for Information Science and Technology* 66, 12: 2596–2608. https://doi.org/10.1002/asi.23356

[20]   Clarke, Victoria. 2017. What is thematic analysis, when is it useful, and what does "best practice" look like? Retrieved November 12, 2018 from https://www.youtube.com/watch?v=4voVhTiVydc

[21]   Oonagh Corrigan. 2003. Empty ethics: the problem with informed consent. *Sociology of Health & Illness* 25, 7: 768–792. https://doi.org/10.1046/j.1467-9566.2003.00369.x

[22]   John W Creswell. 2014. *A Concise Introduction to Mixed Methods Research*. SAGE Publications.

[23]   Department for Environment Food and Rural Affairs. 2020. Rural population 2014/15. *gov.uk*. Retrieved September 17, 2020 from https://www.gov.uk/government/publications/rural-population-and-migration/rural-population-201415

[24]   European Commission. Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection 2020/C 124 I/01E. Retrieved September 16, 2020 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020XC0417%2808%29

[25]   Daisy Fancourt, Andrew Steptoe, and Liam Wright. 2020. The Cummings effect: politics, trust, and behaviours during the COVID-19 pandemic. *The Lancet* 396, 10249: 464–465. https://doi.org/10.1016/S0140-6736(20)31690-1

[26]   Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368, 6491. https://doi.org/10.1126/science.abb6936

[27]   Gabriel Kaptchuk, Daniel G Goldstein, Eszter Hargittai, Jake M Hofman, and Elissa M Redmiles. 2020. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt (preprint). Retrieved from https://arxiv.org/pdf/2005.04343.pdf

[28]   M.-P. Gagnon, P. Ngangue, J. Payne-Gagnon, and M. Desmartis. 2016. M-Health adoption by healthcare professionals: A systematic review. *Journal of the American Medical Informatics Association* 23, 1: 212–220. https://doi.org/10.1093/jamia/ocv052

[29]   Urs Gasser, Marcello Ienca, James Scheibner, Joanna Sleigh, and Effy Vayena. 2020. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. *The Lancet Digital Health* 2, 8: e425–e434. https://doi.org/10.1016/S2589-7500(20)30137-0

[30]   GBD 2016 Meningitis Collaborators. 2018. Global, regional, and national burden of meningitis, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* 17, 12: 1061–1082. https://doi.org/10.1016/S1474-4422(18)30387-9

[31]   George Lakoff and Mark Johnson. 2008. *Metaphors We Live By*. University of Chicago Press.

[32]   Saira Ghafur, Jackie Van Dael, Melanie Leis, Ara Darzi, and Aziz Sheikh. 2020. Public perceptions on data sharing: key insights from the UK and the USA. *The Lancet Digital Health* 2, 9: e444–e446. https://doi.org/10.1016/S2589-7500(20)30161-8

[33]   Ivan J. Gotham, Linh H. Le, Debra L. Sottolano, and Kathryn J. Schmit. 2008. Public Health Preparedness Informatics Infrastructure. A Case Study in Integrated Surveillance and Response: 2004---2005 National Influenza Vaccine Shortage. In *Proceedings of the 2008 International*

Workshop   on   Biosurveillance   and   Biosecurity (BioSecure   '08),   42–55. https://doi.org/10.1007/978-3-540-89746-0_5

[34]   Christine Grady, Steven R. Cummings, Michael C. Rowbotham, Michael V. McConnell, Euan A. Ashley, and Gagandeep Kang. 2017. Informed Consent. *New England Journal of Medicine* 376, 9: 856–867. https://doi.org/10.1056/NEJMra1603773

[35]   Grant   Blank.   2019.   2019   Oxford   Internet   Survey   Questionnaire.   Retrieved   from https://oxis.oii.ox.ac.uk/wp-content/uploads/sites/43/2019/11/OxIS-Questionnaire-2019.pdf

[36]   Quinn Grundy, Kellia Chiu, Fabian Held, Andrea Continella, Lisa Bero, and Ralph Holz. 2019. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. *BMJ* 364. https://doi.org/10.1136/bmj.l920

[37]   Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I. Bosse, Christopher I. Jarvis, Timothy W. Russell, James D. Munday, Adam J. Kucharski, W. John Edmunds, Sebastian Funk, and Rosalind M. Eggo. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet. Global health* 8, 4: e488–e496. https://doi.org/10.1016/S2214-109X(20)30074-7

[38]   Burke JA, Deborah Estrin, Mark Hansen, A Parker, R.Anitha Nithya, S Reddy, and M Srivastava. 2006. Participatory Sensing. *Workshop on World-Sensor-Web (WSW): Mobile Device Centric Sensor Networks and Applications*.

[39]   Johannes Abeler, Sam Altmann, Luke Milsom, Severine Toussaert, and Hannah Zillessen. 2020. Support   in   the   UK   for   app-based   contact   tracing   of   COVID-19.   Retrieved   from https://osf.io/huqtr/

[40]   John M Last (ed.). 2001. *A dictionary of epidemiology*. Oxford University Press.

[41]   M.N. Kamel Boulos and E.M. Geraghty. 2020. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics* 19, 1. https://doi.org/10.1186/s12942-020-00202-8

[42]   Mark Keith, Jeffry Babb, Paul Lowry, Christopher Furner, and Amjad Abdullat. 2016. Limited Information and Quick Decisions: Consumer Privacy Calculus for Mobile Applications. *AIS Transactions on HCI* Forthcoming. https://doi.org/10.17705/1thci.00081

[43]   Don Klinkenberg, Christophe Fraser, and Hans Heesterbeek. 2006. The Effectiveness of Contact Tracing   in   Emerging   Epidemics.   *PLOS   ONE*   1,   1:   1–7. https://doi.org/10.1371/journal.pone.0000012

[44]   Patty Kostkova. 2015. Grand Challenges in Digital Health. *Frontiers in Public Health* 3. https://doi.org/10.3389/fpubh.2015.00134

[45]   T. Lang. 2011. Advancing global health research through digital technology and sharing data. *Science* 331, 6018: 714–717. https://doi.org/10.1126/science.1199349

[46]   Jimmie Leppink. 2017. Revisiting the quantitative–qualitative-mixed methods labels: Research questions, developments, and the need for replication. *Journal of Taibah University Medical Sciences* 12, 2: 97–101. https://doi.org/10.1016/j.jtumed.2016.11.008

[47]   David Lester, Bijou Lester, and Simon James. 2005. A short computer anxiety scale. *Perceptual and motor skills* 100: 964–8. https://doi.org/10.2466/PMS.100.3.964-968

[48]   J. Li, N. Moore, S. Akter, S. Bleisten, and P. Ray. 2010. mHealth for influenza pandemic surveillance in developing countries. https://doi.org/10.1109/HICSS.2010.274

[49]   Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (UbiComp '12), 501–510. https://doi.org/10.1145/2370216.2370290

[50]   Xiaoping Liu, Xiao-Bai Li, Luvai Motiwalla, Wenjun Li, Hua Zheng, and Patricia D. Franklin. 2016. Preserving Patient Privacy When Sharing Same-Disease Data. *J. Data and Information Quality* 7, 4. https://doi.org/10.1145/2956554

[51] Lloyds Bank. 2020. Essential Digital Skills (EDS) data. Retrieved September 17, 2020 from https://www.lloydsbank.com/banking-with-us/whats-happening/consumer-digital-index/essential-digital-skills.html

[52] Sumit Majumder and M. Jamal Deen. 2019. Smartphone Sensors for Health Monitoring and Diagnosis. *Sensors* 19, 9: 2164. https://doi.org/10.3390/s19092164

[53] Kirsten Martin and Katie Shilton. 2016. Why experience matters to privacy: How context-based experience moderates consumer privacy expectations for mobile applications. *Journal of the Association for Information Science and Technology* 67, 8: 1871–1882. https://doi.org/10.1002/asi.23500

[54] B. Martínez-Pérez, I. de la Torre-Díez, and M. López-Coronado. 2015. Privacy and Security in Mobile Health Apps: A Review and Recommendations. *Journal of Medical Systems* 39, 1. https://doi.org/10.1007/s10916-014-0181-3

[55] Matthew Guariglia and Adam Schwartz. 2020. Protecting Civil Liberties During a Public Health Crisis. *Electronic Frontier Foundation*. Retrieved from https://www.eff.org/deeplinks/2020/03/protecting-civil-liberties-during-public-health-crisis

[56] Deven Mcgraw, Robert Belfort, Helen Pfister, and Susan Ingargiola. 2015. Engaging Patients While Addressing Their Privacy Concerns: The Experience of Project HealthDesign. *Personal Ubiquitous Comput.* 19, 1: 85–89. https://doi.org/10.1007/s00779-014-0809-9

[57] Mhairi Aitken, Sarah Cunningham-Burley, Amy Darlington, Stephen Elstub, Oliver Escobar, Kerina Jones, Nayha Sethi, and Rachel Thompson. 2020. Letters to the editor: Why the public need a say in how patient data are Used for Covid-19 Responses. *International Journal of Population Data Science* 5, 2. https://doi.org/10.23889/ijpds.v5i2.1357

[58] Tomasz Miaskiewicz, Tamara Sumner, and Kenneth Kozar. 2008. A Latent Semantic Analysis methodology for the identification and creation of personas. In *Conference on Human Factors in Computing Systems - Proceedings*, 1501–1510. https://doi.org/10.1145/1357054.1357290

[59] Elizabeth L. Murnane, David Huffaker, and Gueorgi Kossinets. 2015. Mobile health apps: adoption, adherence, and abandonment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (UbiComp/ISWC'15 Adjunct), 261–264. https://doi.org/10.1145/2800835.2800943

[60] Neil M Ferguson, David Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick GT Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Retrieved from https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf

[61] Helen Nissenbaum. 2004. Privacy As Contextual Integrity. *Washington Law Review* 79.

[62] Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books.

[63] Nuffield Council on Bioethics. 2020. Research in global health emergencies: ethical issues. Retrieved from https://www.nuffieldbioethics.org/publications/research-in-global-health-emergencies

[64] Office for National Statistics. 2017. Population by age, gender and ethnicity. *ons.gov.uk*. Retrieved August 10, 2020 from https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/populationbyagegenderandethnicity

[65] Office for National Statistics. 2020. Unity and division in Great Britain: 24 April to 28 June 2020. Retrieved   from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/unityanddivisi oningreatbritain/24aprilto28june2020

[66] Aisling Ann O'Kane, Helena Mentis, and Eno Thereska. 2013. Non-static nature of patient consent: Shifting privacy perspectives in health information sharing. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 553–562. https://doi.org/10.1145/2441776.2441838

[67] Open Data Institute. 2020. *Covid-19: Identifying and managing ethical issues around data*. Open Data Institute. Retrieved September 16, 2020 from https://theodi.org/article/covid-19-identifying-and-managing-ethical-issues-around-data/

[68] World Health Organization. 2019. Global influenza strategy 2019-2030.

[69] Chrysanthi Papoutsi, Julie E. Reed, Cicely Marston, Ruth Lewis, Azeem Majeed, and Derek Bell. 2015. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. *BMC medical informatics and decision making* 15: 86. https://doi.org/10.1186/s12911-015-0202-2

[70] Michael J Parker, Christophe Fraser, Lucie Abeler-Dörner, and David Bonsall. 2020. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *Journal of Medical Ethics* 46, 7: 427. https://doi.org/10.1136/medethics-2020-106314

[71] Alexandra L. Phelan. 2020. COVID-19 immunity passports and vaccination certificates: scientific, equitable, and legal challenges. *The Lancet* 395, 10237: 1595–1598. https://doi.org/10.1016/S0140-6736(20)31034-5

[72] Philipp K Masur. 2018. *Situational privacy and self-disclosure: Communication processes in online environments*. Springer, Cham, Switzerland.

[73] Miloslava Plachkinova, Steven Andrés, and Samir Chatterjee. 2015. A Taxonomy of mHealth Apps – Security and Privacy Concerns. In *2015 48th Hawaii International Conference on System Sciences*, 3187–3196. https://doi.org/10.1109/HICSS.2015.385

[74] John Pruitt and Jonathan Grudin. 2003. Personas: Practice and Theory. In *Proceedings of the 2003 Conference on Designing for User Experiences* (DUX '03), 1–15. https://doi.org/10.1145/997078.997089

[75] Elissa M. Redmiles. 2020. User Concerns & Tradeoffs in Technology-Facilitated Contact Tracing. *arXiv:2004.13219 [cs]*. Retrieved September 15, 2020 from http://arxiv.org/abs/2004.13219

[76] Robert A Fahey and Airo Hino. 2020. COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management* 102181. Advance online publication. https://dx.doi.org/10.1016%2Fj.ijinfomgt.2020.102181

[77] Robert Hinch, Will Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferreti, Michael Parker, Areas Meroueh, Bryn Mathias, Scott Stevenson, Daniel Montero, James Warren, Nicole K Mather, Anthony Finkelstein, Lucie Abeler-Dorner, David Bonsall, and Christophe Fraser. 2020. Effective Configurations of a Digital Contact Tracing App: A report to NHSX. Retrieved from https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217

[78] Joni Salminen, Kathleen Guan, Soon-Gyo Jung, Shammur A. Chowdhury, and Bernard J. Jansen. 2020. A Literature Review of Quantitative Persona Creation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 1–14. https://doi.org/10.1145/3313831.3376502

[79] Gabriel Scally, Bobbie Jacobson, and Kamran Abbasi. 2020. The UK's public health response to covid-19. *BMJ* 369. https://doi.org/10.1136/bmj.m1932

[80] Abhinav Sharma, Robert A. Harrington, Mark B. McClellan, Mintu P. Turakhia, Zubin J. Eapen, Steven Steinhubl, James R. Mault, Maulik D. Majmudar, Lothar Roessig, Karen J. Chandross, Eric

M. Green, Bakul Patel, Andrew Hamer, Jeffrey Olgin, John S. Rumsfeld, Matthew T. Roe, and Eric D. Peterson. 2018. Using Digital Health Technology to Better Generate Evidence and Deliver Evidence-Based Care. *Journal of the American College of Cardiology* 71, 23: 2680–2690. https://doi.org/10.1016/j.jacc.2018.03.523

[81] T. Sharon. 2016. The googlization of health research: from disruptive innovation to disruptive ethics. *Personalized Medicine* 13, 6: 563–574. https://doi.org/10.2217/pme-2016-0057

[82] Shaun Gallagher. 2005. *How the Body Shapes the Mind*. Oxford University Press, Oxford, England.

[83] Jeremy Shiffman. 2009. A Social Explanation for the Rise and Fall of Global Health Issues. *Bulletin of the World Health Organization* 87: 608–13. https://doi.org/10.2471/BLT.08.060749

[84] Ben Shneiderman and Catherine Plaisant. 2006. Strategies for Evaluating Information Visualization Tools: Multi-Dimensional in-Depth Long-Term Case Studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (BELIV '06), 1–7. https://doi.org/10.1145/1168149.1168158

[85] Lucy Simko, Ryan Calo, Franziska Roesner, and Tadayoshi Kohno. 2020. COVID-19 Contact Tracing and Privacy: Studying Opinion and Preferences. *arXiv:2005.06056 [cs]*. Retrieved September 16, 2020 from http://arxiv.org/abs/2005.06056

[86] Rashmi Sinha. 2003. Persona Development for Information-Rich Domains. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '03), 830–831. https://doi.org/10.1145/765891.766017

[87] The Scottish Council for Voluntary Organisations. 2018. Essential Digital Skills Toolkit. Retrieved from https://scvo.org.uk/support/digital/participation/skills

[88] Peter Tolmie and Andy Crabtree. 2018. The practical politics of sharing personal data. *Personal and Ubiquitous Computing* 22, 2: 293–315. https://doi.org/10.1007/s00779-017-1071-8

[89] UK Government. 2019. Essential digital skills framework. *GOV.UK*. Retrieved September 16, 2020 from https://www.gov.uk/government/publications/essential-digital-skills-framework

[90] Koen Van Turnhout, Arthur Bennis, Sabine Craenmehr, Robert Holwerda, Marjolein Jacobs, Ralph Niels, Lambert Zaad, Stijn Hoppenbrouwers, Dick Lenior, and René Bakker. 2014. Design Patterns for Mixed-Method Research in HCI. https://doi.org/10.13140/2.1.4701.2643

[91] Sandeep Kumar Vashist, E. Marion Schneider, and John H. T. Luong. 2014. Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management. *Diagnostics* 4, 3: 104–128. https://doi.org/10.3390/diagnostics4030104

[92] Serge Vaudenay. 2020. *Centralized or Decentralized? The Contact Tracing Dilemma*. Retrieved from https://eprint.iacr.org/2020/531

[93] Colin Watson and Jan David Smeddinck. 2020. Unconsented Data Transfusions: Attitudes towards Extracting Personal Device Data for Public Health Emergencies. In *Proceedings of the Conference on Mensch Und Computer* (MuC '20), 205–209. https://doi.org/10.1145/3404983.3409994

[94] Wellcome Trust. 2018. Understanding Patient Data, Public Attitudes to Patient Data Use. Retrieved from https://understandingpatientdata.org.uk/sites/default/files/2018-08/Public%20attitudes%20key%20themes_0.pdf

[95] Sera Whitelaw, Mamas A. Mamas, Eric Topol, and Harriette G. C. Van Spall. 2020. Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health* 2, 8: e435–e440. https://doi.org/10.1016/S2589-7500(20)30142-4

[96] Robyn Whittaker. 2012. Issues in mHealth: Findings From Key Informant Interviews. *Journal of Medical Internet Research* 14, 5: e129. https://doi.org/10.2196/jmir.1989

[97] World Health Organization. 2020. Coronavirus disease (COVID-19): Situation report 112. Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200511-covid-19-sitrep-112.pdf?sfvrsn=813f2669_2

[98] Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-Driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the 2016*

CHI Conference on Human Factors in Computing Systems (CHI '16), 5350–5359. https://doi.org/10.1145/2858036.2858523

[99]   H. Zou, Y. Shu, and T. Feng. 2020. How Shenzhen, China avoided widespread community transmission: A potential model for successful prevention and control of COVID-19. *Infectious Diseases of Poverty* 9, 1. https://doi.org/10.1186/s40249-020-00714-2

[100]   2020. Contact tracing in the context of COVID-19 (Report ref WHO/2019-nCoV/Contact_Tracing/2020.1). Retrieved from https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19

[101]   Unified research on privacy-preserving contact tracing and exposure notification. *Google Docs*. Retrieved September 16, 2020 from https://docs.google.com/document/d/16Kh4_Q_tmyRh0-v452wiul9oQAiTRj8AdZ5vcOJum9Y/edit?usp=embed_facebook