Does Mode of Digital Contact Tracing Affect User Willingness to Share Information? A Quantitative Study

Camellia Zakaria
College of Information & Computer
Sciences, University of Massachusetts
Amherst
USA

Pin Sym Foong
Saw Swee Hock School of Public
Health, National University of
Singapore
Singapore

Chang Siang Lim
Future Health Technologies,
Singapore ETH Center
Singapore

Pavithren V. S. Pakianathan*
Saw Swee Hock School of Public
Health, National University of
Singapore
Singapore

Gerald Koh Choon Huat
Saw Swee Hock School of Public
Health, National University of
Singapore
Singapore

Simon Tangi Perrault
Singapore University of Technology
and Design
Singapore

ABSTRACT

Digital contact tracing can limit the spread of infectious diseases. Nevertheless, barriers remain to attain sufficient adoption. In this study, we investigate how willingness to participate in contact tracing is affected by two critical factors: the modes of data collection and the type of data collected. We conducted a scenario-based survey study among 220 respondents in the United States (U.S.) to understand their perceptions about contact tracing associated with automated and manual contact tracing methods. The findings indicate a promising use of smartphones and a combination of public health officials and medical health records as information sources. Through a quantitative analysis, we describe how different modalities and individual demographic factors may affect user compliance when participants are asked to provide four key information pieces for contact tracing.

CCS CONCEPTS

• **Social and professional topics** → **User characteristics**; • **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

contact tracing, pandemic, willingness, trust, public health

*Also with Singapore University of Technology and Design.

1 INTRODUCTION

Contact tracing is a public health strategy that plays a crucial role in curbing the spread of communicable diseases [10]. Manual contact tracing involves a series of steps, starting with contacting infected persons and then interviewing the patient to gather a log of locations and persons with personal contact within a specified period. These close contacts are identified and notified regarding potential exposure and informed of further containment measures such as testing or isolation. However, an incomplete or incorrect recall of events, locations, and contact persons in the period of interest can deter this process [14]. Often, contact tracing is resource-intensive and time consuming [23].

In the U.S., the COVID-19 pandemic has particularly challenged the limits of manual contact tracing. The novel clinical features of COVID-19 infection, including long incubation period, asymptomatic transmission, and high transmission rate [65] led to a burgeoning number of patients that easily outpaced manual contact tracing efforts. These issues are exacerbated by less than 50% of people being willing to participate in contact tracing efforts [32].

A modeling study by Feretti *et al.* [14] shows that a hypothetical, fully automated digital contact tracing solution can slow or stop the transmission of COVID-19. However, the extent of data gathering that digital contact tracing requires leads to concerns with privacy, user surveillance, and data leaks [25, 45, 50], worsened by the lack of trust in government bodies and technology companies to handle sensitive information required in digital contact tracing [10, 57]. These concerns hamper the adoption of digital contact tracing and place more lives at risk. Unfortunately, a recent scoping review concludes that "there is a dearth of evidence regarding the barriers and facilitators to uptake and engagement with COVID-19 digital contact tracing applications" [61]. Thus it is essential to determine the factors that increase acceptance of digital contact tracing.

Existing studies prospectively probed the desired feature set to influence users' willingness to participate in digital contact tracing. Researchers have begun to identify a few key facilitators. These are: higher perceived public health benefit, perceived individual benefit,

and lower degree of privacy risk [22, 34]. Missing in this body of knowledge about the users' perception of digital contact tracing features is whether the stated *mode of digital contact tracing* may affect willingness to participate. The perception that a particular mode of delivery is more or less threatening to personal privacy may derail digital contact tracing efforts even before actual usage. While the particular features of an application affect a person's willingness to continue using an application, resistance to the mode of delivery can be a barrier to the *first usage*. To our knowledge, this question of the users' lens of digital contact tracing modes has not yet been studied.

In this paper, we sought to **understand how the mode of digital contact tracing affects users' willingness to share different types of data**. We conducted a scenario-based survey of U.S. respondents to understand users' relative willingness to share private information to help contact tracing and examine the modality of data sharing that would be most acceptable by users. We studied three types of information best collected by digital devices to support contact tracing. They are (1) the subject's identity (name and social security number), (2) the subject's contact details (email and contact number), and (3) details of exposure (location or person of interest). We also offered four types of modalities for participants to indicate their willingness to share this information. They are (1) in-person communication with public health officials, (2) providing access to their existing health records, (3) sharing their information collected in their smartphone device (4) providing access to analyze their internet activity. We also examined how such willingness was influenced by demographic variables such as age, parenthood, income, and trust in public health officials. As the study was conducted in the U.S. between November and December 2020, we wanted to ensure that users' responses were not solely based on their experience of the COVID-19 pandemic. Hence, we randomly assigned participants to a different potential epidemic scenario involving one of six infectious agents.

Through a quantitative analysis, we present an overview of users' willingness to share different types of private information to support contact tracing and factors that influence such willingness, supplemented with qualitative responses to help unpack our observations. We found no evidence of the effects of disease types on users' willingness to share their private information for digital contact tracing. Notably, our findings show that participants are most willing to share their private information via smartphones ($p < .01$) and grant public health officials access to individuals' health records ($p < .01$). However, sharing *location* information is significantly impacted by the level of trust in public health officials ($p < .01$). These results add new dimensions to the well-established importance of trust in public health officials in the data collection pipeline for digital contact tracing.

Our work contributes 1) an empirical study of user willingness to adopt contact tracing strategies across various modes of collection, information types, and disease types, and 2) based on these findings, recommendations toward developing a national digital contact tracing strategy in the U.S.

2 RELATED WORK

Our focus here was to summarize existing literature on barriers and facilitators to digital contact tracing. Specifically, we highlight prior work that examined barriers and facilitators of digital contact tracing and the importance of data collection modality.

A recent survey of Americans indicates that just 42% are willing to download and use a contact-tracing app [71]. For automated and partly automated contact tracing to be effective, at least 56% uptake is needed [6] within a population. What would the barriers and facilitators of digital contact tracing be for sufficient mass adoption in the U.S.?

2.1 Background on Contact Tracing

Contact tracing is an essential strategy in public health management, proving itself effective in limiting the spread of infectious disease [58]. This process, however, must be performed expeditiously. As defined by CDC, contact tracing involves trained case investigators by the public health organization to evaluate a patient's close contacts (i.e., person-of-interest, POI) [12]. A POI may expect a case investigator to seek information on their health symptoms and possible exposures (i.e., places, partners) to determine compromised entities during the interview. POI may also receive instructions for isolation and follow-up sessions on their health. Table 1 summarises the types of critical, but sensitive information that serves as a general guideline by CDC in conducting such interviews [9]. This laborious process can be slow [24] and often relies on the POI's willingness and ability to recall the relevant information [43]. The race against time is critically challenged by the shortage of human resources during a health crisis. Hence, more recently, digital tools were explored to play a role in enhancing contact tracing by either automating or semi-automating some of these data collection tasks.

	Questions	Information Type
1	What is their name?	Identity
2	What name do they go by?	
3	What is their gender?	
4	What is their race/ethnicity	
5	What is their primary language?	
6	What is the best way to reach them?	Contact
7	What is their cell number, email, APP and username?	
8	Where do they live?	Exposure Location
9	Who do they live with?	
10	Where do they work?	
11	Where is that located?	
12	When did you see them last?	Exposure Location / Intimate Partner
13	For how long have you spent time with them?	
14	What symptoms might have had?	Health Status
15	What underlying medical conditions might they have?	
16	What do they know about your infectious status?	

Table 1: A list of open-ended questions for contact tracing by the CDC, and grouped according to information type.

Does Mode of Digital Contact Tracing Affect User Willingness to Share Information?

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

Based on this established process, our study examines how answers to these questions can be collected through different data collection modalities and not restricted to human contact tracers alone [9].

2.2 Barriers and Facilitators of Digital Contact Tracing

Various barriers and facilitators to digital contact tracing have been identified. A discrete choice experiment prospective study in the Netherlands suggested less modifiable individual factors specific to COVID-19 [20]. In this study, the factors that correlated with predicted adoption rates were educational attainment, underlying health conditions, and a perceived threat from COVID-19.

More promising, modifiable facilitators can be gleaned with a broader literature review that includes other diseases. Megnin *et al.* [34] conducted a rapid review of qualitative studies on the factors influencing user uptake and engagement with any contact tracing system across various infectious diseases. They identified four modifiable factors that arise from the users' perception of the application. They are a perceived sense of collective responsibility, perceived personal benefit, the presence of community co-production of contact tracing systems, and the perceived capability of reaching contact persons efficiently and effectively. The authors also identified privacy concerns as a key barrier. These concerns were: mistrust with the requester, unmet needs for information and support, fear of stigmatization (due to an identified infection), and what they called "mode-specific challenges". These mode-specific challenges were not communication channel issues per se. They were more akin to accessibility issues (e.g., no smartphone) or usability issues (e.g., difficulty downloading, using the application, or lack of technical proficiency).

These studies suggest two broader research areas that can be brought to bear upon our knowledge about facilitators and barriers of digital contact tracing. The first area covers the perception of trust in public bodies running the digital contact tracing operations. The second area covers studies on multiple alternative channels of user participation and data collection modes.

2.3 Issues of Trust and Data Collection

One commonly cited factor that predicts the adoption of contact tracing is the trust in the entity conducting digital contact tracing (i.e., institutional trust). Mayer's body of work on institutional trust defines it as the following: 'the willingness of a party to be vulnerable to the actions of another party' [31]. In a pandemic, the willingness of users to be vulnerable must be particularly directed towards public health officials, who are providing practical guidelines and general awareness of the disease outbreak. Instead, trust is often referred to as being vulnerable to government bodies and private firms handling the sensitive information required in digital contact tracing.

To overcome the mountain of skepticism towards utilizing contact tracing apps, several public health organizations and researchers [11, 21, 25, 67] have proposed that these systems be transparent and open to public scrutiny. Scientists and researchers across the globe have also recommended that a privacy-by-design approach be adopted whereby only necessary data is collected and stored using secure encryption techniques to preserve the security and privacy of the data [25]. Additionally, the World Health Organization (WHO) and American Civil Liberties Union (ACLU) collectively agree with the CDC's suggestion to make contact tracing voluntary, with full user control over data management [11, 15, 67].

In response to these recommendations, Google and Apple pushed for anonymous contact tracing [16]. The exposure notification framework frequently exchanges anonymous identifier beacons through Bluetooth between and among smartphones whose users are in close proximity; either by installing an official app from their region's government or directly through a verification protocol from public health authorities. NOVID [40], ICheckedIn [59] and SaferMe [53] are similar efforts that followed in pursuit of anonymity with an encrypted framework. For example, users' names and numbers are hashed in a "pin," maintaining relative obscurity in businesses and places of visit [59]. While this information is set to expire in limited duration (e.g., discarded after 30 days), personal information remains accessible to relevant government bodies or pre-approved systems. As might be expected, these efforts did not alleviate the tension associated with the lack of confidence in government and private companies [1, 10, 66, 67].

In a study specific to the U.S., only 37% of respondents found it acceptable to share data with state and local officials, compared to 75% of the sample who preferred sharing data with infectious diseases researchers [32]. Yet, the National Academy for State Health Policy found that contact tracing efforts are either led by state or county (e.g., eight states, including California, are county-led) [38].

The suggestion to empower user autonomy in digital contact tracing places the onus on the users to be sufficiently convinced that the benefits of digital contact tracing outweigh the concerns. In a qualitative study on privacy concerns of a tool for online data collection, Phelan *et al.* [47] drew on dual process theories to describe two kinds of user concern about privacy: *intuitive concern* (when following a gut feeling) and *considered concern* (when weighing pros and cons). In *considered concern*, users regularly recognize the benefits of accepting a privacy intrusion, particularly when there is trust in the requester or low assessed risk. However, *intuitive concern* can override the more cognitive path of *considered concern*. This can happen, for example, when the social presence of the requester provides privacy assurance. Additionally, a study on user responses to surveillance suggests that when these types of concerns are not addressed, individuals may adopt protective coping strategies [54]. These avoidance mechanisms would make it harder to involve them in contact tracing efforts.

Overall the literature on trust indicates issues of trust seem to be centered on *perceptions* of intrusiveness of the tool. Although research in anonymous contact tracing actively pushes for computational and technological protocols that strive to guarantee user anonymity and tool credibility, it remains unclear how willing users are to share critical information that must fundamentally be disclosed to public health authorities upon identifying themselves as infected. Thus, it led us to query if the users' perception of the data collection modality may also affect this intuitive concern. Such knowledge could help pave the way to reducing resistance to privacy intrusions of digital contact tracing for public health.

2.4 Issues of Technology Access

Much research on developing the functions of digital contact tracing is centered on the types of sensing modalities available on a smartphone device. Specifically, investigations have looked at Bluetooth and Global Positioning System (GPS) as two standard sensing techniques utilized by most contact tracing mobile applications. Other investigations for collecting location-based data include using a magnetometer, QR code, phone logs, and the most recent WiFi logs of smartphones. [63, 66, 70].

However, the primary dependence on a smartphone device for data collection may have the disadvantage of disproportionately disenfranchising several demographic groups. In the U.S., the Pew Research Center reported 66% smartphone ownership among users aged 65 and above, and 71% ownership among lower-income earners (less than $30,000) [46]. In response to a similar situation of low smartphone ownership and complaints about data and power usage in some communities, Singapore's digital contact tracing efforts were supplemented by a dedicated hardware-based Bluetooth token. This self-contained token was distributed in pendant form to the entire population, mainly benefiting people who have reservations about using their smartphone for contact tracing or do not own one. The rollout was complex as it additionally required the distribution of scanning hardware [17] at an estimated cost of about USD 4.6 million [69]. Hence different digital contact tracing modalities present challenges to balance accessibility and privacy [11, 15, 50, 66].

Given the cost and complexity of digital contact tracing initiatives, it is important to understand how users perceive the modality being used. By selecting modalities that improve the likelihood of uptake, we may increase the chances of success for a digital contact tracing program.

2.5 Summary

Researchers have identified a series of individual user factors that promote the uptake and engagement with a digital contact tracing program. Additionally, the lack of trust in the entities running a digital contact tracing program can negatively influence the acceptability. Finally, the choice of mode should take into account practical concerns of cost and utility for the user and the supplier. Given these considerations, we propose to complement previous work by quantitatively studying how users' perceptions of the mode of digital contact tracing influences users' willingness to share digital contact tracing data. We regard user identification as critical in contact tracing for enforced containment and not just exposure alert (in anonymous tracing); thus, we explicitly consider disclosing personal data following CDC guidelines despite privacy concerns. We also propose to understand the relative acceptability of the different modalities within a spectrum of manual to automated data collection modes.

3 DATA & METHODOLOGY

We conducted an online scenario-based survey study utilizing Amazon Mechanical Turk (MTurk) to understand how previously identified individual characteristics and modes might affect users' willingness to share contact tracing information in different disease outbreak scenarios. Note: MTurk is a crowdsourcing marketplace hosted by Amazon, allowing us to recruit participants (also known as 'workers') for our survey research. Prior studies in HCI research have leveraged MTurk to provide researchers with a diverse sample of participants from tens [55] to hundreds [5], and thousands [64], based on tasks complexity. Later in Section 3.4, we provide results from conducting a power analysis to test the probability of our design succeeding with the number of samples acquired from our study.

By not limiting the study to users' responses to one specific disease, our goal was to obtain generalizable knowledge regarding contact tracing for infectious diseases and how they are affected by individual characteristics and data collection modalities. Our study was approved by the university's Institutional Review Board (IRB) and took place between November and December of 2020.

3.1 Choice of Scenarios

We selected scenarios from the World Health Organization's (WHO) list of top ten threats to global health, where these communicable diseases account for almost one-third of deaths worldwide [37]. The selection of infectious agents covers a broad range of transmission features and participants' level of familiarity given the outbreak histories in the U.S. [37]. Diseases were categorized by their transmission methods (i.e., fomite/surfaces, animal vector, air/droplets, sexual transmission). We analyzed them by infectivity and curated a mix of novelty diseases. Based on these criteria, we narrowed the scenario list to six diseases – Human Immunodeficiency Virus (HIV), Novel Coronavirus (nCov), Zika Virus, Ebola Virus, Methicillin-resistant Staphylococcus Aureus (MRSA), and Hepatitis causing virus (Hep). Table 2 shows the feature spread for this selection of diseases.

Disease Name	Trans mission	Infect iousness	Detrimental Effects	Comment(s)
HIV	Sexual contact	Varies by health care context	High lifetime cost	Need sexual partner data for contact tracing
nCov	Air/Droplets	High	Age and health profile dependent	Current pandemic
Zika	Vector (Mosquitoes)	Moderate	Specifically to pregnant mothers and fathers-to-be	-
Ebola	Direct Contact, bodily liquids	High	Highly fatal	Small U.S. out -break in 2014
MRSA	Surfaces	High	High, Currently no Treatment	May occur in hospitals
Hep	Sexual contact, bodily liquids	Moderate	High	Novel virus but older vaccine hints to possible new variant outbreak

Table 2: Disease feature spread selected for the scenarios.

3.2 Scenario-based Survey

We chose to identify each disease in the given scenarios. We considered white-labeling (not naming) each disease scenario to reduce users' existing perceptions of the particular disease. However, after discussion, we realized that presenting the specific characteristics of the disease would still cause participants to recall their understanding or experiences of that characteristic. In contrast, naming the actual disease would improve the likelihood that the responses

(a) HIV Scenario.

(b) nCov Scenario.

Figure 1: Two of six disease scenario surveys prepared on Qualtrics and integrated with MTurk.

are real-world perceptions of the disease and the scenarios are NOT associated with COVID-19, which would have had high saliency during the study period.

A participant is randomly assigned to a scenario describing one particular type of infectious disease outbreak. As shown in Figure 1, each scenario has the following structure: First, it describes the disease's transmission and severity/lived experience. Then, we provide the current method of confirming infection and, finally, the measures for curbing the disease spread. It is important to note that the reader's/ participant's infection status is not stated in the scenario.

Based on the established contact tracing interview process (see Table 1), we designed each scenario requesting three critical pieces of private information relevant to contact tracing. They are:

(1) Subject's identity defined by their name and social security number,
(2) Subject's contact details defined by their phone number and email address,
(3) Details of exposure are either defined by the locations visited or information of their sexual partners to identify contact transmission.

Participants rated their preference on four types of modalities where private information will be collected. These choices represent a spectrum of data collection modes, from an entirely manual collection to a fully automated collection on the user's end. They are:

(1) Communicating directly with public health officials on the relevant data,
(2) Granting public health officials access to extract the data from existing medical health records,
(3) Providing the data through their smartphone,
(4) Automatically extracting the data from their internet browsing activities.

Our primary consideration for smartphones is with the assumption that the utilization of smartphones is the most common across user groups [46]. We regard a smartphone as central to the usage of other digital device types, including a wearable tracker. Standalone options such as Bluetooth tokens may be used for digital contact tracing [17]. However, the smartphone remains the most immediately viable option for digital contact tracing in the U.S. during the COVID-19 outbreak. Our focus is not to address the usage of possible new modalities.

To understand better user expectations when faced with a disease outbreak event, we asked for a set of multiple-choice questions and open-ended questions. In order of presentation, they are:

- A randomly assigned disease scenario as per Figure 1.
- Q1: A hygiene question meant to test participants' understanding of the assigned disease scenario.
- Q2-5: A set of questions rating participants' willingness to share the three types of private information for digital contact tracing using four types of modality (from 1:Strongly Disagree to 5:Strongly Agree).
- Q6: A set of questions asking users to state their most preferred way among the four types of modality for collecting three types of private information for digital contact tracing.
- Q7: A set of questions to evaluate participants' trust in public health officials (from 1:Strongly disagree to 5:Strongly Agree). This scale is adapted from McKnight *et al.* [33] and is used to measure an individual's trusting beliefs towards public health officials across three constructs; competence, benevolence, and integrity.
- Q8-9: A set of questions asking participants about their experience with the disease (0:None, 1: Friends/Family, 2:Personally) and degree of concern with the assigned disease (0:Low, 1:Medium, 2:High).
- Q10-12: A set of open-ended questions querying participants' general concerns with (a) the disease, (b) the selected contact tracing procedure, and (c) their understanding of why that procedure is essential for the disease.
- Others: Participants' demographics (age, gender, income, education, marital status, and the number of children. Refer to Appendix A for supplementary materials.

3.3 Survey Item Development

The co-authors of this paper wrote the first drafts of each question, aiming to use phrasing that is brief, relevant, unambiguous, specific, and objective, as recommended by Peterson *et al.* [44]. The survey questions were then reviewed by an expert panel of two public health researchers, and items were modified based on their feedback. Finally, the questions were prepared in Qualtrics [48], where they passed the check for readability at 9th-11th grade reading levels.

3.4 Participant Recruitment

The survey was then integrated with MTurk, where we selectively defined the qualifications for potential respondents based on the Number of Human Intelligence Tasks (HITs) Approved, HIT Approval Rate, and Location (U.S.). Each participant was randomly assigned to a scenario survey related to one of six disease groups. Each scenario survey was estimated to take 10-15 minutes to complete, with compensation of US$1.50.

Power Analysis. Our user study was designed to have the same sample size for each of the six disease groups. However, the exclusion of unreliable data unavoidably led to unequal sample sizes. One likely explanation is that we detected plagiarism (high similarity to content searched from the web) within the open-ended responses more frequently in entries that were not related to nCov. This resulted in more exclusion of responses in the non-nCov disease scenarios.

However, the relative difference between groups was 4% or less (Table 4).
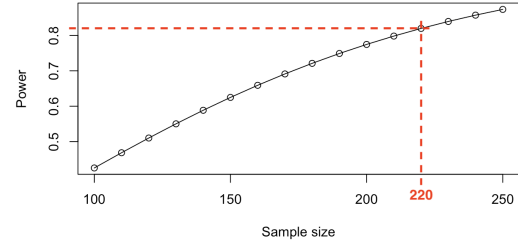
Figure 2: Power curve plots the relationship between sample size and power. A sample size of 220 participants yields a power of 0.8.

Disease	n	power
HIV	35	0.80
MRSA	37	0.82
nCov	43	0.89
Zika	37	0.82
Hep	35	0.80
Ebola	33	0.77

Table 3: Power calculation per disease group.

Our analyses aimed to study the relationship between individual factors and user willingness to share information via various contact tracing methods. A power of 0.8 is often desired. The general rule is to assume a medium effect size of 0.25 without information to our experiments [4]. Based on our analysis in Figure 2, our total sample size of 220 would achieve power above 0.8. Table 3 summarizes the power calculation per disease group. Given our unbalanced design (i.e., the number of samples in each group is not constant), the power for our samples in the Ebola group was slightly under 0.8. However, as will be explained in Section 4, our experiments moving forward would consider all groups as a whole.

3.5 Data Collection

Table 4 summarizes the key details for two rounds of data collection.

Parti-cipants	421 total, 314 (verified US National/Resident), 220 (verified entries)		
Gender	M (128), F (89), Not stated (3)	Period	Nov-Dec 2020
Status	Married (91), Single (107), Others (22)	Parent	Parent (98), No child (122)
Age	18-24 (8), 25-34 (94), 35-44 (54), 45-54 (27), 55-64 (20), > 65 (17)	Edu-cation	Skilled (13), Degree (135), Higher Degree (72)
Ethn-icity	Asian (12), Caucasian (171), Afr-ican American (17), Others (20)	Region	Northeast (40), South (96) West (48), Midwest (36)
Disease Type	HIV (35), MRSA (37), nCov (43), Zika (37), Hep (35), Ebola (33)	Income	< 25k (102), 25-100k (99), > 100k (17), NA (2)

Table 4: Summary of 220 verified participants' demographics.

Verification Procedure. We recruited a total of 421 participants with 314 participants stating their nationality and region of residence in the U.S. Our final analyses relied on verified responses from 220 participants. Verification was performed by our research team manually inspecting the qualitative data answering open-ended questions.

Ninety-four responses were discarded. The removal steps were: First, if a participant supplied an incorrect answer to the hygiene test (Q1), we verified the survey's legitimacy through their qualitative responses (Q10-12). Secondly, we removed entries if any response to the qualitative questions had plagiarized content searched on the web or repeated exact words. 19 out of 220 valid responses provided a wrong answer to the hygiene question. They were still considered valid as their qualitative questions were not plagiarized in any way, suggesting the presence of personal input by the participant.

Coding Procedure. Thematic analysis [7] was used to analyze the open-ended answers. One researcher generated codes and themes by reviewing all participants' responses, and disagreements on themes were iteratively resolved with the team (refer to Appendix B). Two researchers applied the pre-established code to our data

set and achieved a Cohen's $\kappa = 0.816$ (strong level of agreement), where 1 is perfect agreement.

4 EXPLORATORY ANALYSIS

Table 5 provides a summary of data (N=220) grouped by six disease types in order of survey presentation. Despite applying data transformation, results from conducting a Shapiro-Wilk test for data normality achieved statistical significance ($p < .001$) implying that the distribution of the data are significantly different from normal distribution and thus, we cannot assume normality. We compared the variances of six groups with Levene test, which is less sensitive to departures from normality based on the median. The test results in no significant differences (location: $p = .119$, contact: $p = .147$, and identity: $p = .127$). With no evidence suggesting that the variance of information sharing is significantly different for the disease groups, we test the effects of diseases and willingness to share information (location: $p = .088$, contact: $p = .055$, and identity: $p = .098$). The Kruskal-Wallis test revealed no significant difference ($p > .05$), thus did not provide strong evidence to reject the null hypothesis that these disease groups have identical means. Separately, post-tests are more conservative to find differences between groups. Since the data are not normally distributed,

Data	HIV	MRSA	nCov
Hygiene	TRUE:30, FALSE:5	TRUE:37, FALSE:0	TRUE:42, FALSE:1
Share location data	-	$\bar{x}$: 3.55, *M*: 4, *s*: 1.22	$\bar{x}$: 3, *M*: 3.25, *s*: 1.08
Share sex partner data	$\bar{x}$: 2.43, *M*: 2, *s*: 1.31	-	-
Share contact data	$\bar{x}$: 2.62, *M*: 2.75, *s*: 1.39	$\bar{x}$: 3.48, *M*: 4, *s*: 1.31	$\bar{x}$: 2.99, *M*: 3.25, *s*: 1.11
Share personal data	$\bar{x}$: 2.54, *M*: 2.5, *s*: 1.36	$\bar{x}$: 3.22, *M*: 3.75, *s*: 1.24	$\bar{x}$: 2.88, *M*: 3, *s*: 1.06
CT by public health officials	$\bar{x}$: 2.63, *M*: 3, *s*: 1.38	$\bar{x}$: 3.80, *M*: 4, *s*: 1.3	$\bar{x}$: 3.4, *M*: 4, *s*: 1.21
CT through health records	$\bar{x}$: 2.75, *M*: 2.67, *s*: 1.5	$\bar{x}$: 3.68, *M*: 4, *s*: 1.36	$\bar{x}$: 3.22, *M*: 3.67, *s*: 1.3
CT through smartphone	$\bar{x}$: 2.54, *M*: 2.67, *s*: 1.42	$\bar{x}$: 3.44, *M*: 4, *s*: 1.38	$\bar{x}$: 2.97, *M*: 3.33, *s*: 1.27
CT through internet	$\bar{x}$: 2.19, *M*: 1.67, *s*: 1.41	$\bar{x}$: 2.74, *M*: 3.17, *s*: 1.43	$\bar{x}$: 2.24, *M*: 2, *s*: 1.26
Trust in public health officials	$\bar{x}$: 3.62, *M*: 3.82, *s*: 0.99	$\bar{x}$: 3.99, *M*: 4.27, *s*: 1.03	$\bar{x}$: 3.63, *M*: 3.91, *s*: 0.91
Disease Experience	0: 27, 1: 7, 2: 1, 3: 0	0: 27, 1: 6, 2: 4, 3: 0	0: 19, 1: 18, 2: 3, 3: 3
Disease Concern	High: 6, Med: 9, Low: 20	High: 5, Med: 19, Low: 13	High: 18, Med: 20, Low: 5
Data	**Hep**	**Zika**	**Ebola**
Hygiene	TRUE:34, FALSE:1	TRUE:31, FALSE:6	TRUE:27, FALSE:6
Share location data	$\bar{x}$: 2.91, *M*: 3, *s*: 1.14	$\bar{x}$: 2.97, *M*: 3.38, *s*: 1.42	$\bar{x}$: 3.31, *M*: 3.5, *s*: 1.04
Share contact data	$\bar{x}$: 2.81, *M*: 3, *s*: 1.13	$\bar{x}$: 2.92, *M*: 3.25, *s*: 1.32	$\bar{x}$: 3.25, *M*: 3.5, *s*: 1.08
Share personal data	$\bar{x}$: 2.62, *M*: 2.5, *s*: 1.11	$\bar{x}$: 2.68, *M*: 2.88, *s*: 1.18	$\bar{x}$: 3.14, *M*: 3.25, *s*: 1.01
CT by public health officials	$\bar{x}$: 3.29, *M*: 3.67, *s*: 1.27	$\bar{x}$: 3.07, *M*: 3.67, *s*: 1.38	$\bar{x}$: 3.60, *M*: 4, *s*: 0.99
CT through health records	$\bar{x}$: 3.06, *M*: 3.33, *s*: 1.28	$\bar{x}$: 3.04, *M*: 3.5, *s*: 1.29	$\bar{x}$: 3.33, *M*: 3.67, *s*: 1.27
CT through smartphone	$\bar{x}$: 2.77, *M*: 3, *s*: 1.38	$\bar{x}$: 2.97, *M*: 3.33, *s*: 1.33	$\bar{x}$: 3.33, *M*: 4, *s*: 1.24
CT through internet	$\bar{x}$: 2.01, *M*: 1.67, *s*: 1.14	$\bar{x}$: 2.34, *M*: 2.33, *s*: 1.3	$\bar{x}$: 2.67, *M*: 2.33, *s*: 1.37
Trust in public health officials	$\bar{x}$: 3.68, *M*: 3.82, *s*: 0.88	$\bar{x}$: 3.34, *M*: 3.68, *s*: 1.01	$\bar{x}$: 3.71, *M*: 3.73, *s*: 0.73
Disease Experience	0: 26, 1: 8, 2: 1, 3: 0	0: 34, 1: 3, 2: 0, 3: 0	0: 23, 1: 6, 2: 4, 3: 0
Disease Concern	High: 5, Med: 9, Low: 21	High: 2, Med: 10, Low: 25	High: 8, Med: 8, Low: 17

Table 5: Summary statistics in order of survey presentation. $\bar{x}$ denotes sample mean, *M* is sample median, and *s* is the standard deviation of a sample. TRUE refers to participants correctly answering the hygiene question. Willingness to share data, modes for Contact Tracing (CT), and Trust in public health officials scores are on a 1-5 Likert scale. Disease experience scores 0 for no experience, 1 for friends and family, 2 for personal experience, 3 for more than one experience. Disease concern counts participants expressing high, medium and low concern.

we employed the Dunn's test with FDR adjustment as a nonparametric post-hoc alternative. The tests yield no statistical difference in means ($p > .05$), indicating the data can be treated as a whole, except for an HIV scenario which requires a different set of contact tracing information (see Section 3.2). Appendix C reports the statistical results for these tests; Figures 9, 10 and 11 chart the differences between groups, while Tables 11, 12 and 13 summarize the p-values between two disease groups for every data type.

Despite the similarity between disease scenarios, in the following paragraphs we point out some data of note. Note: when reporting the results, M stands for median and $\bar{x}$ for mean.

Overall, 91% of our participants indicated a correct understanding of the disease's nature by providing the true answer to the hygiene question (Q1). All participants were asked about their experience with the disease (Q8). All disease groups except nCov

had most participants reporting no experience of the disease (Q8). One participant reported personal experience with HIV, while 24 participants noted some personal experience with nCov.

In expressing their concerns on the diseases (Q9), participants expressed the lowest concern for a Zika outbreak (n=25). Most participants expressed deep concern for a nCov outbreak (n=18). These statistics are consistent with the current state of knowledge on disease threats. Specifically, only four Zika transmissions were reported to have occurred in the U.S. since 2020 [8]. On the other hand, with the most recent coronavirus first identified in January 2020 (COVID-19) and the country battling over 30 million cases (as of when this study was conducted), it is reasonable to expect the concern for this disease to be high [60].

Participants were asked about their trust in public health officials, as defined by the modified "trusting belief" construct from [33] (see
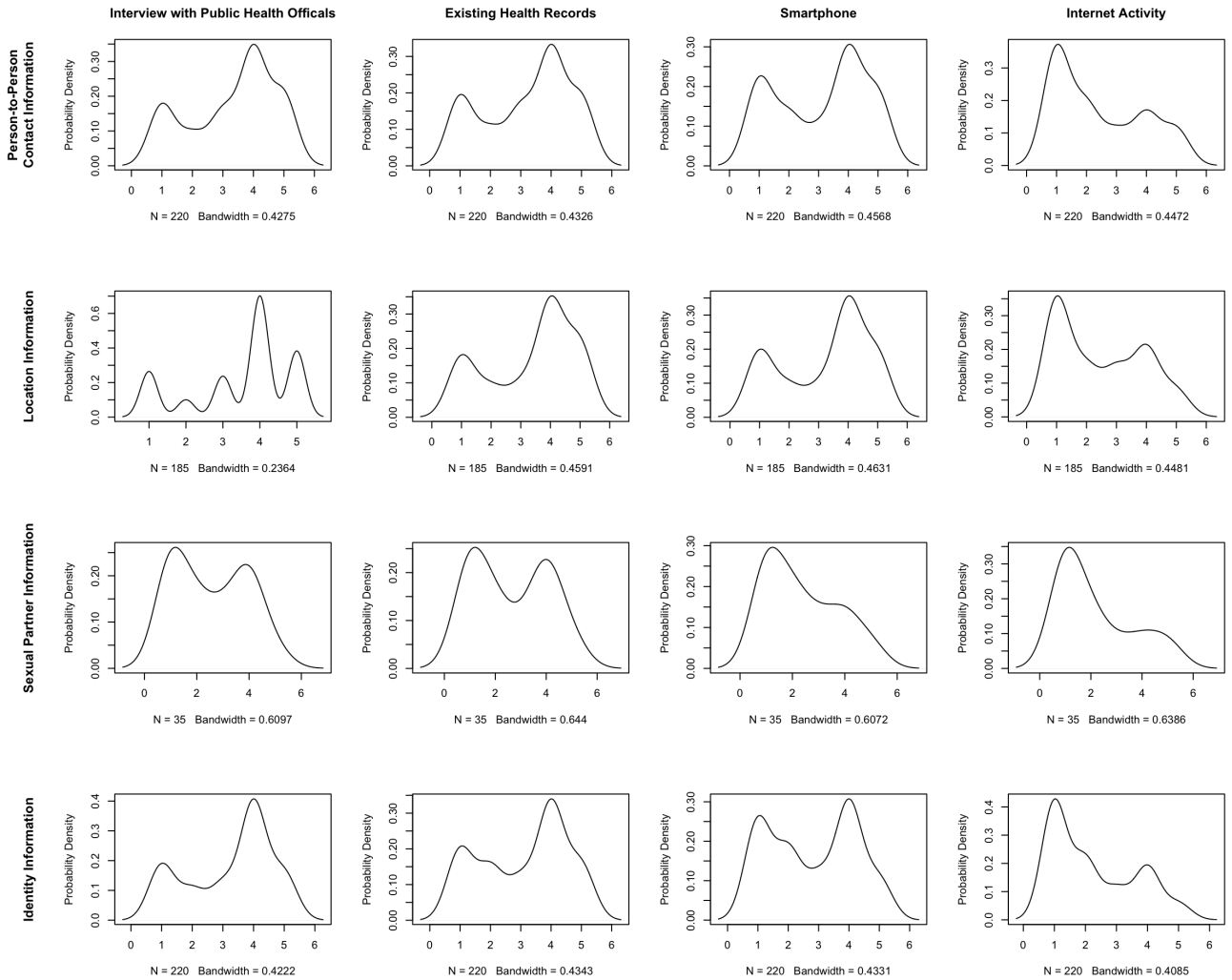

Figure 3: Probability distribution of willingness to share information through different modality.

Section 3.2). Trust in public health officials is moderately high among all participants across disease groups ($\bar{X}$: 3.66/5, M: 3.91/5, SD: 0.94). This result is consistent with Pew Research's reporting that a majority of Americans are somewhat confident with public health organizations [45].

Participants expressed more willingness to share private information for tracing MRSA (e.g., location: M:4, contact: M:4, personal: M:3.75) and Ebola (e.g., location: M:3.5, contact: M:3.5, personal: M:3.25) compared to other diseases. It is worth noting that users' willingness to share private information is lowest among our HIV sample, particularly in providing sex partner information. Our results support Njozing *et al.* findings that highlight the challenges in persuading patients to disclose their HIV status to their partners after testing [39].

While more participants generally preferred communicating directly with a public health officer for contact tracing, scores for other modalities are consistently higher among our MRSA sample (e.g., M for contact tracing methods – smartphone: 4, internet: 3.17, public health officer: 4, health record: 4). From our review of the open-ended responses, it seems that MRSA is perceived as being a high threat. For example, participant P1 responded, "MRSA is one of the most common antibiotic-resistant bacteria. [It] can occur between food[,] animals and people," and P2 described it as "a very bad situation [to have] a resistance bacterial outbreak." On the other end of the scale, participants in the HIV scenario rated the modality preferences lowest. These results are consistent with prior findings on contact tracing pushback in response to HIV, primarily due to the stigma of the virus [13]. P3 expressed, "I think that personal information and your status [should not] be attached to your identity. There is a negative stigma with HIV positive individuals." P4 wrote, "If I was infected, I would be concerned about suffering

discrimination from those assuming the worst about how I got infected."

4.1 Mixing Modalities and Reducing Data Dimensionality

We plotted Figure 3 to examine if user willingness to share different types of information can improve through different data collection modalities. We observed an overall decrease in user willingness when the data collection method shifted from manual to increasingly automated methods. Lying on the most extreme of user willingness, sharing any kind of private information through internet browsing histories remains the least acceptable method among users in all scenarios.

Taking a step further, we found that reducing data dimensionality for each modality offered more promising results. As shown in Figures 4 and 5, we observed increasing users' acceptance of a tool when it was not set to collect all three critical types of contact tracing information. Instead, most of our participants preferred using a mix of modalities to share their private information, and these preferences differ with information types. Specifically, most users would prefer to either communicate with a public health officer about their identity and contact information (>30%) or grant access to their health records of this information ($\approx$40%). Collecting location information using a smartphone remains favorable. However, user willingness decreased to less than 20% when identity and contact information are collected alongside location histories.

In the context of an HIV scenario, our results showed a higher preference for utilizing health records to retrieve information related to their sex partners (>50%). This preference is in contrast to non-sexually transmitted infections, where participants were more positive towards communicating private information directly with
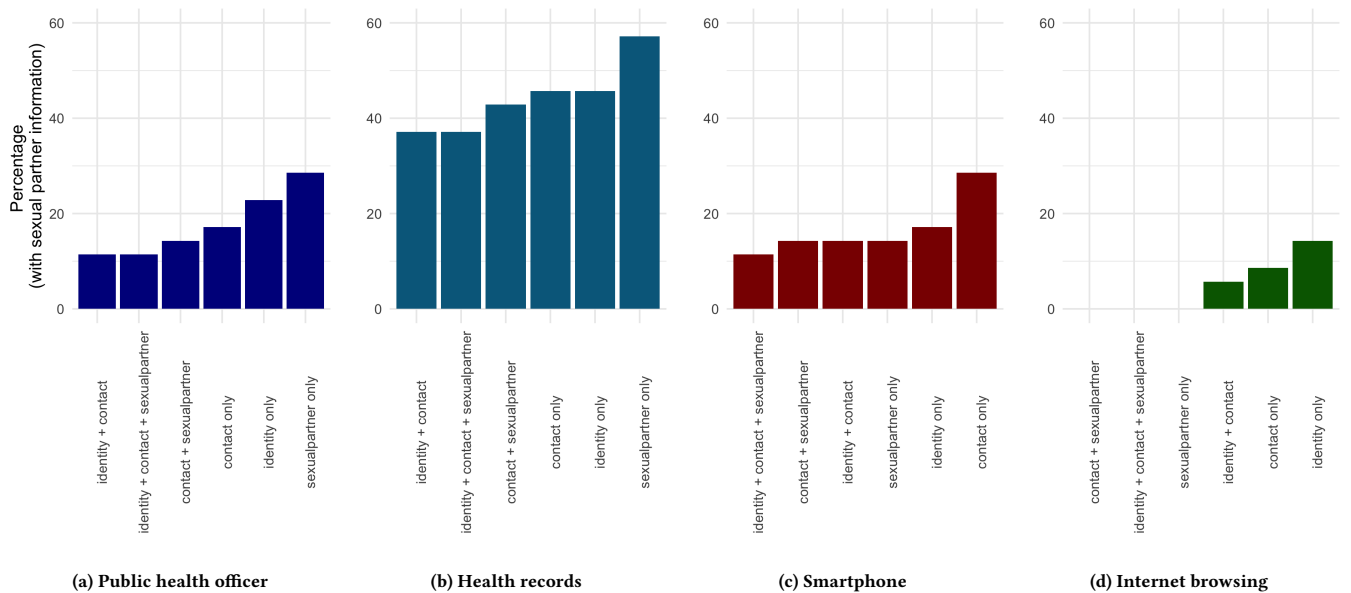

Figure 4: Percentage of users preference to share data, a combination of personal identifiable, contact details and sexual partner PII for HIV disease, through different methods.
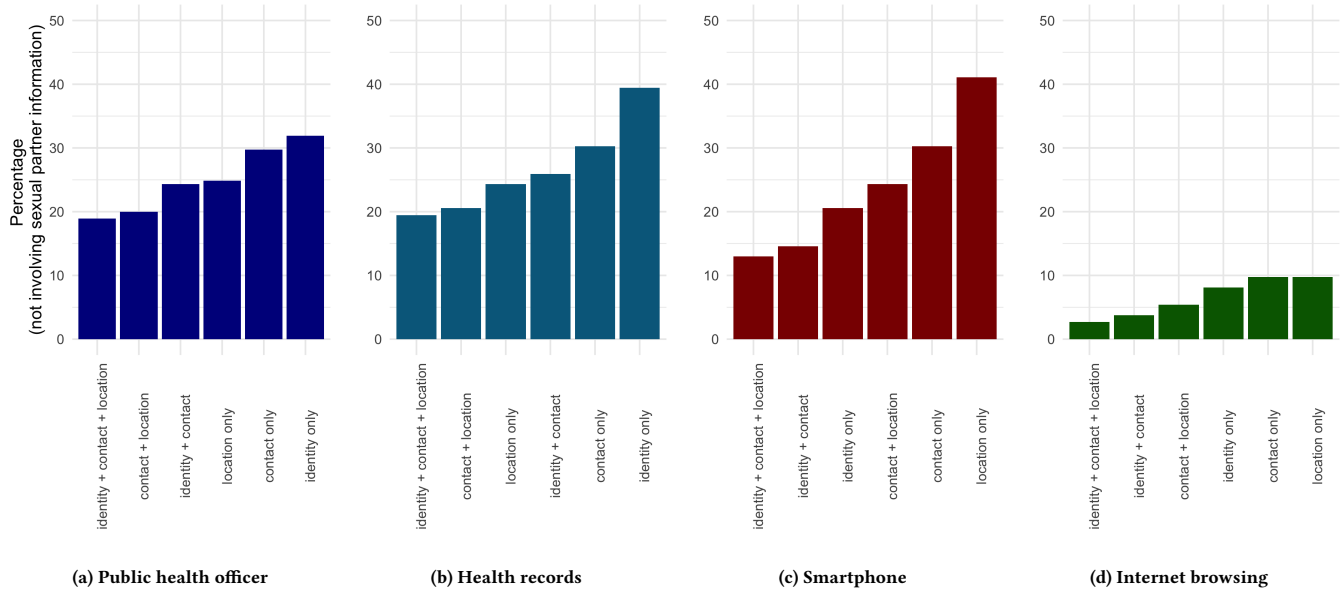
Figure 5: Percentage of users preference to share data, a combination of personal identifiable information, contact details and location information for most diseases, through different methods.

a public health official and utilizing a smartphone to share such information.

4.2 Visualization of Qualitative Responses

As a final step, we generated word clouds to visualize our qualitative responses in Figure 6. Participants were asked about their concerns about the disease type, sharing their private information for contact tracing and whether contact tracing is important. It is important to note that while these representations do not capture the meaning of the content, they help identify the most salient issues in this exploratory phase.
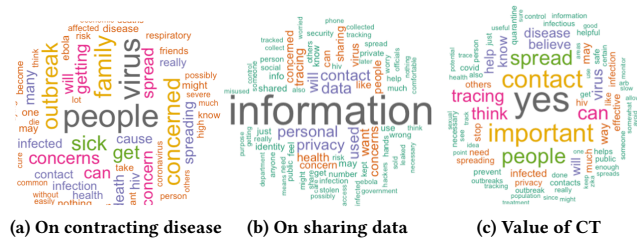

Figure 6: Summary of the most salient issues captured in users' qualitative responses

Firstly, our users' two most prominent words are 'people' and 'virus' when asked about their general concerns about the disease. These words are commonly associated with *fear of contracting disease*, followed by *fear of death*. Despite expressing extreme worries, their concerns with sharing contact tracing information for the disease are mainly related to *loss of privacy* and *misuse of data*.

Words such as 'information,' 'personal privacy,' 'contact,' and 'used' from their expressed concerns conform with our understanding of why contact tracing faced strong community resistance during the COVID-19 pandemic [57]. Some participants expressed worry that their data would fall into the hands of the wrong organization. P5 reported, "I [am] concerned about [a] data breach that we hear about with other companies having our data stolen and lost. I [do not] want identity theft. I had it before[,] and [it is] no fun." P6 responded, "I [do not] want the information compromised or sold to companies to stalk me."

Nonetheless, in terms of understanding whether contact tracing is essential, the word 'yes' arises as a significant term. Some lower-order terms include 'spread' and 'stop.' These terms collectively reveal an unexpected determination to leverage contact tracing as a way of *curbing virus spread*. When asked if they believed in the importance of contact tracing, P7 replied, "Absolutely - it is the only way to stop the spread." P8 responded, "Yes if it's done right and safe to use for the users. It will immensely stop the disease." In the case of a HIV spread, P9 expressed, "No, because I don't see much of it in my community."

Our work aims to unpack the strong user resistance on digital contact tracing and discover better means of facilitating critical information sharing. These preliminary steps hint toward promising insights into how information sharing in the contact tracing process can be reshaped to balance user concerns.

5 REGRESSION ANALYSIS

We carried out multiple linear regression to investigate the relationship between users' willingness to share private information for contact tracing (DV: dependent variables) and the preferred

Model	DV	modality	demographic	AIC	R^2	adj. R^2	Std.e	df	F	$Sig.$
$M^{identity}$	identity	smartphone, PH-HR	parenthood,income	-293.45	0.819	0.816	0.501	(4, 211)	240.1	0'***'
$M^{contact}$	contact details	smartphone, PH-HR	education,income,age	-393.73	0.899	0.896	0.396	(5, 210)	374.4	0'***'
$M^{location}$	location	smartphone, PH-HR	parenthood, income,trust	-278.88	0.857	0.853	0.461	(5, 178)	214.1	0'***'
$M^{partner}$	sex partner's details	smartphone, PH-HR	education,income	-51.98	0.903	0.889	0.413	(4, 27)	63.13	0'***'

Table 6: Model summary for user willingness to share critical contact tracing information.

modality (IV: independent variables). We also utilized *trust in public health* and a range of previously identified user demographics variables such as *income* and *education* as potential predictors.
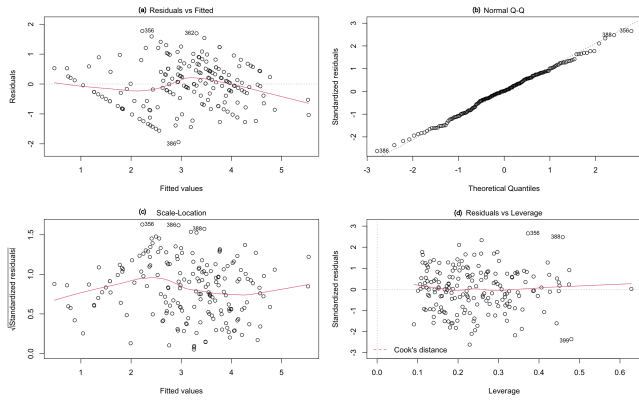

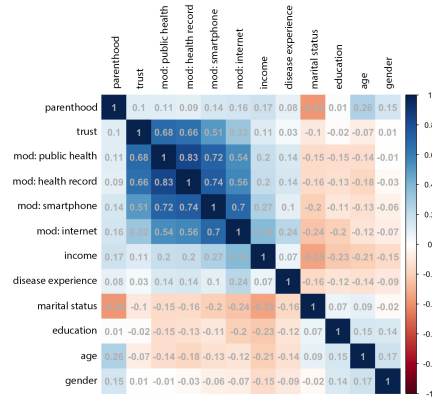**Figure 7: Regression diagnostic.**

Figure 8: Correlations between variables.

Our regression diagnostics plots in Figure 7 confirmed assumptions such as normality of errors, homoscedasticity, and multicollinearity were met for multiple regression to be reliable. The horizontal line (a) indicates a close to linear relationship without distinct patterns, the residual points follow the dashed line in (b), the horizontal line (c) indicates homoscedasticity, and (d) identifies few data points with high leverage. However, the standardized

residuals of these observations are no more than 3 in absolute value as suggested by [19].

The next step was to ensure no one independent variable is strongly correlated (r=0.8) with another. As common practice, highly correlated variables must be removed. Figure 8 shows the correlations between demographic variables are no more than r=0.3. However, there is a high correlation between communicating with public health officers and retrieving health records for different modalities (r=0.83).

We examined the phrasing of the questions (Questions 8.4 and 8.5 in Appendix A), as well as the qualitative responses in order to understand this high correlation. There is a possibility that participants perceive the mode of collecting information "personally from a public health official" as bearing similarity with "allowing such information to be collected from your doctor's medical health record." However, we are unable to unpack the meaning of this. Moving forward, we have decided to account for this high correlation by using the data for "health records" modality and to discuss the combined item as "public health-health record" (PH-HR).

We carried out stepwise regression to procedurally determine the best combination of independent variables (IV). Table 6 lists the model summary for each of our dependent variables (DV). We iteratively performed a backward selection to achieve a simplified model with the lowest Akaike Information Criteria (AIC), without impacting model performance; the lowest value quantifies the amount of information loss due to model simplification.

Overall, the results suggest that more than 80% of the variance in the data can be explained by all models (R^2 for $M^{location}$: 0.857, $M^{partner}$: 0.903, $M^{identity}$: 0.819, $M^{contact}$: 0.899). We summarize the coefficient estimates for all four models in Tables 7 to 10.

Choice of Modality In these models, we found the use of smartphone and PH-HR consistent as significant predictors. Specifically in predicting user willingness to share location data ($M^{location}$), the positive coefficient indicates an increase by 0.533 ($p < .001$) given a one-unit shift in the use of smartphone with other variables held constant (see Table 7). This result allows us to assess the effect of each modality isolated from other variables, confirming the use of smartphone to gather location information to be higher than other modalities.

In contrast, we found participants who responded to the HIV disease scenario to be more willing to share their sexual partner information through the use of PH-HR. As shown in Table 9, the mean in willingness to provide sexual partner information ($M^{partner}$) increases by 0.559 given a one-unit shift in the independent variable

Coefficient	β	est.	Std.e	t	Sig.
(Intercept)		-0.09	0.160	-0.563	0.573
smartphone	0.533	0.483	0.038	12.686	< 2e-16 ***
PH-HR	0.369	0.338	0.043	7.739	7.23e-13 ***
parenthood	-0.048	-0.053	0.031	-1.677	0.095 .
income	0.082	0.087	0.031	2.802	0.005 **
trust	0.100	0.129	0.050	2.562	0.011 *

Table 7: Coefficient summary to predict user willingness to share location information, $M^{location}$.

Coefficient	β	est.	Std.e	t	Sig.
(Intercept)		0.040	0.140	0.286	0.775
smartphone	0.485	0.445	0.030	14.576	< 2e-16 ***
PH-HR	0.510	0.468	0.030	15.438	< 2e-16 ***
education	-0.063	-0.077	0.027	-2.817	0.005 **
income	0.052	0.055	0.025	2.216	0.027 *
age	0.045	0.042	0.021	1.994	0.047 *

Table 8: Coefficient summary to predict user willingness to share contact information, $M^{contact}$.

Coefficient	β	est.	Std.e	t	Sig.
(Intercept)		1.083	0.370	2.927	0.006 **
smartphone	0.419	0.380	0.100	3.805	0.000 ***
PH-HR	0.559	0.479	0.091	5.222	1.68e-05 ***
education	-0.184	-0.218	0.083	-2.608	0.014 *
income	-0.163	-0.169	0.075	-2.245	0.033 *

Table 9: Coefficient summary to predict user willingness to share sexual partner information, $M^{partner}$.

with the smartphone modality (and other variables) held constant. Our findings in Table 8 indicate user willingness in utilizing PH-HR to extract contact information ($M^{contact}$, β=0.510, $p < .001$). This result supports the feasibility in extracting characteristics of household contacts from the electronic medical record, as suggested by Metlay *et al.* [35].

Finally, in Table 10, our results seem to suggest that participants might be less particular of the modality choice (β smartphone=0.449, $p < .001$, health record=0.489, $p < .001$) for providing information exposing their identity ($M^{identity}$).

Demographic Details. In all models, users' *income* is one demographic data that commonly yields significance. The positive coefficient indicates that as income increases, the mean of the sharing location, contact and personally identifiable information for contact tracing tend to increase.

Education is also a significant demographic predictor for sharing contact and sexual partner information. It should be noted that education level is an ordinal independent variable (i.e., 1: Skilled, 2: Degree, 3: Higher Degree). Education is observed to have a negative relationship in these models, suggesting that users with lower educational attainment are less willing to provide contact tracing information.

Separately, *parenthood* reveals inverse relationships between sharing location and personal particulars. Non-parent users are less willing to provide location data, while parents are more willing to provide personal data for contact tracing.

Coefficient	β	est.	Std.e	t	Sig.
(Intercept)		0.005	0.120	0.043	0.966
smartphone	0.449	0.390	0.038	10.084	<2e-16 ***
PH-HR	0.489	0.426	0.038	11.207	<2e-16 ***
parenthood	0.042	0.046	0.032	1.441	0.1511
income	0.075	0.076	0.031	2.474	0.014 *

Table 10: Coefficient summary to predict user willingness to share personal identity, $M^{identity}$.

Finally, it is worth noting that *trust in public health* plays a significant role in users providing location information, but not other kinds of information. The mean willingness to provide location data increases by 0.1, given a one-unit shift in trust as a predictor (with all else held constant). We reviewed codes for open-ended answers in response to the question, "What are your concerns with sharing contact tracing information for the disease?" Respondents unwilling to share location information (1 on the Likert scale) seemed to be most concerned with privacy issues and concerns arising from cybercrimes or abuse of power. For example, P10 said, "It might be misused and the data might reach third parties and endanger my privacy." P11 perceived contact tracing as a "trojan horse for access[ing] private information of location, activities, and constitutional rights being infringed."

5.1 Summary of Findings

In this section, we ran a quantitative analysis to explore the predictors affecting users' willingness to share their private information for contact tracing. Our results showed that leveraging internet browsing histories as a data collection modality is a redundant predictor to collect any information (see Table 6). The evidence indicates strong support for using smartphones. The role of trust in public health organizations is particularly critical when users are asked to share location data through their smartphones.

Furthermore, this analysis suggests the importance of a modality *PH-HR*, that may be a configuration of public health officials extracting relevant data from medical health records. The preference for sharing more private information such as contact data (for all disease types) and sexual partner information (for HIV) is higher through this modality.

Regardless of data collection modality, some previously documented demographic, non-modifiable factors were found to significantly influence user willingness to provide these data. They are primarily *income* and *education* [20]. *Parenthood* emerged as an intriguing possible new demographic predictor.

6 DISCUSSION

Our study's objectives were to better understand how individual factors and modes of data collection affect user acceptance in digital contact tracing. Here we discuss the key takeaways of our findings.

6.1 How Mode Influences User Willingness to Share Contact Tracing Information

Since the COVID-19 outbreak, much focus on mobilizing contact tracing efforts has fundamentally relied on communicating with

public health officials and utilizing smartphone sensing modalities for partial automation of contact tracing. Our study sought to examine these and alternative methods of data collection.

6.1.1 Influence of Disease Type on Contact Tracing Preferences. While our study observed six different diseases, each depicting a unique feature spread, our statistical findings did not demonstrate strong evidence of disease groups influencing users' willingness to share different types of critical contact tracing. Accordingly, we had no basis to pursue our regression analysis by disease type. It is important to note that our study did not investigate the threat perception of each disease, which may have exhibited differentiation [42, 52]. Further, we cannot determine if the lack of differentiation by disease resulted from studying users' responses to the mode of contact tracing efforts and not to each unique disease. Future work may require a systematic examination of threat perception for different disease states and whether it is a moderating factor in users' responses to the mode of digital contact tracing.

6.1.2 User Resistance towards Internet Modality. Despite some success in using automated data collection of search results, our findings suggest that acceptability will be low. In a study related to COVID-19, Li *et al.* highlights internet data as a reliable source that can accurately predict the COVID-19 outbreak one to two weeks earlier before the first laboratory confirmation was announced. Word searches such as 'coronavirus' and 'pneumonia' peaked on Google Trend [28] five days before the COVID-19 was announced as a global health emergency by WHO [68]. Similarly, Shin *et al.* found internet data to monitor MERS-CoV (Middle East respiratory syndrome) three days before laboratory confirmations [56]. Accordingly, prior work has proposed internet surveillance data to complement tracking infection trends in the COVID-19 pandemic [30]. In this work, we examined the idea of collecting contact tracing information through internet browsing histories as a possible modality. As shown in Figure 4, utilizing search data to gather any contact tracing information is not likely to reach beyond a 20% adoption, thus, insufficiently sized for mass adoption. This result is not surprising, given high concerns over behavioral advertising through surreptitious tracking and targeting of users. Such utilization of internet data is precisely expressed as a concern. P12 explained, "My concerns would be using the internet as a source of information. It would give access to information that does not pertain to health."

6.1.3 Potential of Medical Health Records in Coordination with Public Health Personnel as a Modality. Our findings support the use of medical health records for collecting contact tracing information. Specifically, our exploratory analysis found approximately 75% users willing to share contact data and close to 90% users willing to provide personal information through medical health records. Indeed, this approach was reported by Metlay *et al.* to be effective in assisting with COVID-19 contact tracing. The authors successfully identified immediate contacts of a patient through using electronic medical health records, leading to crucial information for estimating household transmission risk [35]. However, medical health records are generally a poor source of current location information. There may be outdated home addresses or miscounting of immediate family contacts. However, since our findings suggest utilizing medical

health records enjoys high user acceptance as a data source for gathering contact and personal information, it may be worth discussing further how to employ this mode of digital contact tracing.

6.2 Users' Concerns within Each Mode

Our regression analysis revealed user's perception of different modalities and the activity of information exchange.

6.2.1 Smartphone Location-sensing Needs Trust. Our participants were more willing to share location data when the mode of digital contact tracing was smartphones. However, this finding does not adhere with the low application uptake for digital contact tracing mobile apps deployed in the U.S. during the COVID-19 pandemic[62]. These systems relied on alert notification systems (for example, a joint initiative by Apple and Google) that solely track exposure through utilizing Bluetooth proximity [3, 38].

Our study cannot precisely explain the disjunct between the stated preference to use a smartphone in sharing location data and the current poor acceptance of digital contact tracing. The open-ended answers suggests that this push back is concerned with restrictions on individual liberties and with personal safety. P13 asserted, "sharing my personal whereabouts with [the] government [will infringe on] my freedoms." P14 said, "[Tracking location] will lead to restrictions in people's ability to move about and go places once controls are implemented." Finally, P15 explained, "I am concerned about the [metadata] that can be gleaned from contact tracing apps then later used against me. What if I am circumstantially somewhere where a crime is committed, and my cell phone was the only one in the area [*sic*] I just [do not] want government having my [metadata] even though they can probably get it anyway."

At the same time, our findings allude to the significance of trust in public health officials in moderating these intuitive concerns [47] of privacy on digital contact tracing. Specifically, users with higher trust in public health officials may be more willing to supply their location data[18, 41]. Additionally, we learned from our qualitative findings that more participants had communicated confidence in public health agencies to lead the contact tracing effort. P16 expressed, "I do not have any worry about sharing my contact tracing information because I know the public health officials have my best interest at heart."

6.2.2 Addressing Limitations of Direct Communication with a Public Health Official. In most scenarios, the users preferred to communicate directly with a public health official. However, in the HIV scenario, users were more cautious towards sharing sensitive sexual partner and contact information. This difference is not statistically significant but suggests different sensitivities might need to be addressed. Additionally, direct contact by public health officials in the US is complicated by a majority of Americans screening unknown calls and voicemails [45]. Manual contact tracing is also limited by the difficulty of scaling up.

In the U.S. today, the mechanism of authorizing information sharing with primary care physicians is recommended for COVID-19 vaccine administration to maintain patient safety [2]. We suggest the same kind of information sharing could be modified to allow for digital contact tracing. Upon consent, a user could preemptively

authorize public health officials to access their medical records by providing information of the primary care physician.

6.3 Influence of Demographics on the Type of Information in Each Mode

Strategizing for high community engagement is a critical component for whole nation contact tracing to be successful. CDC and WHO have emphasized the need to improve marketing campaigns to targeted user groups [10, 67]. In this discussion point, we use our findings to shed light on the potential hard-to-reach groups.

6.3.1 Socioeconomic status (SES). Our regression results yield income as a significant demographic factor in all models. It suggests that lower income users' can be less willing to participate. This finding corresponds with prior work that users with lower-income and less formal education tend to be less comfortable sharing information [32]. This effect seemed to be particularly strong in sharing sexual partner information under an HIV outbreak.

Recent work about contact tracing in Australia [29] suggests the age of the user may further moderate the effect of socioeconomic status, where older users with low socioeconomic status are willing to participate in digital contact tracing tools. Separately, a cross-section study in France found a positive correlation between acceptance of digital contact tracing and one's health literacy level in understanding the disease [49]. While our study did not study the effects of health literacy, much research has revealed the significance of SES inequalities, including lower-income and less formal education, to negatively impact one's understanding of basic health information and its appropriate services [51]. Roughly one-half of American adults were reported to exhibit low health literacy in the U.S. [36], but the same study promisingly indicates a positive association with trust in public health. More work is required to understand if the presence of mechanisms of co-production of trust, as suggested by [34] may be particularly useful among lower SES users. One key takeaway is that public health messaging within lower SES communities may focus on basic disease information and privacy assurance when using digital contact tracing tools.

6.3.2 Domestic Characteristics and Parenthood. Previous work has identified demographic characteristics that influence contact tracing adoption [20, 29]. Our findings add to this body of work by suggesting parenthood as a new facilitator.

A household is considered a significant venue for transmission in a pandemic situation, putting everyone under the same roof at high risk [27, 35]. With contact tracing bringing higher health security for all, our findings suggest that a user with family members to consider could display a more significant commitment to following safety measures, including embracing contact tracing mechanisms. This finding may also further contextualize [34] findings regarding perceived responsibility and perceived benefit as facilitators of acceptance of contact tracing efforts.

6.3.3 Messaging Opportunities for Promoting Digital Contact Tracing. Our results identified demographic predictors of willingness to share information such as parenthood, education, and income. These results could also be used to inform the messaging for contact tracing initiatives. For example, lower education is a predictor of users' unwillingness to share contact, identity, and sexual partner

information. In the case of an outbreak of diseases where such information is needed (e.g., STIs), contact tracing campaigns may need to enhance outreach among populations with lower education or lower literacy levels. On the other hand, parenthood showed promise in terms of predicting willingness to share information. In the case of COVID-19, while young children tend to show milder symptoms [26], starting digital contact tracing through K-12 schools could be a good way to start a larger-scale initiative on a community level. Parents would be indirectly targeted by the initiative, hopefully ensuring a more extensive user base in the long run. Due to the variances in willingness to share information amongst different population groups, targeted messaging could be a promising method for promoting the uptake of digital contact tracing.

7 LIMITATIONS AND FUTURE WORK

As depicted in Appendix C, although we do not find statistical differences in the willingness to participate in contact tracing by different disease groups, we visually observe certain tendencies in some disease groups compared to others. Our approach to combining different groups into a single analysis may affect the differences in each disease group. Future work should identify the nuances of each disease scenario and understand how these nuances influence the user's willingness to participate in contact tracing.

Our work was done through Amazon Mechanical Turk, with a specific focus on the U.S. As such, the results might not be generalizable to other countries with significant cultural differences. The survey is limited by the nature of self-reported data, where responses are based on a hypothetical scenario. Future work should systematically assess users' acceptance of a widespread digital contact tracing solution. Some critical factors include users' demographics, perception, and health literacy.

Another limitation is our respondents' overall high level of trust due to the platform used to collect data. By definition, crowd workers are more willing to provide information online. Future work may be required to replicate this survey using paper-and-pen methods. Finally, our study was limited to sharing critical but sensitive information as per CDC guidelines. The transformation of such information from manual practice to digital forms could adopt privacy-preserving operations. With the evolving role of digital technologies, future work should alternatively investigate users' willingness to use different modes for anonymous tracing. A deeper qualitative evaluation is required to better understand users' attitudes towards these modalities.

8 CONCLUSION

We conducted a survey study where more than 200 respondents from the U.S. participated in one of six infectious disease scenarios (i.e., HIV, nCov, Zika, Ebola, MRSA, and Hep), listed as top global health threats by the WHO. The findings are similar across these diseases, suggesting strong generalizability of results beyond the current COVID-19 pandemic. We have presented our first steps in identifying the facilitators and barriers of contact tracing in a U.S. setting. Our study is set apart from prior work by examining user perception of contact tracing across six different disease types – without the procedure tied to a particular application, but the mode of information exchange itself. First, the study showed how

user willingness to provide contact tracing information could be improved by incorporating various modalities, specifically medical health records and smartphones. Second, trust in public health agencies is moderately high, suggesting a more coordinated effort of digital contact tracing overseen by public health institutions rather than government and technology companies in the U.S. Third, and finally, the study shed more light on the types of individual demographic factors that need targeted strategies to improve contact tracing uptake.

ACKNOWLEDGMENTS

REFERENCES

[1] Altmann, S., Milsom, L., Zillessen, H., Blasone, R., Gerdon, F., Bach, R., Kreuter, F., Nosenzo, D., Toussaert, S., and Abeler, J. Acceptability of app-based contact tracing for covid-19: Cross-country survey study. *JMIR mHealth and uHealth*, 8(8):e19857, 2020.
[2] American Academy of Family Physicians. Covid-19 vaccine.
[3] Association of Public Health Laboratories. Exposure notifications, 2020.
[4] Bausell, R. B. and Li, Y.-F. *Power analysis for experimental research: a practical guide for the biological, medical and social sciences*. Cambridge University Press, 2002.
[5] Bentley, F. R., Peesapati, S. T., and Church, K. " i thought she would like to read it" exploring sharing behaviors in the context of declining mobile web use. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1893–1903, 2016.
[6] Braithwaite, I., Callender, T., Bullock, M., and Aldridge, R. W. Automated and partly automated contact tracing: a systematic review to inform the control of covid-19. *The Lancet Digital Health*, 2020.
[7] Braun, V. and Clarke, V. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
[8] Centers for Disease Control and Prevention. 2020 case counts in the us.
[9] Centers for Disease Control and Prevention. Case investigation and contact tracing: Part of a multipronged approach to fight the covid-19 pandemic, 2020.
[10] Centers for Disease Control and Prevention. Digital contact tracing tools, 2020.
[11] Centers for Disease Control and Prevention. Preliminary criteria for the evaluation of digital contact tracing tools for covid-19, 2020.
[12] Centers for Disease Control and Prevention. Training case investigators and contact tracers, 2021.
[13] Davis, S. Contact tracing apps: Extra risks for women and marginalized groups. *Health and Human Rights Journal*, 2020.
[14] Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., and Fraser, C. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491), 2020.
[15] Gillmor, D. K. Principles for technology-assisted contact-tracing. *ACLU white paper*, 2020.
[16] Google. Exposure notifications: Using technology to help public health authorities fight covid-19.
[17] Government of Singapore. Tracetogether, safer together, 2020.
[18] Hassandoust, F., Akhlaghpour, S., and Johnston, A. C. Individuals' privacy concerns and adoption of contact tracing mobile applications in a pandemic: A situational privacy calculus perspective. *Journal of the American Medical Informatics Association*, 28(3):463–471, 2021.
[19] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An introduction to statistical learning*, volume 112. Springer, 2013.
[20] Jonker, M., de Bekker-Grob, E., Veldwijk, J., Goossens, L., Bour, S., and Rutten-Van Mölken, M. Covid-19 contact tracing apps: Predicted uptake in the netherlands based on a discrete choice experiment. *JMIR mHealth and uHealth*, 8(10):e20741, 2020.
[21] Kahn, J. P. et al. *Digital contact tracing for pandemic response: Ethics and governance guidance*. Johns Hopkins University Press, 2020.
[22] Kaptchuk, G., Goldstein, D. G., Hargittai, E., Hofman, J., and Redmiles, E. M. How good is good enough for covid19 apps. *The influence of benefits, accuracy, and privacy on willingness to adopt*, 2020.
[23] Keeling, M. J., Hollingsworth, T. D., and Read, J. M. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). *J Epidemiol Community Health*, 74(10):861–866, 2020.
[24] Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., van Boven, M., van de Wijgert, J. H., and Bonten, M. J. Impact of delays on effectiveness of contact tracing strategies for covid-19: a modelling study. *The Lancet Public Health*, 5(8):e452–e459, 2020.
[25] Larus, J. et al. Joint statement on contact tracing: Date 19th april 2020, 2020.
[26] Lee, B. and Raszka, W. V. Covid-19 transmission and children: the child is not to blame. *Pediatrics*, 146(2), 2020.
[27] Lewis, N. M., Chu, V. T., Ye, D., Conners, E. E., Gharpure, R., Laws, R. L., Reses, H. E., Freeman, B. D., Fajans, M., Rabold, E. M., et al. Household transmission of sars-cov-2 in the united states. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 2020.
[28] Li, C., Chen, L. J., Chen, X., Zhang, M., Pang, C. P., and Chen, H. Retrospective analysis of the possibility of predicting the covid-19 outbreak from internet searches and social media data, china, 2020. *Eurosurveillance*, 25(10):2000199, 2020.
[29] Lockey, S., Edwards, M. R., Hornsey, M. J., Gillespie, N., Akhlaghpour, S., and Colville, S. Profiling adopters (and non-adopters) of a contact tracing mobile application: insights from australia. *International Journal of Medical Informatics*, 149:104414, 2021.
[30] Lu, T. and Reis, B. Y. Internet search patterns reveal clinical course of covid-19 disease progression and pandemic spread across 32 countries. *NPJ digital medicine*, 4(1):1–9, 2021.
[31] Mayer, R. C., Davis, J. H., and Schoorman, F. D. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.
[32] McClain, C. and Rainie, L. The challenges of contact tracing as us battles covid-19. *Pew research center*, 2020.
[33] McKnight, D. H., Choudhury, V., and Kacmar, C. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3):334–359, 2002.
[34] Megnin-Viggars, O., Carter, P., Melendez-Torres, G., Weston, D., and Rubin, G. J. Facilitators and barriers to engagement with contact tracing during infectious disease outbreaks: A rapid review of the evidence. *PloS one*, 15(10):e0241473, 2020.
[35] Metlay, J. P., Haas, J. S., Soltoff, A. E., and Armstrong, K. A. Household transmission of sars-cov-2. *JAMA Network Open*, 4(2):e210304–e210304, 2021.
[36] Michael, M., Amanda, M.-F., Sara, C., Erin E, D., and Pounders, K. Health literacy and health information technology adoption: The potential for a new digital divide. *Journal of Medical Internet Research*, (10):e264, 2016.
[37] Michaud, C. Global burden of infectious diseases. *Encyclopedia of Microbiology*, page 444, 2009.
[38] National Academy for State Health Policy. State approaches to contact tracing during the covid-19 pandemic.
[39] Njozing, B. N., Edin, K. E., San Sebastián, M., and Hurtig, A.-K. " if the patients decide not to tell what can we do?"-tb/hiv counsellors' dilemma on partner notification for hiv. *BMC international health and human rights*, 11(1):1–11, 2011.
[40] NOVID. New feature: 100% anonymous contact tracing, 2020.
[41] Nurgalieva, L., Ryan, S., and Doherty, G. Attitudes towards covid-19 contact tracing apps: a cross-national survey. *IEEE Access*, 2021.
[42] Perez-Fuentes, M. D. C., Molero Jurado, M. d. M., Oropesa Ruiz, N. F., Martos Martinez, A., Simon Marquez, M. d. M., Herrera-Peco, I., and Gazquez Linares, J. J. Questionnaire on perception of threat from covid-19. *Journal of Clinical Medicine*, 9(4):1196, 2020.
[43] Peterman, T. A., Toomey, K. E., Dicker, L. W., Zaidi, A. A., Wroten, J. E., and Carolina, J. Partner notification for syphilis: a randomized, controlled trial of three approaches. *Sexually transmitted diseases*, 24(9):511–518, 1997.
[44] Peterson, R. *Constructing Effective Questionnaires*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 2000.
[45] Pew Research Center. Key findings about americans' views on covid-19 contact tracing.
[46] Pew Research Center. Mobile fact sheet.
[47] Phelan, C., Lampe, C., and Resnick, P. It's creepy, but it doesn't bother me. *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5240–5251, 2016.
[48] Qualtrics. *Qualtrics*. Available at: https://www.qualtrics.com.
[49] Rajae, T., Emilien, S., Seth M, H., Stéphanie, V., Pierre, A., Francis, G., Dominuque, R., Alexandra, R., Anne-Déborah, B., and Julien, M. Early acceptability of a mobile app for contact tracing during the covid-19 pandemic in france: National web-based survey. *JMIR mHealth and uHealth*, (7):e27768, 2021.
[50] Redmiles, E. M. User concerns & tradeoffs in technology-facilitated contact tracing. *arXiv preprint arXiv:2004.13219*, 2020.
[51] Rikard, R., Thompson, M. S., McKinney, J., and Beauchamp, A. Examining health literacy disparities in the united states: a third look at the national assessment of adult literacy (naal). *BMC Public Health*, 16(1):1–11, 2016.
[52] Sadique, M. Z., Edmunds, W. J., Smith, R. D., Meerding, W. J., De Zwart, O., Brug, J., and Beutels, P. Precautionary behavior in response to perceived threat of pandemic influenza. *Emerging infectious diseases*, 13(9):1307, 2007.
[53] SaferMe. A new approach to pandemics, 2021.

[54] Sanches, P., Tsaknaki, V., Rostami, A., and Brown, B. Under surveillance: Technology practices of those monitored by the state. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[55] Sannon, S. and Cosley, D. Privacy, power, and invisible labor on amazon mechanical turk. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[56] Shin, S.-Y., Seo, D.-W., An, J., Kwak, H., Kim, S.-H., Gwack, J., and Jo, M.-W. High correlation of middle east respiratory syndrome spread with google search and twitter trends in korea. *Scientific reports*, 6(1):1–7, 2016.

[57] Sinha, P. and Paterson, A. E. Contact tracing: Can big tech come to the rescue, and if so, at what cost? *EClinicalMedicine*, 24, 2020.

[58] Swanson, K. C., Altare, C., Wesseh, C. S., Nyenswah, T., Ahmed, T., Eyal, N., Hamblion, E. L., Lessler, J., Peters, D. H., and Altmann, M. Contact tracing performance during the ebola epidemic in liberia, 2014-2015. *PLoS neglected tropical diseases*, 12(9):e0006762, 2018.

[59] The Associated Press. Anonymous contact tracing without the tracking, 2020.

[60] The New York Times. Coronavirus in the u.s.: Latest map and case count.

[61] Thorneloe, R., Epton, T., Fynn, W., Daly, M., Stanulewicz, N., Kassianos, A., Shorter, G. W., Moll, S.-J., Campbell, M., Sodergren, S., et al. Scoping review of mobile phone app uptake and engagement to inform digital contact tracing tools for covid-19. 2020.

[62] TIME. Contact tracing apps were big tech's best idea for fighting covid-19. why haven't they helped?

[63] Trivedi, A., Zakaria, C., Balan, R., Becker, A., Corey, G., and Shenoy, P. Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–26, 2021.

[64] Uzor, S., Jacques, J. T., Dudley, J. J., and Kristensson, P. O. Investigating the accessibility of crowdwork tasks on mechanical turk. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[65] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. *Jama*, 323(11):1061–1069, 2020.

[66] Wirth, F. N., Johns, M., Meurers, T., and Prasser, F. Citizen-centered mobile health apps collecting individual-level spatial data for infectious disease management: Scoping review. *JMIR mHealth and uHealth*, 8(11):e22594, 2020.

[67] World Health Organization. Ethical considerations to guide the use of digital proximity tracking technologies for covid-19 contact tracing: interim guidance, 28 may 2020. Technical report, World Health Organization, 2020.

[68] World Health Organization. Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov), 2020.

[69] Yuen, S. Parliament: $13.8 million spent on SafeEntry, TraceTogether digital contact tracing tools. *The Straits Times*, Nov. 2020.

[70] Zakaria, C., Trivedi, A., Cecchet, E., Chee, M., Shenoy, P., and Balan, R. Analyzing the impact of covid-19 control policies on campus occupancy and mobility via passive wifi sensing. *ACM Transactions on Spatial Algorithms and Systems*, 1(1), 2022.

[71] Zhang, B., Kreps, S., McMurry, N., and McCain, R. M. Americans' perceptions of privacy and surveillance in the covid-19 pandemic. *Plos one*, 15(12):e0242652, 2020.

A   SCENARIO-BASED SURVEY: FULL QUESTIONNAIRE

Q1. **Novel coronavirus (nCov)** can be transmitted directly from a person to another.
○ True
○ False

Q2. Based on the information needs highlighted in yellow above, how willing would you be in allowing such information to be collected **automatically from your smartphone**?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
Your Identity	○	○	○	○	○
Contact Data	○	○	○	○	○
Location of places you have visited	○	○	○	○	○

Q3. Based on the information needs highlighted in yellow above, how willing would you be in allowing such information to be collected **automatically from your doctor's medical health record**?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
Your Identity	○	○	○	○	○
Contact Data	○	○	○	○	○
Location of places you have visited	○	○	○	○	○

Q4. Based on the information needs highlighted in yellow above, how willing would you be in allowing such information to be collected **automatically from your Internet browsing activities**?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
Your Identity	○	○	○	○	○
Contact Data	○	○	○	○	○
Location of places you have visited	○	○	○	○	○

Q5. Based on the information needs highlighted in yellow above, how willing would you be in allowing such information to be collected **personally from a public health official**?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
Your Identity	○	○	○	○	○
Contact Data	○	○	○	○	○
Location of places you have visited	○	○	○	○	○

Q6. Please state your most preferred way of providing the following information:

	Most Preferred Access Method
Your Identity	Select ▽
Contact Data	Select ▽
Location Visited	Select ▽

Q7. What is your experience with **novel coronavirus (nCov)**? (You may select more than one option)
☐ Personally
☐ Friends/family(close circle)
☐ None

Q8. Using the following scale, please indicate how much you agree or disagree with the following statements:

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I believe that the public health officials would act in my best interest.	○	○	○	○	○
If I required help, the public health officials would do their best to help me.	○	○	○	○	○
The public health officials are interested in my well-being, not just their own.	○	○	○	○	○
The public health officials are truthful in their dealings with me.	○	○	○	○	○
I would characterise the public health officials as honest.	○	○	○	○	○
The public health officials would keep their commitments.	○	○	○	○	○
The public health officials are sincere and genuine.	○	○	○	○	○
The public health officials are competent and effective in providing public health and safety advice.	○	○	○	○	○
The public health officials perform their role of giving public health and safety advice very well.	○	○	○	○	○
Overall, the public health officials are capable and proficient safety providers.	○	○	○	○	○
In general, the public health officials are very knowledgeable about public health and safety.	○	○	○	○	○

Q9. How concerned are you with **novel coronavirus(nCov)**?

○ Low
○ Medium
○ High

Q10. What are your concerns with sharing contact tracing information for **novel coronavirus (nCov)**?

_____

Q11. What concerns you most about an outbreak of **novel coronavirus (nCov)**?

_____

Q12. Do you believe contact tracing is important for **novel coronavirus (nCov)**? Why?

_____

B CODEBOOK FOR OPEN-ENDED RESPONSES

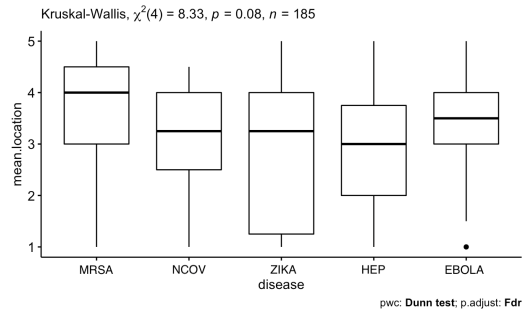| Code	Code Definition
Data Leakage/Theft	Respondent is fearful of data being leaked to unauthorised parties
Identity Theft	Respondent is concerned that their identity might be stolen
Misuse of Data	Respondent is fearful of data used for unauthorized reasons
Trust - Negative	Respondent is concerned about trust towards entity collecting the data
Trust - Positive	Respondent trusts entity collecting the data
Loss of Privacy	Respondent is concerned about loss of personal privacy
Data Confidentiality	Respondent is concerned about how the data is stored confidentially
No Concern	Respondent demonstrates no concerns about sharing contact tracing information
Surveillance	Respondent is fearful of data used for surveillance purposes
Stigma	Respondent is concerned about stigma of contracting the disease
Financial implications to self	Respondent is fearful of loss of job, income or livelihood
Financial implications to society	Respondent demonstrates concerns about widespread economic downturn due to disease
Fear of contracting disease	Respondent demonstrates fear of contracting disease
Fear of death from disease	Respondent demonstrates fear of dying from the disease
Fear of family members contracting disease	Respondent demonstrates fear of family members contracting disease
Fear of spreading disease to social circle	Respondent demonstrates fear of contracting and spreading the disease to people in social circle
No Concern	Respondent demonstrates no fear of disease
Misinformation	Respondent demonstrates concerns about misinformation due to disease
Fear of Disease Control	Respondent demonstrates fear over how well the disease can be controlled
Supportive	Respondent is supportive of contract tracing
Not Supportive	Respondent is not supportive of contract tracing
Unsure/Subjective	Respondent is unsure about effectiveness of contact tracing or would want more convincing reasons
Warns	Respondent feels that contact tracing can help to warn about virus spread
Curb Virus Spread	Respondent feels that contact tracing can help to curb virus spread
Privacy Concerns	Respondent is concerned about sharing personal data for contact tracing
Public Health benefit	Respondent feels that contact tracing is beneficial for public health
Doesn't concern individual	Respondent feels that contact tracing isn't important as the disease doesn't impact the individual or is not serious
Unlikely contraction	Respondent feels that it's unlikely to get infected.
Politics	Respondent believes that the virus can be used for political reasons
Effectiveness/Infectivity	Respondent is skeptical about effectiveness/infectivity

C PAIRWISE COMPARISON BETWEEN DISEASE GROUPS

Kruskal-Wallis, $\chi^2(4) = 8.33$, $p = 0.08$, $n = 185$

pwc: **Dunn test**; p.adjust: **Fdr**

Figure 9: Sharing location data by disease

Kruskal-Wallis, $\chi^2(5) = 10.81$, $p = 0.055$, $n = 220$

pwc: **Dunn test**; p.adjust: **Fdr**
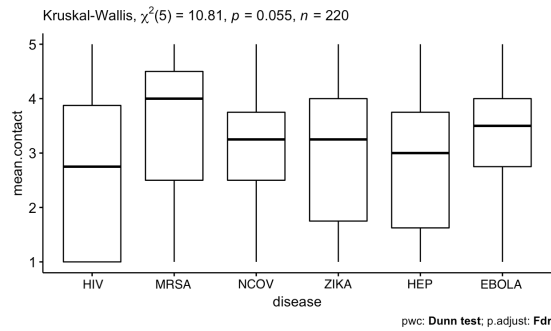
Figure 10: Sharing contact data by disease

Kruskal-Wallis, $\chi^2(5) = 9.29$, $p = 0.098$, $n = 220$

pwc: **Dunn test**; p.adjust: **Fdr**
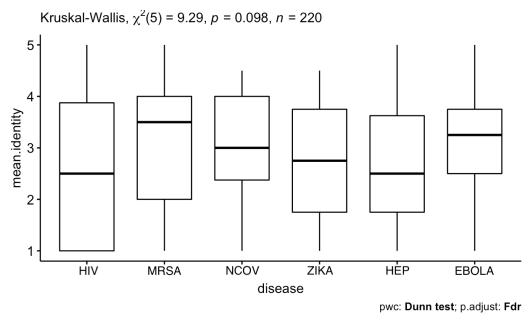
Figure 11: Sharing identity data by disease

Group 1	Group 2	n1	n2	statistic	p	p.adj	Sig
MRSA	nCov	37	43	-2.35	0.019	0.094	ns
MRSA	Zika	37	37	-1.95	0.051	0.171	ns
MRSA	Hep	37	35	-2.51	0.012	0.094	ns
MRSA	Ebola	37	33	-1.19	0.234	0.469	ns
nCov	Zika	43	37	0.329	0.742	0.777	ns
nCov	Hep	43	35	-0.284	0.777	0.777	ns
nCov	Ebola	43	33	1.05	0.296	0.493	ns
Zika	Hep	37	35	-0.586	0.558	0.697	ns
Zika	Ebola	37	33	0.703	0.482	0.688	ns
Hep	Ebola	35	33	1.26	0.206	0.469	ns

Table 11: Dunn's test p-value summary for location

Group 1	Group 2	n1	n2	statistic	p	p.adj	Sig
HIV	MRSA	35	37	2.88	0.003	0.059	ns
HIV	nCov	35	43	1.05	0.292	0.502	ns
HIV	Zika	35	37	0.958	0.338	0.502	ns
HIV	Hep	35	35	0.512	0.609	0.702	ns
HIV	Ebola	35	33	1.91	0.055	0.167	ns
MRSA	nCov	37	43	-1.96	0.049	0.167	ns
MRSA	Zika	37	37	-1.95	0.050	0.167	ns
MRSA	Hep	37	35	-2.36	0.018	0.136	ns
MRSA	Ebola	37	33	-0.900	0.368	0.502	ns
nCov	Zika	43	37	-0.062	0.950	0.950	ns
nCov	Hep	43	35	-0.516	0.606	0.702	ns
nCov	Ebola	43	33	0.969	0.332	0.502	ns
Zika	Hep	37	35	-0.439	0.661	0.708	ns
Zika	Ebola	37	33	0.996	0.319	0.502	ns
Hep	Ebola	35	33	1.41	0.159	0.397	ns

Table 12: Dunn's test p-value summary for contact

Group 1	Group 2	n1	n2	statistic	p	p.adj	Sig
HIV	MRSA	35	37	2.38	0.017	0.152	ns
HIV	nCov	35	43	1.06	0.287	0.524	ns
HIV	Zika	35	37	0.454	0.650	0.740	ns
HIV	Hep	35	35	0.055	0.956	0.956	ns
HIV	Ebola	35	33	1.73	0.083	0.282	ns
MRSA	nCov	37	43	-1.42	0.156	0.389	ns
MRSA	Zika	37	37	-1.95	0.051	0.255	ns
MRSA	Hep	37	35	-2.32	0.020	0.152	ns
MRSA	Ebola	37	33	-0.590	0.555	0.694	ns
nCov	Zika	43	37	-0.603	0.546	0.694	ns
nCov	Hep	43	35	-1.01	0.314	0.524	ns
nCov	Ebola	43	33	0.766	0.444	0.666	ns
Zika	Hep	37	35	-0.398	0.691	0.740	ns
Zika	Ebola	37	33	1.30	0.192	0.411	ns
Hep	Ebola	35	33	1.67	0.094	0.282	ns

Table 13: Dunn's test p-value summary for identity