Harvard Data Science Review • Special Issue 1 - COVID-19: Unprecedented Challenges and Chances

Individual Acceptance of Using Health Data for Private and Public Benefit: Changes During the COVID-19 Pandemic

Frederic Gerdon, Helen Nissenbaum, Ruben L. Bach, Frauke Kreuter, Stefan Zins

Published on: Apr 06, 2021

DOI: 10.1162/99608f92.edf2fc97

ABSTRACT

While the COVID-19 pandemic has been devastating, data collected in this context has unprecedented opportunities for data scientists. The stunning breadth of data obtained through new gathering systems put in place to manage the pandemic offers a richly textured view of a transformed world. Looking forward, privacy researchers worry that these new data-gathering systems risk running afoul of societal norms regarding the flow of information. Looking back at pre-pandemic public preferences with respect to data sharing may provide us some idea of what to expect in the future. In July of 2019, we happened to conduct a vignette study in Germany to examine the public's willingness to share data for fighting an outbreak of an infectious disease. In April of 2020, during the first peak of the pandemic, we repeated the study to examine crisis-driven changes in respondents' willingness to share data for public health purposes with three different samples. Public acceptance of the use of individual health data to combat an infectious disease outbreak increased notably between the two measurements, while acceptance of data use in several other scenarios barely changed over time. This shift aligns with the predictive framework of contextual integrity theory, and the data presented here may serve as a good reminder for policymakers to carefully consider the intended purpose of and appropriate limitations on data use.

Keywords: privacy attitudes, contextual integrity, COVID-19, data sharing for the public good

1. Introduction

While the COVID-19 pandemic has been devastating for individuals, global health, and the economy, it has created unprecedented opportunities for data scientists. The stunning breadth of data, collected through new systems installed to manage the pandemic, offers a richly textured window into a transformed world (e.g., COVID-19 Data Exchange, 2020). These new systems repurpose data from familiar services and platforms, such as phone companies, operating system providers, and social media platforms, and deploy them in the service of efforts to increase information about people's movements and predict the spread of COVID-19 (e.g., Apple, 2020; Google, 2020). New smartphone applications track patterns of actions relevant to the spread of disease, and people are donating data from other digital devices (e.g., data4life, 2020; Ferretti et al., 2020; O'Neill et al., 2020; Robert-Koch-Institut, 2020; Whittaker, 2020).

Predictably, and understandably, privacy researchers have thrown up red flags concerning these developments, given they will likely persist long after immediate threats pass (Morley et al., 2020; Sanfilippo et al., 2020). Researchers worry that existing norms regarding privacy and data sharing in the population are being ignored, and state the public's willingness to accept data transmission, far from signifying widespread assent to the sacrifice of privacy across the board, is, in fact, confined to specific purposes (e.g., Martin & Nissenbaum, 2016).

In recent years, the framework of "contextual integrity" (CI) (Nissenbaum, 2010, 2018) has been proposed as a rubric with which to best judge—or encourage others to judge—the conditions under which a data-handling practice is appropriate. Contextual integrity posits that data transmissions meet privacy expectations when they conform with privacy norms, contingent upon the types and circumstances of information collected, as well as the actors involved.

While we cannot predict people's future preferences with respect to sharing their data, we can gather some insights from attitudes expressed prior to the COVID-19 outbreak that may help us clarify what is at stake in this area. In the summer of 2019, we happened to conduct a vignette study in Germany, the primary purpose of which was to test public willingness to share data for a public purpose vs. a private purpose through a survey experiment. Serendipitously, one of the public purpose examples was fighting an infectious disease.

We repeated the experiment in April of 2020 during the first wave of the pandemic with three samples. Once equipped with the set of additional experimental data collections, we addressed our original questions from the 2019 study: "Are people willing to share their individual data for a public purpose or are they more willing to share their data to benefit privately?" and "Are people equally willing to share their data for a public purpose across different areas such as public health, energy consumption, or traffic infrastructure?" We addressed a new question as well: "Did the public's attitude towards sharing individual information for the purpose of promoting public health change due to the COVID-19 pandemic?" While looking back at a potential attitude shift can only provide us with suggestive insights regarding a possible post-pandemic attitude shift, such a comparison between past and future shifts, when seen through the lens of contextual integrity theory, may enrich the debate about the incorporation of sunset clauses into new technical developments for data collection.

We start out with a brief review of the contextual integrity framework, before describing the pre-COVID-19 experimental data collection, as well as our efforts to replicate the study and to collect additional data for bias assessments. After presenting cross-sectional and longitudinal analyses of the data, we discuss the political importance of this study, as well as implications for future research.

2. Contextual Integrity and Shifts in Acceptance

Technological innovation has enabled an unprecedented advance in our capacity to acquire, analyze, communicate, and disseminate data. This advance has forced us to rethink our shifting understandings of and expectations concerning privacy. The concept of privacy, of course, has a complicated history, but many contemporary accounts of privacy reflect a focus on two dominant notions: namely, privacy as control and privacy as secrecy. Given the historical background of notions of privacy (Mulligan et al., 2016), this is not surprising. Yet arguably, the venerable notions of privacy as secrecy and control fail to capture what privacy means in a world of widely adopted digital information systems.

The theory of contextual integrity (Nissenbaum, 2010) offers a new way to think about privacy in our current situation. This approach defines privacy as *appropriate flow of data* where appropriateness is a function of conformity with contextual informational norms. These norms are derived from particular social domains, or contexts, where they attain legitimacy by prescribing flows that judiciously serve stakeholder interests and promote the purposes and values of the respective social domains (Nissenbaum, 2018). Contextual informational norms prescribe flow in terms of five key parameters: (1) the sender of the information, (2) the recipient of the information, (3) the attribute or type of information, (4) the subject of the information, and (5) a transmission principle that states the condition under which the information flow is permitted.

In order to assess whether a given practice respects or violates privacy, information flows associated with that practice are described by assigning values to each of these five parameters. For example, in the health care context, it is commonly accepted that patients (sender and subject) provide their doctors (recipient) with health information (attribute) in confidence (transmission principle). A practice that generates conforming data flows is unproblematic. However, if a practice diverts medical information to a different recipient, such as a patient’s employer, a red flag is raised, even if all other factors remain the same. Equally critically, if any of the parameters is left unspecified, the description is ambiguous.

A series of empirical studies in which respondents were presented with different descriptions of data-sharing scenarios demonstrated that the approval of data sharing is contingent on situational parameters (Martin & Nissenbaum, 2017a, 2017b; Martin & Shilton, 2016). Martin and Shilton (2016), for instance, show that secondary use of tracking data for commercial purposes has a large negative impact on perceived appropriateness of data sharing, and Martin and Nissenbaum (2016) find that secondary data use driven by commercial interests meets individuals' privacy expectations less than the use of data in other contexts in which they were collected (for example, the use of information entered into a search platform to improve the search results vs. the use of this information to decide on advertisement shown when visiting other sites).

Over the past few decades, tremendous shifts in data collection practices on digital devices and online platforms have contributed to significant discontinuity between those practices and user privacy expectations. The COVID-19 pandemic adds to this misalignment, requiring quick decisions under intense conditions. Here, CI provides a useful analytic framework, allowing us to first fine-tune multiple factors influencing privacy perception and tailor necessary adjustment.

To empirically investigate the factors that influence the acceptance of data-sharing scenarios, we draw on the situational parameters suggested by CI to design descriptions of situations in which data are being shared. We focus on comparing the acceptance of public purposes and private purpose uses for different data types. Next, we provide details on our data-collection procedure and survey questionnaire.

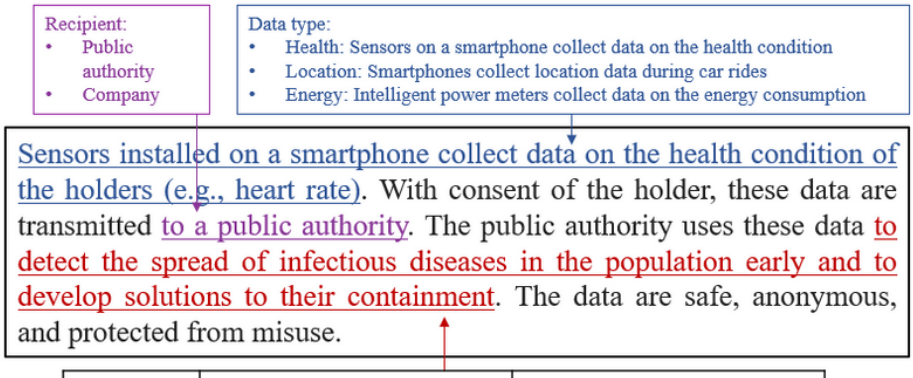
3. A Vignette Study to Measure Public's Willingness to Share Data

In 2019, we designed a vignette study or *factorial survey experiment* (Auspurg & Hinz, 2015) to experimentally test the public's willingness to share data for a public purpose vs. a private purpose. Each participant in this survey experiment was asked to rate one randomly chosen data-sharing scenario ('vignette') out of a total of twelve scenarios regarding the acceptability of data collection and use. Each scenario was followed by the question: "How acceptable is it to you to use these data for this purpose?" The answer scale had five points, ranging from 1 (Not acceptable) to 5 (Very acceptable) (see Appendix B). The answer to this question serves as the outcome in our analyses.

The descriptions presented to respondents were structured according to the theory of contextual integrity, that is, we specified values for the five key parameters (i.e., the data sender, data subject, data recipient, information type, and transmission

principles; Nissenbaum, 2018). The vignettes varied along two of the five CI parameters: the information type to be transmitted and the recipient of the data. In addition, we varied the purpose of the data. Regarding information types, we investigated health, location, and energy consumption (see Horne & Kennedy, 2017) data. The recipient was either a company or public administration. For each data type, we constructed a public purpose and a private purpose (to the data recipient) of the data. For example, the suggested purpose in the health data vignettes was either personal recommendations for health behavior (private purpose) or contribution to the containment of infectious diseases (public purpose). We held the remaining three CI parameters constant across vignettes. The sender and the data subject were referred to as an unspecified individual (e.g., the "holder" of a smartphone or the "driver" of a car). The transmission principle was described as "with consent" and defined that the "data are safe, anonymous, and protected from misuse." The focus on the parameters we experimentally varied follows our substantive interests and the practical requirement to limit the number of total vignettes. Presenting all respondents with relatively safe and cautious transmission principles should reduce effects of situation-specific privacy breach concerns.

For illustrative purposes, Figure 1 shows the survey vignette that asked about health data used by a public authority for a public health purpose (translated from German). In addition, the survey asked for respondents' age and gender, as well as information on their general privacy concerns. We also collected additional variables in the survey that we do not analyze in this article, such as the perceived sensitivity of several data types, and how much respondents trusted companies and public authorities. The latter variables were placed after the vignette in the questionnaire. The full questionnaires in English and German as well as a list of the vignettes are available in the Supplementary Materials (Appendices B and C).



Depending on data type:	Private purpose	Public purpose
Health	… personal recommendations on health behavior	…to detect outbreaks of infectious diseases early and to develop solutions to their containment.
Location	… personal recommendations on driving behavior and routes	…to develop improvements of the local infrastructure
Energy	… personal recommendations on the optimization and reduction of the own energy consumption	… to develop a more efficient energy distribution system

Figure 1. Example vignette as well as dimensions and levels of the other vignettes. The vignettes varied along the indicated data type, recipient, and data use.

4. Sample Design and Data Collection in 2019

We implemented the factorial survey experiment in a cross-sectional survey fielded from July 9 to July 18, 2019, among individuals of age 18 to 69 in Germany (*cross-section 2019*). This was the original study we designed to experimentally test public's willingness to share data for a public purpose vs. a private purpose. A total of 1,401 people[1] participated in this study and responded to all questions.

The sample for this first study was drawn from an opt-in panel maintained by respondi AG, a survey vendor that maintains a pool of individuals interested in participating in market and social research studies. Individuals registered in such panels are usually recruited through banner ads placed on websites or on social media, and participation is usually open to everyone interested. For this reason, such panels are often referred to as *nonprobability online panels*. Researchers can buy access to a sample of participants from the survey vendor and ask them questions through online surveys. The survey vendor remunerates participants who successfully complete surveys with small financial incentives.

Samples from these nonprobability online panels are often drawn using river or quota sampling, hoping that the sample will mimic the population of a country. They offer a fast, cheap, and increasingly popular method for conducting experimental studies with high internal validity (Cornesse et al., 2020). Nonprobability online panels face a number of challenges, though. For example, when interested in obtaining accurate estimates of public opinion, bias may arise because people without internet access are not covered in participant pools and because samples consist of volunteers who self-select into participation in these panels (Bethlehem, 2017). Therefore, it is difficult to infer population totals from such data without relying on strong additional and often untestable assumptions regarding the data-generating process (Kohler et al., 2019). The focus of the 2019 study was thus on comparing the acceptance of public purpose and private purpose uses for different data types, and our experimental design allows us to obtain results with high internal validity. Due to the nonprobability sample, we cannot guarantee that our findings also represent broader public opinion in Germany, that is, that they have high external validity.

Nevertheless, to achieve a sample of respondents that represents the German adult population with regard to several predefined characteristics, we selected our sample from the vendor's pool using quota sampling. Quotas were based on age and gender population benchmarks for Germany, provided by Eurostat for 2018. Quotas were applied separately and not crossed. In addition, we weighted the final analysis sample using raking (Deville et al., 1993) to population benchmarks obtained from the German micro census for 2019. Age, gender, and state were used in the weighting procedure. While weighting procedure can reduce some of the bias that arises from using a sample from a nonprobability online panel, it is likely that more factors exist that influenced participation in our study.

5. Three Additional Surveys to Study the Effect of the 2020 Pandemic

After the outbreak of the COVID-19 pandemic, we replicated the 2019 study to investigate the question we raised in the introduction (whether the public's attitude toward sharing individual information for the purpose of promoting public health changed as a result of the pandemic). For an ideal research design, we would have interviewed all of the 2019 respondents for a second time in 2020. Ignoring attrition, such a longitudinal sample would have allowed us to eliminate bias due to differences in the composition of the 2019 and the 2020 samples and to unobserved individual heterogeneity, for example, by using fixed effects regression modeling. Unfortunately,

we planned the 2019 study as a single cross-sectional survey as, at the time, the pandemic was not contemplated. Therefore, we took several sampling approaches to combat potential biases. We selected a second cross-sectional quota sample from the nonprobability online access panel that we also used in 2019. This second survey was fielded from March 31 to April 5, 2020 (*cross-section 2020*), and we collected responses from 970 respondents who were not selected for the cross-section 2019 survey. We used the same experimental survey design and asked respondents the set of questions that we described here. In order to achieve a maximum of comparability of the two surveys over time, we selected the cross-section 2020 survey with the same quotas. However, we cannot exclude the possibility that the two surveys differ in their composition as the age and gender quotas were in both surveys applied separately and not crossed. We also weighted the cross-section 2020 survey using again the raking approach, but we note that differences remain in the distribution of age and gender between the cross-section 2019 and the cross-section 2020 surveys (see Table A1 in the Appendix).

There may also be unobserved confounders that could result in bias when we use the two surveys to study change in acceptability of data collection and use between 2019 and 2020. For example, the pool of potential participants maintained by the survey vendor may have changed over time, and the factors driving individuals into participation may have changed from 2019 to 2020.

To address biases resulting from unobserved differences between the 2019 and the 2020 cross-section samples, we ran a third survey on the respondi survey platform (*longitudinal sample*). The survey vendor was able to identify and reinterview 627 participants of the 2019 survey. These respondents were still registered in the vendor's participant pool in 2020. Identification was based on unique participant IDs assigned to each participant by the vendor. We interviewed these participants for a second time in 2020, parallel to the cross-section 2020 survey using the experimental survey design and the set of questions described in the previous section. Each of these respondents received the same vignette they received in the survey of 2019. These 627 respondents who were interviewed in both 2019 and 2020 form a true longitudinal sample, which we used to assess the robustness of our analyses with respect to both observed and unobserved individual heterogeneity.

Furthermore, we collected responses to a fourth online survey that we ran with a different survey vendor (forsa) between April 2 and April 7, 2020. Forsa runs a similar online panel of participants interested in answering survey questions. The design of

the panel is, however, fundamentally different (Baker et al., 2010). Forsa panelists are originally recruited through a probability-based telephone survey. Therefore, it should be less affected by bias due to individuals self-selecting into the participant pool, but we note that it may still be affected by biases due to differential nonresponse, for example. We refer to this sample as *benchmark 2020*. We used the experimental design and the set of questions described in the previous section also in the benchmark 2020 survey.

We used a similar quota-sampling approach to select the benchmark sample ($N = 801$). Crossed age-gender quotas that mimic the German adult population were provided by forsa. We also weighted the benchmark 2020 sample using the raking procedure and the population benchmarks mentioned here.

We collected the benchmark 2020 sample to assess the robustness of the estimates obtained from the nonprobability survey cross-section 2020. While there is no guarantee that using a quota sample selected from a probability sample and weighting the data will remove bias due to, for example, differential nonresponse, using a probability-based online survey weighted to census data is backed by statistical theory that provides justification for confidence, and continuously performed well when compared to population benchmarks (Cornesse et al., 2020). Table 1 presents a summary of the characteristics of our data collections and indicates which questions we answer with each survey.

Table 1. Characteristics of the Analysis Samples

Survey:	**Cross-section 2019**	**Cross-section 2020**	**Longitudinal sample**	**Benchmark 2020**
Purpose	1. Sharing individual data for a public purpose vs. benefitting privately 2. Sharing individual data for a public purpose across data types	Changes in sharing individual data for a public purpose (public health) in response to COVID-19 pandemic	Assess robustness of results with respect to sample composition over time	Assess robustness of results with respect to sample recruitment

Field period	7/9 – 7/18 2019	3/31 – 4/5 2020	7/9 – 7/18 2019 and 3/31 – 4/5 2020	4/2 – 4/6 2020
Number of complete responses (unweighted)	1,401	970	1,254 (627 respondents)	801
Recruitment of participant pool	Quota based sample from nonprobability online access panel	Quota based sample from nonprobability online access panel	Quota based sample from nonprobability online access panel	Quota based sample from probability online panel with initial phone recruitment

6. Analytical Strategy

We use the cross-section 2019 data to answer our first research question (whether people are willing to share their data for a public vs. private purpose) and our second research question (whether people are equally willing to share data for a public purpose across different data types). We examine responses to the 5-point Likert-scale question asking for respondents' acceptance to use their data. The variable ranges from 1 ("Not acceptable") to 5 ("Very acceptable").

Our analytical strategy to answer the third research question (whether the public's attitude toward sharing individual information for the purpose of promoting public health changed due to the COVID-19 pandemic) is inspired by the difference-in-differences (DiD) approach (Wooldridge, 2010, ch. 6). DiD is a popular technique for evaluating policy interventions in economics and in the social sciences. DiD designs require four groups (see Figure 2). First, a treatment group measured prior to treatment, and second, a control group measured prior to treatment. Third, we need a treatment group measured *after* it was treated and, fourth, a control group that did not get the treatment but was also measured *after* treatment was given to the treated.

We think of the pandemic as the treatment, therefore, the cross-section 2019 survey as the pretreatment measurement and the cross-section 2020 survey as the post-treatment measurement. Furthermore, we think of those who were asked about health data as the treated group and those who were asked about non-health data as the

control group. The rationale for this is that the health data vignettes described scenarios directly related to the pandemic (sharing health data for personal health behavior recommendations and the detection of an outbreak of an infectious disease), while the non-health data vignettes described scenarios completely unrelated to the pandemic (e.g., sharing data for improving energy-saving measures). We assume that the pandemic influenced privacy attitudes related to health data while leaving attitudes toward sharing other data types mostly unchanged. Of course, it is possible that the pandemic also affected attitudes toward sharing other data types. However, we assume that such effects should be much smaller than the effect of the pandemic on sharing health data.

We apply the same logic to our analysis of the question of whether the pandemic affected respondents' acceptance of health-data sharing *for public purposes*. In two of the four health vignettes, we described a scenario where the transmitted data were used for a public purpose. Specifically, we asked how acceptant respondents were of transmitting their health data to help "detect outbreaks of diseases early and to develop solutions to their containment" (see above section for details). We treat these two scenarios as the treated conditions in our analysis of change over time.

The control group conditions are restricted to the two health-data–sharing scenarios with a private purpose ("provide the holders with personal recommendations on their health behavior"). These did not mention a public health crisis. It is not unlikely that the pandemic also affected control-group participants' data-sharing attitudes as the vignette mentioned recommendations on health behavior. However, we assume that the pandemic had a larger effect on participants' acceptance to share health data for public purposes. That is, we restrict the data to those respondents who answered a health vignette with either public or private purpose (cross-sectional samples: $N = 784$, longitudinal sample: $N = 203$ per wave).

In the traditional DiD logic, we are interested in comparing the difference between the mean outcome of the pretreatment treatment and control groups with the difference in the mean outcomes of the treatment and control groups after treatment has been assigned. Thereby, pretreatment differences between the treatment and the control groups will be removed from post-treatment comparisons of the treatment and control groups.

The key assumption for our design is the parallel trends assumption. That is, we need to assume that had there been no treatment (i.e., had there been no pandemic), the outcomes of the treatment and the control groups would have evolved similarly. In

other words, we need to assume that there is no event in Germany between 2019 and 2020 that changed attitudes toward *only one* data type (health data but not non-health data and public purpose but not private purpose and vice versa). In addition, we need to assume that the two cross-sectional samples are truly comparable such that we can attribute any difference in privacy attitudes between the treatment and the control groups in 2020, after adjusting for differences observed between the two groups in 2019, to the pandemic alone. Figure 2 illustrates the idea of the design.

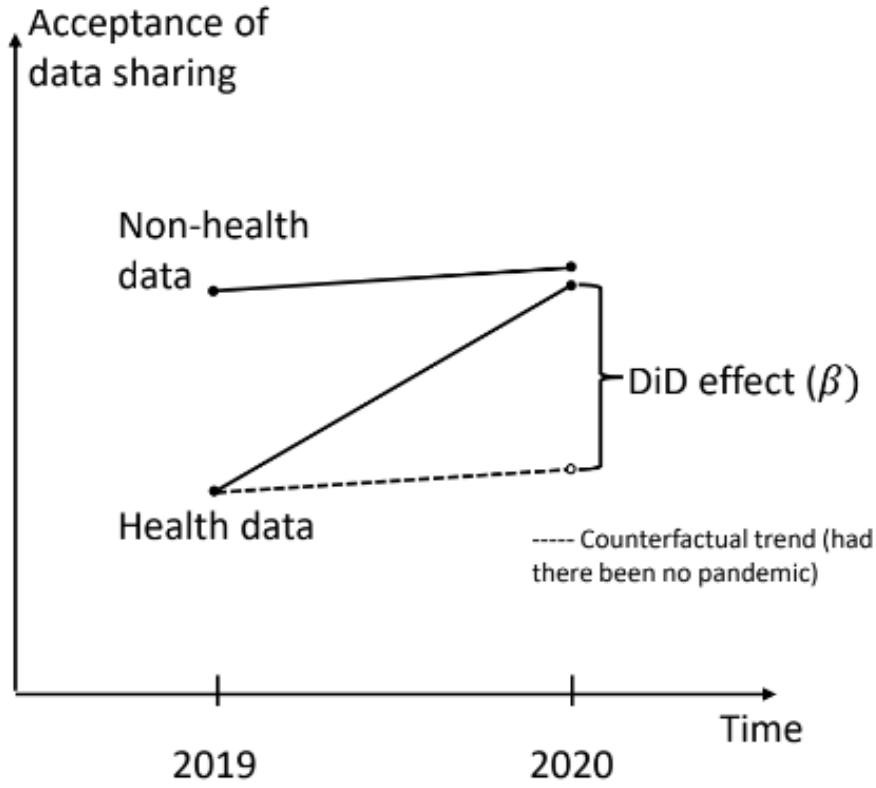


Figure 2. Difference-in-differences (DiD) identification strategy. Schematic representation of a mean comparison.

With continuous outcomes, the DiD effect is defined as the difference between the means of the control group outcome and the treatment group outcome *after* treatment has been assigned, subtracted from the difference between the means of the control group outcome and the treatment group outcome *before* treatment has been assigned (Wooldridge, 2010, ch. 6). Athey and Imbens (2006) and Yamauchi (2020) used DiD-like procedures for discrete outcomes for simple random samples. To avoid further assumptions on our outcome variable (treating it as continuous) and to allow for the

proper use of survey weights, we conduct a series of Kolmogorov–Smirnov (KS) tests for two discrete samples following the logic described above. The KS test is a nonparametric test that does not require the estimation of standard errors for the test statistic. This is an advantage, as it would be difficult to infer the distribution of most statistics of interest under our survey estimation strategy. Since the distribution of the test statistic of a KS test is also unknown for weighted survey data, we implemented a KS permutation test. We simulate the distribution of the test statistic under the null hypothesis (the data from the two samples are independent and identically distributed, e.g., there is no effect of the pandemic) and we implement the following. In a first step we resample the observations in each sample proportional to their respective weights by sampling from a list of indices. Each index of the list corresponds to one sample element and one element only and is repeated proportional to the weight of the element it corresponds to. Random unbiased rounding is used to coerce noninteger weights into integers. In a second step the indices selected in step 1 are completely randomly permuted. In a third step we calculate the KS test statistic as the maximum distance between the empirical cumulative distribution function (ECDF) of the values corresponding to the first n_1 indices and the last n_2 indices, where n_1 and n_2 are the sizes of the two resamples selected in the first step. Steps 1 to 3 are repeated 1,000 times. We then calculate the proportion of the KS test statistics, calculated in step 3, that are larger than the test statistic based on the original samples and our survey weights. This proportion is the *p*-value for our (one-sided) test. Because the permutation test may tend to reject a null hypothesis too easily for small sample sizes, we compare our test results with those of a more conservative KS test where we estimate the ECDFs using our survey weights. The *p*-values for these tests are obtained from the theoretical distribution of the KS test statistic for two simple random samples. Numerical examples showed that this simple random sample assumption resulted in consistently more conservative *p*-values than with the permutation test. We use these conservative KS tests as robustness checks for our test decisions based on the permutation test.

For our analysis, we use the software *R* (R Core Team, 2020) with the packages *ggpubr* (Kassambara, 2020), *gridExtra* (Auguie, 2017), *sampling* (Tillé & Matei, 2021), *scales* (Wickham & Seidel, 2020), *srvyr* (Ellis & Schneider, 2020), *survey* (Lumley, 2020), *tidyverse* (Wickham, 2017), and *viridis* (Garnier, 2018). All analyses report weighted estimates.

7. Results

In this section, we describe the empirical findings from our four surveys. We first present results from the cross-section 2019 survey and answer the questions regarding differences in sharing data for a public vs. private purpose and sharing data for a public purpose across data types. Second, we report descriptive findings of changes in sharing individual information for a public purpose (public health) in response to the COVID-19 pandemic before turning to results of the KS permutation tests. We conclude this section with several sensitivity and robustness analyses.

7.1. Contextual Integrity Matters for Acceptability of Data Transmission

Figure 3 presents acceptance levels for each data type by recipient (public agency vs. private company) and use (public vs. private purpose) using the weighted cross-section 2019 data. We show mean values to provide a quick and simple descriptive impression of the results, while the distributions for all groups are shown in the Appendix (Tables A2 and A3, Figures A2a–e). We find clear evidence that context matters when individuals judge the appropriateness of data transmission. Overall, respondents find the use of health data less acceptable than the use of location or energy data. Furthermore, the figure shows that respondents find it equally acceptable but often more acceptable to transmit data to a company than to a public authority or agency. However, transmission of data seems also to depend on the intended use of the data. Individuals find it in many scenarios more appropriate to transmit data for private purpose to a company than to a public agency. Regarding sharing individual data for a public purpose vs. sharing such data for private benefit, we do not find a consistent pattern across data types.

Looking at each data type separately, we find some evidence that individuals deem it more acceptable to transmit health data for a private purpose (here, personal recommendations on health behavior) to a company than to transmit health data to a public authority or agency for a public purpose (containment of infectious diseases). In fact, rather strikingly, transmitting health data to a public agency for a public purpose is least accepted. For location data, individuals find it equally acceptable to transmit data for a public purpose (here, develop improvements of the local infrastructure) to an agency or a private company. Transmitting data to an agency for a private purpose (personal recommendations on driving behavior and route) is least accepted. Regarding energy data, differences do not seem as pronounced. It seems that only transmitting data to an agency for a private purpose (personal recommendations on optimization of energy consumption) is less accepted than the other scenarios.

Therefore, regarding differences in sharing data for a public purpose vs. benefitting privately, we find a strong dependency on data type, but also on the recipient of the data.

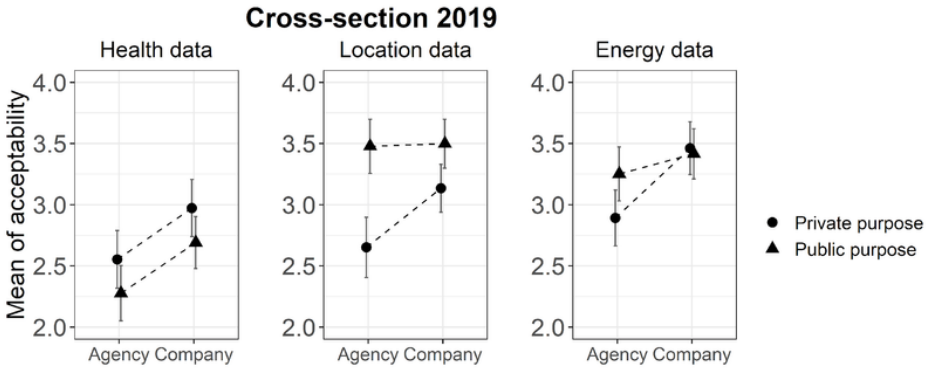


Figure 3. Mean acceptability of different data transmissions, depending on data type, data use, and recipient of the data. Vertical bars indicate 95% confidence intervals. N = = 1,401. Weighted analysis.

7.2. Longitudinal Analysis and the Effect of the Pandemic on Sharing of Health Data

Next, we compare the distribution of the outcome variables over time and between the groups defined in Section 6. The top row of panels in Figure 4 shows that acceptance to transmit data changed for both health and non-health scenarios from 2019 to 2020. Overall, respondents were more likely in 2020 to judge the transmission of health data as acceptable. This effect seems to be mainly driven by fewer respondents choosing the extreme category "1 – Not acceptable" in 2020 than in 2019. At the same time, respondents found it less acceptable to transmit non-health data over time. The KS permutation tests indicate that both changes over time are statistically significant ($p <$ .05, see rows three and four in Table A4 in Appendix A). The more conservative KS tests indicate insignificant differences in both cases. Visual inspection of the distributions suggests that the change in health data over time is much more pronounced than the change in non-health data over time, however.

The longitudinal sample confirms this finding (Figure 4, bottom row). With this sample, differences between change in health data over time and the change in non-health data are even more pronounced. Transmitting health data became more acceptable, while

transmitting non-health data did not change much. Here, the KS permutation tests indicate that the change over time for health data is statistically significant, while it is not statistically significant for non-health data (rows seven and eight in Table A4 in Appendix A). The conservative KS tests confirm these findings. It is likely that the results obtained with the longitudinal sample are more accurate, as the two cross-sectional samples differ in their compositions while the longitudinal sample does not (see Section 5).

Looking at changes over time *within* health data, we find that the increased levels of acceptance we reported are mainly driven by increased acceptance to share health data for a *public* purpose. Respondents chose the lowest acceptance category less often and the two highest categories more often for public purpose health data (Figure 5, top, right panel). At the same time, visual inspection suggests that sharing health data for a private purpose changed to a much smaller degree and in the opposite direction. Indeed, our KS permutation tests show that the change over time in acceptance to share health data for a public purpose was significant, while it was not significant for private purpose health data (rows 11 and 12 in Table A4 in Appendix A). Overall, these findings are supported by the longitudinal sample. Sharing health data for a public purpose was more accepted in 2020 than in 2019, while sharing health data for a private purpose changed to a smaller degree. This is confirmed by the KS permutation tests, which indicate significant changes over time for a public purpose but not for a private purpose (rows 13 and 14 in Table A4 in Appendix A). The conservative KS tests confirm the findings for both groups.

As we discussed, our research design is inspired by the DiD approach. Therefore, one would ideally net out the change over time in the non-health data / private purpose scenarios (our control groups) from the change in the health data / public purpose scenarios (our treatment groups) over time to adjust for baseline shifts. Given that we find substantial changes over time for health data and public purpose health data, respectively, but only mild shifts for non-health data and private purpose health data, we are confident that the findings reported here would also hold when controlling for baseline shifts.

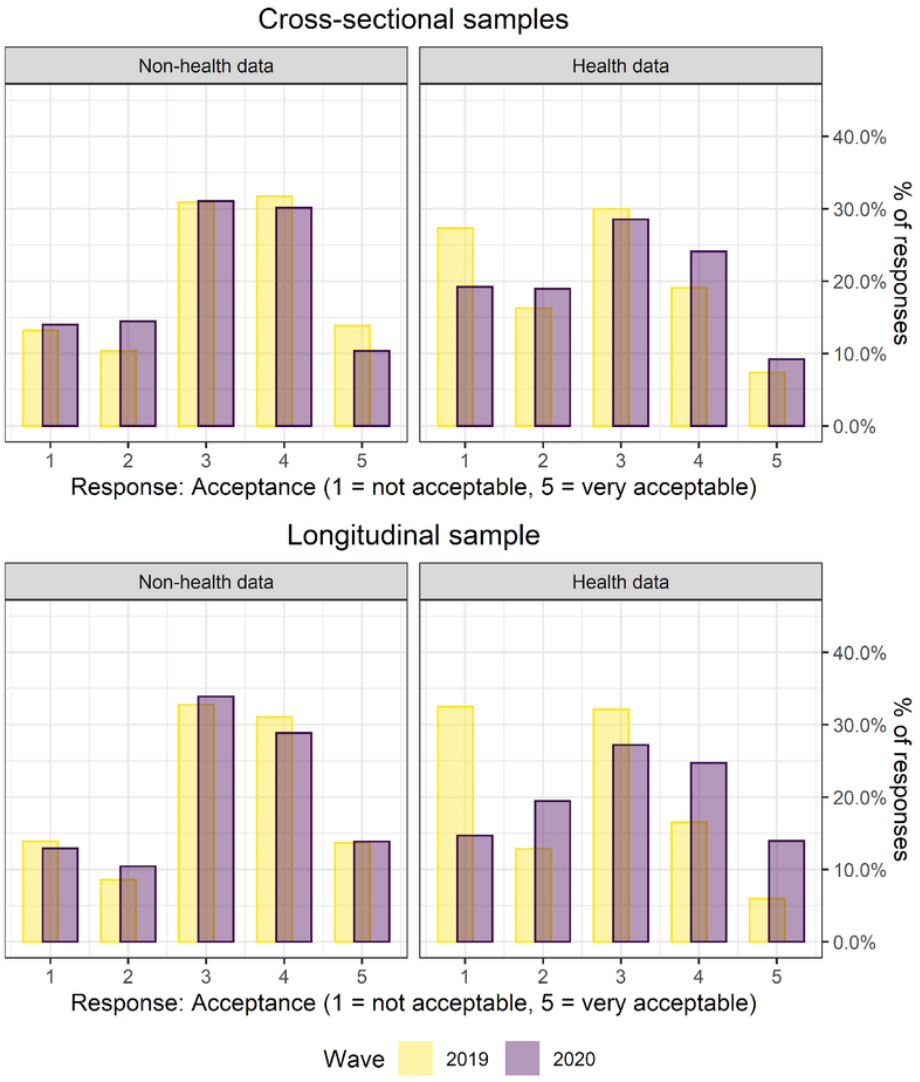


Figure 4. Relative frequency of acceptance for respondents shown health or non-health vignettes, by wave. Cross-sectional samples: $N = 2{,}371$. Longitudinal sample: $N = 627$ per wave. Weighted analysis.

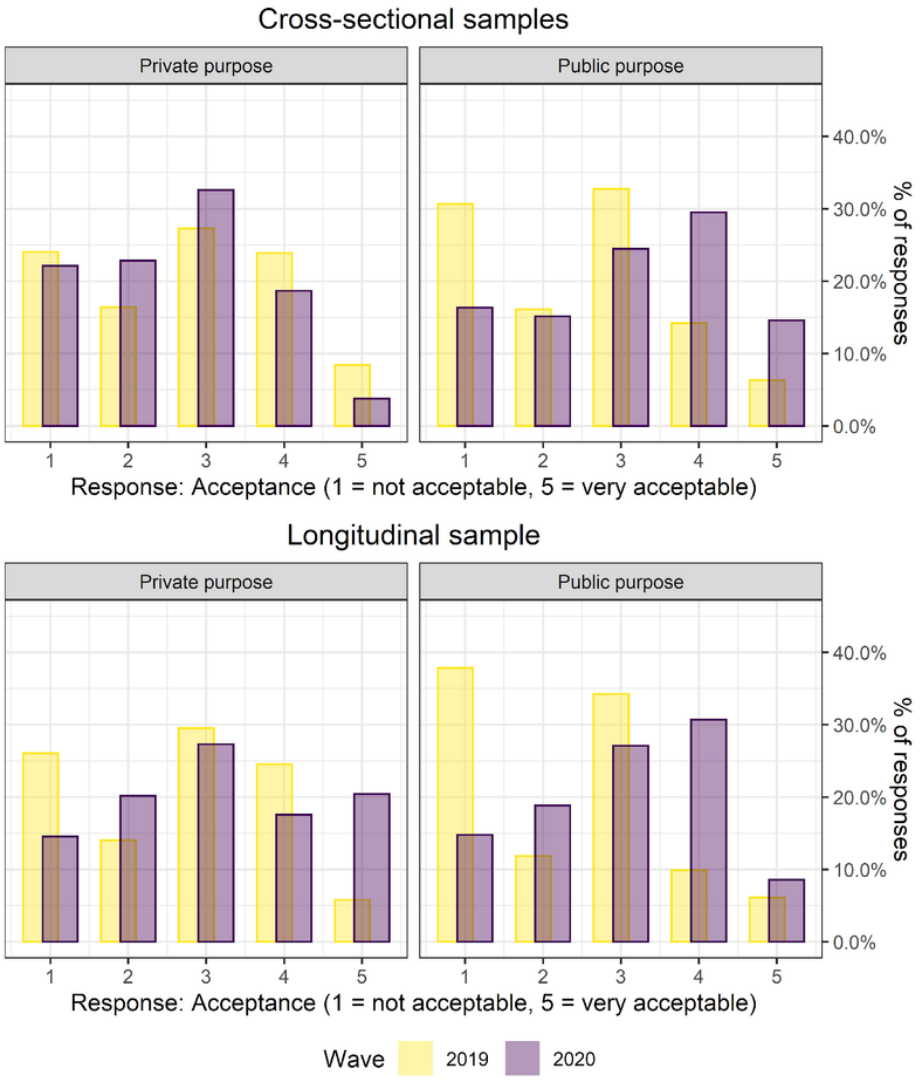


Figure 5. Relative frequency of acceptance for respondents shown a health vignette with a public purpose or a health vignette with a private purpose, by wave. Cross-sectional samples: $N = 784$. Longitudinal sample: $N = 203$ per wave. Weighted analysis.

For the longitudinal sample, we additionally test differences in the number of respondents who changed or did not change their answer from 2019 to 2020. That is, we calculate how many respondents chose a lower response category in 2020 than in 2019, how many did not change their answer, and how many chose a higher response category (Figure 6). We then compare the distributions of these three categories (lower in 2020, same answer, higher in 2020) between respondents who answered to a health data scenario and respondents who answered to a non-health data scenario

using the KS permutation test. In addition, we conduct this test for the comparison between private purpose health data sharing and public purpose health data sharing. Note that it is not possible to run these analyses with the cross-sectional samples, as we do not observe the same respondents in the two samples.

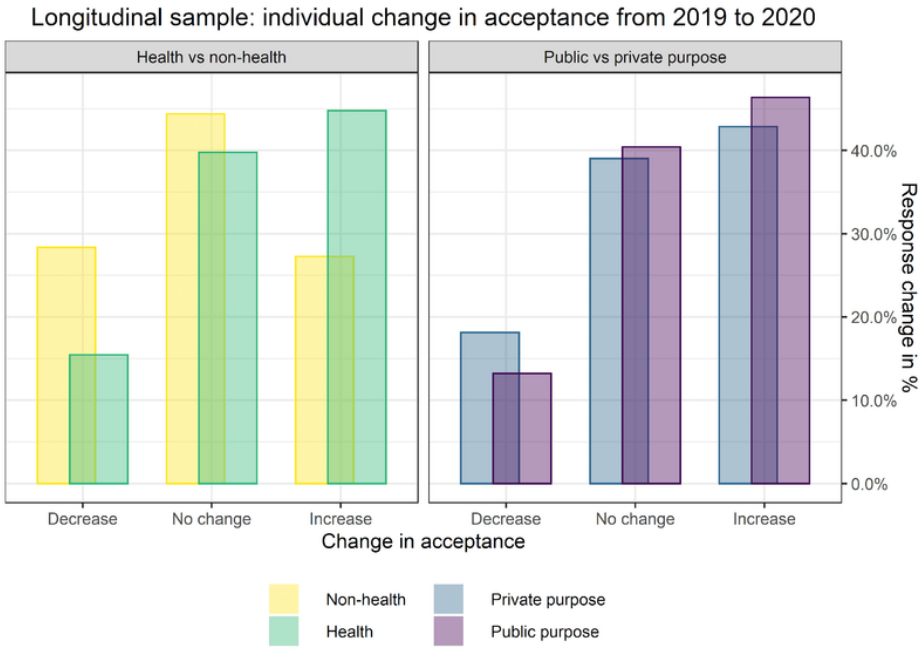


Figure 6. Changes in response category chosen by the respondents from 2019 to 2020 in the longitudinal sample. Left panel: Cross-sectional samples: $N = 2{,}371$. Longitudinal sample: $N = 627$. Right panel: Cross-sectional samples: $N = 784$. Longitudinal sample: $N = 203$. Weighted analysis.

The left panel of Figure 6 shows a clear pattern: the share of respondents choosing a higher acceptance category in 2020 than in 2019 is much larger for health vignette respondents than for non-health vignette respondents. Vice versa, the share of respondents choosing a lower acceptance category in 2020 than in 2019 is much smaller for health vignette respondents than for non-health vignette respondents. The KS permutation test also indicates that the distributions are in fact different between health vignette and non-health vignette respondents ($p = 0$). Regarding differences between the change in acceptance to share public purpose health data and private purpose health data, the right panel of Figure 6 shows a similar pattern. The share of respondents changing their response toward a more favorable answer in 2020 compared to 2019 is higher among public purpose respondents. At the same time, the

share of respondents who chose a less favorable answer in 2020 than in 2019 is higher among private purpose respondents than among public purpose respondents. The differences between the two groups are less pronounced than those between health and non-health vignette respondents, and our KS permutation does not indicate that the distributions are different in a meaningful way ($p = 1$). The conservative KS tests confirm the results of permutation tests for both cases of public and private purpose use of health data.

8. Discussion

When we first designed this study, we set out to empirically investigate the factors that influence the acceptance of data-sharing scenarios through a survey experiment and by drawing on the situational parameters suggested by CI theory. One of the most striking results of this experiment is that individuals in Germany perceive the sharing of health data with a public agency, irrespective of a private purpose or a public purpose, as least acceptable among a series of data types. With this result in mind, the signs for public support of tracking, predicting the spread of, and fighting a pandemic like COVID-19 with data on people's movements and contacts but also information about their health were far from positive.

It may be possible that, back then, the idea of a pandemic such as COVID-19 with its devastating consequences for individuals, global health, and the economy was too abstract for individuals to fully evaluate the potential benefits that sacrificing some privacy might generate. Amid the influence of the COVID-19 pandemic, public opinion toward the acceptability of sharing health data for private purpose but also for a public purpose changed, resulting in increased levels of acceptability. That is, we may conclude that individuals judge the flow of information for fighting a public health crisis as more appropriate when both the devastating consequences of a public health crisis but also the benefits of sharing data become apparent.

We should be careful when considering the question of whether individuals will judge the flow of information as equally appropriate once the pandemic has ended. We suspect, from looking back at pre-pandemic times, it is likely the public's judgment of the appropriateness may decrease again. Future work should replicate our data collection as the pandemic proceeds and eventually ends. Moreover, future data collections may be designed to study additional questions such as whether individuals' judgment of appropriate data flows is a function of the severity of the pandemic. In addition, more work will be needed to learn whether and how increased levels of

acceptance during exceptional times might generalize to other contexts and, more interestingly, to circumstantial changes that might suggest shifts in expectations.

From a policy perspective, our analysis and application of contextual integrity theory suggest the need to reevaluate practices post-pandemic. For these reasons, we call for government policymakers, software developers, and the general public to pay attention to the contextual purposes served by given data practices (sometimes enabled by technical systems) and be ready to adapt data use and storage policies accordingly.

However, we also need to consider that our findings and the implications discussed here are derived from a study that, originally, was never intended to include a longitudinal perspective. In 2019, we could not anticipate that a pandemic would change circumstances in such meaningful ways that we would run a second survey just a few months after the original 2019 study. As a result, several limitations arise. First, we observe that there are differences in the compositions of the two cross-sectional surveys. Although both samples were selected from the same survey platform and with the same specifications, our quota sampling specifications did not cross age and gender quotas but applied them separately, resulting in differences in age and gender compositions. We addressed these differences by weighting both cross-sectional samples to population benchmarks obtained from the German micro census. Unfortunately, weighting could not remove all differences between the two samples. In addition, our analyses rely on the assumption that had there been no pandemic, outcomes of the health data scenarios and the non-health data scenarios would have evolved in a similar way. Unfortunately, we can neither test this assumption itself nor assess its plausibility by, for example, analyzing temporal leads of the outcome variable (see, e.g., Autor, 2003).

Second, it is likely that there are additional unobserved differences between the two cross-sectional samples that may bias our analyses of change in the outcome over time. We did not collect information beyond respondents' age, gender, and state. Since we already observe that there are differences on these two observed confounders, it is likely that additional (unobserved) variables could also differ between the two samples, thereby biasing our analyses of change in data-sharing acceptance.

We addressed these differences by identifying a true longitudinal sample of respondents interviewed in both 2019 and 2020. In general, results obtained with this sample point in a similar direction as the results obtained from the two cross-sectional samples.

Regarding the size of the effects identified, we note that the shift in acceptance to transmit data is small. However, this is not completely unexpected as other studies investigating, for example, the public's willingness to install apps developed to facilitate the tracing of potentially infected people find high levels of support for such apps, but a fair number of individuals not willing to use such apps due to privacy concerns (see, e.g., Altmann et al., 2020). Moreover, uptake of such apps in various countries indicate that actual use of such technologies is likewise far from universal (Mosoff et al., 2020).

Overall, our results indicate a favorable shift toward the idea of using individuals' data for efforts designed to fight the COVID-19 pandemic. This is good news for data scientists and the public health system if these attitudes translate into a high rate of access to the data needed to address the crisis. Whether these attitudes prevail over the course of the pandemic and beyond will be interesting to watch, and we hope research will continue as well. In the meantime, however, public policymakers and researchers should keep in mind that the public's approval of these activities is limited to specific contexts and purposes.

Disclosure Statement

This research was partially funded by the Volkswagen Foundation: "Consequences of Artificial Intelligence for Urban Societies," as well as the Deutsche Forschungsgemeinschaft (DFG, project numbers 396057129 and 139943784, SFB 884). This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences. Supporting H Nissenbaum, we gratefully acknowledge US National Security Agency (The Science of Privacy: Implications for Data Usage, H98230-18-D-006) and US National Science Foundation (SaTC: CORE: Medium: Collaborative: Contextual Integrity: From Theory to Practice, CNS-1801501).

Acknowledgments

We thank the editor, Xiao-Li Meng, Stephanie Eckman, Felix Henninger, Christoph Kern, Florian Keusch, Pascal Kieslich, Johannes Ludsteck, Sonja Malich, Ido Sivan-Sivilia, and Patrick Schenk for helpful comments on earlier versions of this paper, and Ann Sarnak, Suzanne Smith, and Jason McMillan for editing help.

References

Altmann, S., Milsom, L., Zillessen, H., Blasone, R., Gerdon, F., Bach, R., Kreuter, F., Nosenzo, D., Toussaert, S., & Abeler, J. (2020). Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. *JMIR mHealth and uHealth, 8*(8), Article e19857. https://doi.org/10.2196/19857

Apple. (2020). *Mobility trends reports*. https://covid19.apple.com/mobility

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica, 74*(2), 431–497. https://doi.org/10.1111/j.1468-0262.2006.00668.x

Auguie, B., (2017). *gridExtra*: Miscellaneous functions for "grid" graphics (R package version 2.3) [Computer software]. R Foundation. https://CRAN.R-project.org/package=gridExtra

Auspurg, K., & Hinz, T. (2015). *Factorial survey experiments*. SAGE. https://doi.org/10.4135/9781483398075

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics, 21*(1), 1–42. https://doi.org/10.1086/344122

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D. A., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., & Zahs, D. (2010). AAPOR report on online panels. *Public Opinion Quarterly*, *74*(4), 711–781. https://doi.org/10.1093/poq/nfq048

Bethlehem, J. G. (2017). *Understanding public opinion polls*. CRC Press Taylor & Francis Group. https://doi.org/10.1201/9781315154220

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., de Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., & Wenz, A. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology, 8*(1), 4–36. https://doi.org/10.1093/jssam/smz041

COVID-19 Data Exchange. (2020). *Support & contribution*. https://www.covid19-dataexchange.org/support-contributors

data4life. (2020). *COVID-19 survey*. https://www.data4life.care/en/corona/pulsecheck/

Deville, J. C., Särndal, C. E., & Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, *88*(423), 1013–1020. https://doi.org/10.1080/01621459.1993.10476369

Ellis, G. F., & Schneider, B. (2020). *srvyr*: "*dplyr*"-like syntax for summary statistics of survey data (R package version 0.4.0) [Computer software]. R Foundation. https://CRAN.R-project.org/package=srvyr

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., & Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science, 368*(6491), Article eabb6936. https://doi.org/10.1126/science.abb6936

Garnier, S. (2018). *viridis*: Default color maps from "*matplotlib*" (R package version 0.5.1) [Computer software]. R Foundation. https://CRAN.R-project.org/package=viridis

Google. (2020). *COVID-19 community mobility reports*. https://www.google.com/covid19/mobility

Horne, C., & Huddart Kennedy, E. (2017). The power of social norms for reducing and shifting electricity use. *Energy Policy, 107*, 43–52. https://doi.org/10.1016/j.enpol.2017.04.029

Kassambara, A. (2020). *ggpubr*: "*ggplot2*" based publication ready plots (R package version 0.2.5) [Computer software]. R Foundation. https://CRAN.R-project.org/package=ggpubr

Kohler, U., Kreuter, F., & Stuart, E. A. (2019). Nonprobability sampling and causal analysis. *Annual Review of Statistics and Its Application, 6,* , 149–172. https://doi.org/10.1146/annurev-statistics-030718-104951

Lumley, T. (2020). *survey*: Analysis of complex survey samples (R package version 4.0) [Computer software]. R Foundation. https://CRAN.R-project.org/package=survey

Martin, K., & Nissenbaum, H. (2017a). Measuring privacy: An empirical test using context to expose confounding variables. *The Columbia Science & Technology Law Revie*w, *18*, 176–218. https://doi.org/10.7916/stlr.v18i1.4015

Martin, K., & Nissenbaum, H. (2017b). Privacy interests in public records. An empirical investigation. *Harvard Journal of Law & Technology, 31*(1), 111–143. http://jolt.law.harvard.edu/articles/pdf/v31/31HarvJLTech111.pdf

Martin, K., & Shilton, K. (2016). Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society, 32*(3), 200–216. https://doi.org/10.1080/01972243.2016.1153012

Morley, J., Cowls, J., Taddeo, M., & Floridi, L. (2020). Ethical guidelines for COVID-19 tracing apps. *Nature, 582*(7810), 29–31. https://doi.org 10.1038/d41586-020-01578-0

Mosoff, R., Friedlich, T., Scassa, T., Bronson, K., & Millar, J. (2020). *Global Pandemic App Watch (GPAW): COVID-19 Exposure notification and contact tracing apps*. GPAW. https://craiedl.ca/gpaw/

Mulligan, D. K., Koopman, C., & Doty, N. (2016). Privacy is an essentially contested concept: A multi-dimensional analytic for mapping privacy. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences, 374*(2083), Article 20160118. https://doi.org/10.1098/rsta.2016.0118

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press. https://www.sup.org/books/title/?id=8862

Nissenbaum, H. (2018). Respecting context to protect privacy: Why meaning matters. *Science and Engineering Ethics, 24*(3), 831–852. https://doi.org/10.1007/s11948-015-9674-9

O'Neill, P. H., Ryan-Mosley, T., & Johnson, B. (2020, May 7). A flood of coronavirus apps are tracking us: Now it's time to keep track of them. *Technology Review*. https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robert-Koch-Institut. (2020). *Corona-Datenspende*. https://corona-datenspende.de/science/en/

Sanfilippo, M. R., Shvartzshnaider, Y., Reyes, I., Nissenbaum, H., & Egelman, S. (2020). Disaster privacy/privacy disaster. *Journal of the Association for Information Science and Technology, 59*(9), 1–13. https://doi.org/10.1002/asi.24353

Tillé, Y., & Matei, A. (2021). *sampling*: Survey sampling (R package version 2.9) [Computer software]. R Foundation. https://CRAN.R-project.org/package=sampling

Whittaker, J. (2020, September 22). Data from your FitBit could help predict COVID: Research suggests wearable devices could help in virus fight. *Cayman Compass*. https://www.caymancompass.com/2020/09/22/data-from-your-fitbit-could-help-predict-covid

Wickham, H. (2017). *tidyverse*: Easily install and load the "*tidyverse*." (R package version 1.2.1) [Computer software]. R Foundation. https://CRAN.R-project.org/package=tidyverse

Wickham, H., & Seidel, D. (2020). *scales*: Scale functions for visualization (R package version 1.1.1) [Computer software]. R Foundation. https://CRAN.R-project.org/package=scales

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Yamauchi, S. (2020). *Difference-in-differences for ordinal outcomes: Application to the effect of mass shootings on attitudes toward gun control*. ArXiv. https://arxiv.org/abs/2009.13404

Appendices

pdf	AppendixA_Tables_and_Figures.pdf	1 MB
pdf	AppendixB_Questionnaires.pdf	281 KB
pdf	AppendixC_Vignettes.pdf	67 KB

indicated with respect to particular material included in the article. The article should be attributed to the authors identified above.

Footnotes

1. All sample sizes refer to those respondents who responded to all questions and for which we have information on all weighting variables. ↩