Original Paper

US Privacy Laws Go Against Public Preferences and Impede Public Health and Research: Survey Study

Cason Schmit[1*], JD; Theodoros Giannouchos[1,2], PhD; Mahin Ramezani[1], BS; Qi Zheng[3], PhD; Michael A Morrisey[1], PhD; Hye-Chung Kum[1*], PhD, MSW

[1]Population Informatics Lab, Department of Health Policy and Management, Texas A&M University, College Station, TX, United States

[2]Pharmacotherapy Outcomes Research Center, College of Pharmacy, University of Utah, Salt Lake City, UT, United States

[3]Department of Epidemiology and Biostatistics, Texas A&M University, College Station, TX, United States

[*]these authors contributed equally

Corresponding Author:
Cason Schmit, JD
Population Informatics Lab
Department of Health Policy and Management
Texas A&M University
1266 TAMU
College Station, TX, 77843
United States
Phone: 1 (979)436 0277
Email: schmit@tamu.edu

Abstract

Background: Reaping the benefits from massive volumes of data collected in all sectors to improve population health, inform personalized medicine, and transform biomedical research requires the delicate balance between the benefits and risks of using individual-level data. There is a patchwork of US data protection laws that vary depending on the type of data, who is using it, and their intended purpose. Differences in these laws challenge big data projects using data from different sources. The decisions to permit or restrict data uses are determined by elected officials; therefore, constituent input is critical to finding the right balance between individual privacy and public benefits.

Objective: This study explores the US public's preferences for using identifiable data for different purposes without their consent.

Methods: We measured data use preferences of a nationally representative sample of 504 US adults by conducting a web-based survey in February 2020. The survey used a choice-based conjoint analysis. We selected choice-based conjoint attributes and levels based on 5 US data protection laws (Health Insurance Portability and Accountability Act, Family Educational Rights and Privacy Act, Privacy Act of 1974, Federal Trade Commission Act, and the Common Rule). There were 72 different combinations of attribute levels, representing different data use scenarios. Participants were given 12 pairs of data use scenarios and were asked to choose the scenario they were the most comfortable with. We then simulated the population preferences by using the hierarchical Bayes regression model using the ChoiceModelR package in R.

Results: Participants strongly preferred data reuse for public health and research than for profit-driven, marketing, or crime-detection activities. Participants also strongly preferred data use by universities or nonprofit organizations over data use by businesses and governments. Participants were fairly indifferent about the different types of data used (health, education, government, or economic data).

Conclusions: Our results show a notable incongruence between public preferences and current US data protection laws. Our findings appear to show that the US public favors data uses promoting social benefits over those promoting individual or organizational interests. This study provides strong support for continued efforts to provide safe access to useful data sets for research and public health. Policy makers should consider more robust public health and research data use exceptions to align laws with public preferences. In addition, policy makers who revise laws to enable data use for research and public health should consider more comprehensive protection mechanisms, including transparent use of data and accountability.

(J Med Internet Res 2021;23(7):e25266) doi: 10.2196/25266

KEYWORDS

privacy; law; medical informatics; conjoint analysis; surveys and questionnaires; public health; information dissemination; health policy; public policy; big data

Introduction

Cleaning, integrating, and managing the uncertainty in chaotic real data is essential for reproducible science and to unleash the potential power of big data for biomedical research. This often requires access to very detailed data that inevitably raise privacy concerns. Despite the widespread use of personal information for big data purposes (eg, marketing, intelligence gathering, political campaigns), big data analytics are still challenged in health applications owing to concerns about privacy and complex and differing federal and state laws [1,2]. The patchwork of federal and state data protection laws poses a substantial challenge to leveraging data to promote health outcomes [1,2]. Data protection laws have 5 fundamental elements: (1) a definition of protected data, (2) definition of a regulated person or entity, (3) data use or disclosure restrictions, (4) data use or disclosure exceptions, and (5) penalties for violating the law. It is common for data protection laws to vary wildly in these 5 elements [2-6]. Consequently, it can be exceptionally difficult to understand which law (or laws) apply to a data project and whether a proposed data use is permitted. Often, the only commonality between different data protection laws is that they usually protect only identifiable data. Data that do not identify a person typically are not protected by US data protection laws. However, legal definitions for “identifiable” data or deidentification standards vary considerably [6]. This inconsistency encourages highly conservative measures to strip data of potential identifiers, which can severely limit data utility [6]. This reality poses a substantial barrier to cross-sectoral and cross-jurisdictional data uses relevant to health outcomes, including exploration of social determinants of health, retrospective database research studies, informatics research on decision support systems, digital ethology, and big data analytics in health (eg, precision public health) [2,7]. These barriers challenge efforts to rapidly leverage data in public health emergencies (eg, COVID-19).

An increasing number of published stories highlight the fact that different privacy protections apply in different contexts. For example, popular news stories have addressed how health information is treated differently when it is collected by health care providers as opposed to commercial companies such as Fitbit, Apple, or Ancestry.com [8,9]. Data use by health care providers is regulated by the Health Insurance Portability and Accountability Act of 1996 whereas data use by Fitbit or Apple is regulated by the Federal Trade Commission Act, which permits data use so long as they are neither unfair nor deceptive (ie, disclosed in a lengthy privacy policy) [10].

Recent high profile breaches (eg, Equifax) and scandals (eg, Facebook and the 2016 US election) have raised awareness of these different privacy standards [10,11]. Moreover, new data protection regulations in some jurisdictions have provoked debate and congressional inquiry into new federal privacy legislation [12-15]. Any new federal privacy law will necessarily address the 5 fundamental elements of data protection laws and will inevitably address the trade-off between privacy and utility [16,17]. Privacy risk in database studies is best minimized through a holistic approach that involves security technology (eg, encryption), data manipulation (eg, differential privacy), as well as good data governance models (eg, transparency) and legal protections. Legal protections can shield against a variety of harms. Alternatively, permitting certain data use can promote social benefits, including advancing knowledge and science, promoting public health, facilitating law enforcement, and enabling economic activity [18]. In the United States, the decision to permit or restrict certain data uses is determined by elected officials. These are policy choices with significant consequences for both individual interests (eg, privacy) and public benefit. Consequently, constituent input is critical to finding the right balance between individual privacy and public benefits.

The purpose of this paper is to report on the results of a nationally representative survey examining US residents’ preferences for which of their identifiable personal data should be available for use, by whom, and for what purposes. Prior research focusing on Americans’ attitudes on data use and privacy shows strong support for socially beneficial uses such as research [19-23]. However, few US privacy laws provide specific exceptions for data uses for research or public health [1,2]. Thus, information on how the US population views certain data uses in relation to other data uses is valuable, especially if one data use is currently restricted under US laws and the other is permitted. Such data would be extremely useful to US policy makers as they deliberate new data protection frameworks.

Methods

Study Design and Recruitment

In February 2020, we conducted a web-based survey to explore the comfort levels and the preferences of the US population when individually identifiable data is reused for different purposes without their consent. Potential participants were recruited via a third-party research company (Dynata) that specializes in deploying surveys by using nationally representative sampling. We sought to balance the sample on 6 targets based on population characteristics used by the census (gender, race/ethnicity, age, education, household income, and region) where possible. Our goal was to recruit 500 adult (≥18 years) US residents fluent in English to enable reasonable sample balancing [24]. To provide a demographic context for participants’ baseline privacy concerns, we included the validated Concern for Information Privacy instrument [25,26]. The Concern for Information Privacy instrument has 15 seven-point Likert scale questions (1=strongly disagree to 7=strongly agree) and provides a composite score with 4 subscales for privacy [25]. We estimated participants’ preferences on the potential reuse of their data with a choice-based conjoint analysis [27-30]. Choice-based conjoint analysis is built on the premise that an individual places different values on an option according to its characteristics and makes trade-off choices among alternatives based on the combination of characteristics. Conjoint analysis is a decomposition method because the implicit value for a characteristic is derived from some overall score for a profile consisting of 2 or more (conjoint) characteristics [27-29,31]. Choice-based conjoint analysis is commonly used in health care and economics research to understand clinical decision making, to assess patients' preferences, evaluation, and willingness to accept new treatments and health states, and to promote shared decision making among patients and stakeholders by quantifying the relative importance that individuals place on different attributes and levels within those attributes [27-30,32-34].

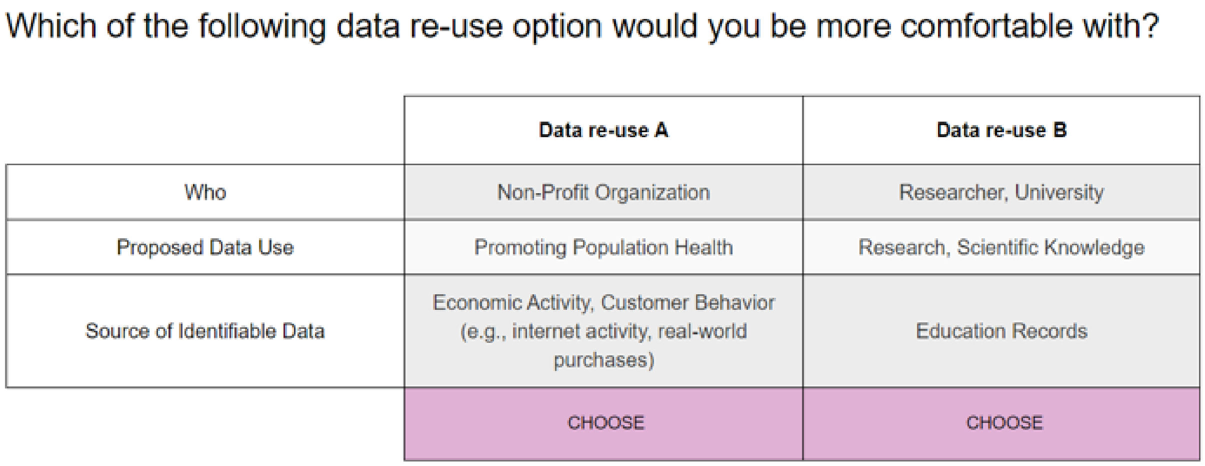
We selected attributes based on 4 of the 5 elements of the data protection laws (excluding violation penalties) (Table 1). The "source of identifiable data" is related to the definition of protected data, "who" is related to the regulated entity, and the "proposed data use" is related to 2 different elements: legal restrictions and exceptions for data use or disclosure [2]. We selected attribute levels to correspond to various legal provisions permitting or restricting data reuse and to identify the nuances within the categories (eg, business vs nonprofit organization), which resulted in 80 different scenarios comprising different attribute levels (4×4×5) [10,35-38]. Of those, we excluded 8 as implausible or likely to confuse survey respondents (eg, government or nonprofit conducting profit-driven activity), leaving 72 different scenarios.

Table 1. Attributes and levels for data reuse scenarios.

Who	Purpose	Source of identifiable data
Researcher, University	Research, scientific knowledge dissemination	Education records
Nonprofit Organization	Promoting population health	Health records
Government	Identify criminal activity	Government program or activity
Business	Marketing, recruitment	Economic activity, customer behavior
	Profit-driven activity	

Since it is not feasible and manageable to present all the possible combinations of each scenario to the participants, a fractional factorial design was used to randomly generate subsets of all the combinations, which were sufficient to obtain robust and meaningful differences in preferences through a standard web-based platform called "conjoint.ly", similar to that reported in previous work [33,34,39,40]. This resulted in 72 choice sets, with each set consisting of 12 pairs of data use scenarios that would allow for simulating participant preferences in the full space of data use scenario permutations. Each participant was randomly assigned to respond to one of the choice sets, and we asked each participant to select the data use scenario that they were the most comfortable with for each of their assigned 12 scenario pairs (Figure 1).

Figure 1. Sample pair scenario question.



Which of the following data re-use option would you be more comfortable with?

	Data re-use A	Data re-use B
Who	Non-Profit Organization	Researcher, University
Proposed Data Use	Promoting Population Health	Research, Scientific Knowledge
Source of Identifiable Data	Economic Activity, Customer Behavior (e.g., internet activity, real-world purchases)	Education Records
	CHOOSE	CHOOSE

Statistical Analysis

To estimate the parameters, we used a hierarchical Bayes regression model, and in estimating the parameters at the individual level, we generated 10,000 posterior draws by using the Markov chain Monte Carlo simulation [29,41,42]. This approach allowed the estimation of attributes and levels with the small amount of data collected from each respondent, while simultaneously accounting for the heterogeneity of preferences across and within individuals, the nested structure of the choices, and thus, the nonrandom preference variation of the respondents [29,34,41-43]. The value of 3 attributes was scaled to sum up to 100%, while the values of the levels within each attribute (part worth utilities) sum up to zero, with negative values indicating decreased and positive values indicating increased preferences. Finally, we used the sum of the estimated relative values (utilities) of different levels to identify and rank the alternative scenarios from the most to the least preferable. All measures were estimated at the individual level, which were then averaged and reported as the population mean with standard deviation in the results. Data analyses were conducted using the ChoiceModelR package in R [44]. This study was approved by our university institutional review board.

Results

The survey was distributed to 687 individuals. Of them, 22 individuals declined to participate (3.2%), 157 did not fully complete the survey (22.8%), and 4 participant responses (0.6%) were marked as low quality based on detected participant behavior (eg, rapidly clicking through without mouse movement). This resulted in 504 respondents who fully completed the web-based survey (response rate 74.4%), which was our final analytic sample. Generally, we were able to meet our census sampling targets for gender, race/ethnicity, age, education, income, and census region (Table 2). In addition, although we did not try to balance for it, our sample's health insurance coverage is similar to data published by the US Census Bureau [45]. Around half of the respondents had used a health care provider in the past year, around one-third had at least one chronic condition, and around 19.8% (100/504) of the respondents visited an emergency department in the past year. The overall privacy score was 5.8 (SD 1.1), which is consistent with the Concern for Information Privacy validation samples (ie, scores ranging from 5 to 6) [25].

Table 2. Sociodemographic data, clinical characteristics, and privacy attitude scores of the participants (N=504).

Participant characteristics	Values	Target sample percentage[a,b]
Age categories (years), n (%)		
18-24	41 (8.1)	13.1
25-34	75 (14.9)	17.5
35-44	100 (19.8)	17.5
45-54	101 (20.0)	19.2
55-64	68 (13.5)	15.6
65 or older	89 (17.7)	17.2
Gender, n (%)		
Male	224 (44.4)	48.5
Female	278 (55.2)	50.5
Other/prefer not to answer	2 (0.4)	—[c]
Race categories, n (%)		
White	315 (62.5)	63.7
African American	77 (15.3)	12.2
Hispanic	51 (10.1)	16.4
Asian	46 (9.1)	4.7
Other	15 (3.0)	3.0
Income categories, n (%)		
$20,000 or less	103 (20.4)	19.9
$20,000 to $49,999	149 (29.6)	30.6
$50,000 to $99,999	137 (27.2)	29.1
$100,000 to $149,999	67 (13.3)	12.0
$150,000 or more	48 (9.5)	8.3
Educational level, n (%)		
High school or less	172 (34.1)	32.0
Some college completed	99 (19.6)	19.0
College degree	191 (37.9)	31.0
Master’s	37 (7.3)	—
PhD/doctoral	5 (1.0)	—
Region, n (%)		
Midwest	95 (18.8)	22.0
Northeast	126 (25.0)	18.2
South	174 (34.5)	36.2
West	109 (21.6)	23.6
Health insurance coverage, n (%)[b]		
Private	169 (33.5)	64.7
Medicare	112 (22.2)	17.7
Medicaid	83 (16.5)	17.9
Uninsured	52 (10.3)	8.5
VA/TRICARE	10 (2.0)	3.6
Multiple	78 (15.5)	14.5

Participant characteristics	Values	Target sample percentage[a,b]
Any chronic condition, n (%)		
No	319 (63.3)	—
Yes	181 (35.9)	—
Use of health care provider in the past year, n (%)		
No	93 (18.5)	—
Yes	256 (50.8)	—
At least one emergency department visit in the past year, n (%)		
No	404 (80.2)	—
Yes	100 (19.8)	—
Respondent is a primary care giver for someone else, n (%)		
No	423 (83.9)	—
Yes	77 (15.3)	—
Concern for information privacy scores, mean (SD)	5.8 (1.1)	—

[a] Survey sampling targets based on census data.

[b] Insurance data were not used as the sampling target. These data show 2018 insurance statistics from the US census for survey sampling comparisons [45]. Our survey solicited mutually exclusive responses in contrast to the US census data, which do not exclude persons with multiple insurance types from these groups.

[c] Not available.

Figure 2 presents the relative importance for the different levels within each attribute. Positive values indicate preference with higher values and reveal greater importance, while negative values indicate nonpreferred levels associated with potential data reuse. Participants were most comfortable with the reuse of identifiable data if the proposed data use was intended to promote population health (10.1%, SD 11.6) or promote science or research (8.2%, SD 6.5), if the data were used by university-affiliated researchers (6.4%, SD 10.7) or nonprofit organizations (2.5%, SD 16.1), and if the source of the data included educational (2.2%, SD 11.3) or health care records (1.4%, SD 10.4). In contrast, participants were least comfortable with data reuse by businesses (–4.5%, SD 13.7) or the government (–4.3%, SD 16.8) mainly for profit-driven (–11.7%, SD 12.3) or marketing (–4.2%, SD 11) activities based on governmental (–1.7%, SD 10.1) or economic activity data (–1.8%, SD 11.4). Overall, we observed higher differences in the values between the levels of the proposed data use attribute compared to other 2 attributes, particularly with the attribute related to the source of the identifiable data.

Figure 2. Relative importance by level within each attribute in percentage (SD).

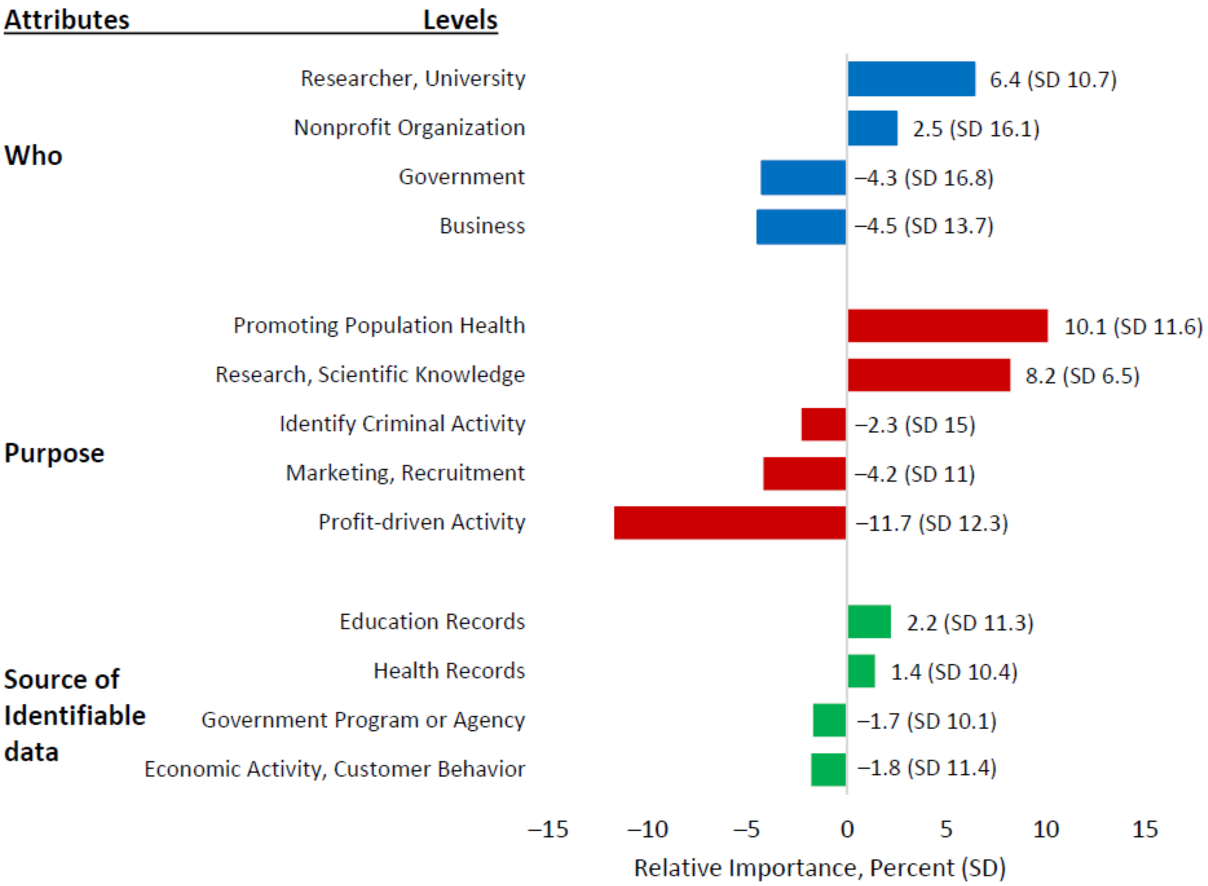


Figure 3 presents the relative importance of the different data use purposes by users. All data use activities were more preferred when conducted by universities or nonprofit organizations than when conducted by government or business. Public health and research activities received positive relative importance values regardless of who conducts the activity. Conversely, for-profit and marketing activities received near uniform negative relative importance values. Universities conducting marketing activities was the lone exception, which had a small but positive relative importance value (2.6%). The relative importance of identifying criminal activity was either positive (university and nonprofit) or negative (government and business) depending on the user. We also estimated the overall values that participants placed on each scenario. Figure 4 presents only the 10 highest and lowest ranked scenarios (Figure 4). The 4 lowest ranked scenarios all involved businesses using data for profit-driven purposes. The remainder of the lowest ranked scenarios involved business or government organizations engaging in marketing or identifying criminal activity. Eight of the 10 highest ranked scenarios involved universities/researchers engaging in scientific research or public health activities. Nonprofit organizations conducting population health programs represented the seventh and eighth highest ranked scenarios. We checked model validity by comparing the actual choices made by each participant with the estimated choices made for at least 90% of the last 50 iterations of the Markov chain Monte Carlo simulation. Precision (% of correct estimates) was good at 92% for the simulated model.

Figure 3. Public preferences for use of data by users and purpose in percentage (SD). Our survey did not pair “for-profit” purposes with government or nonprofit users because these pairings were implausible and likely to confuse survey respondents.

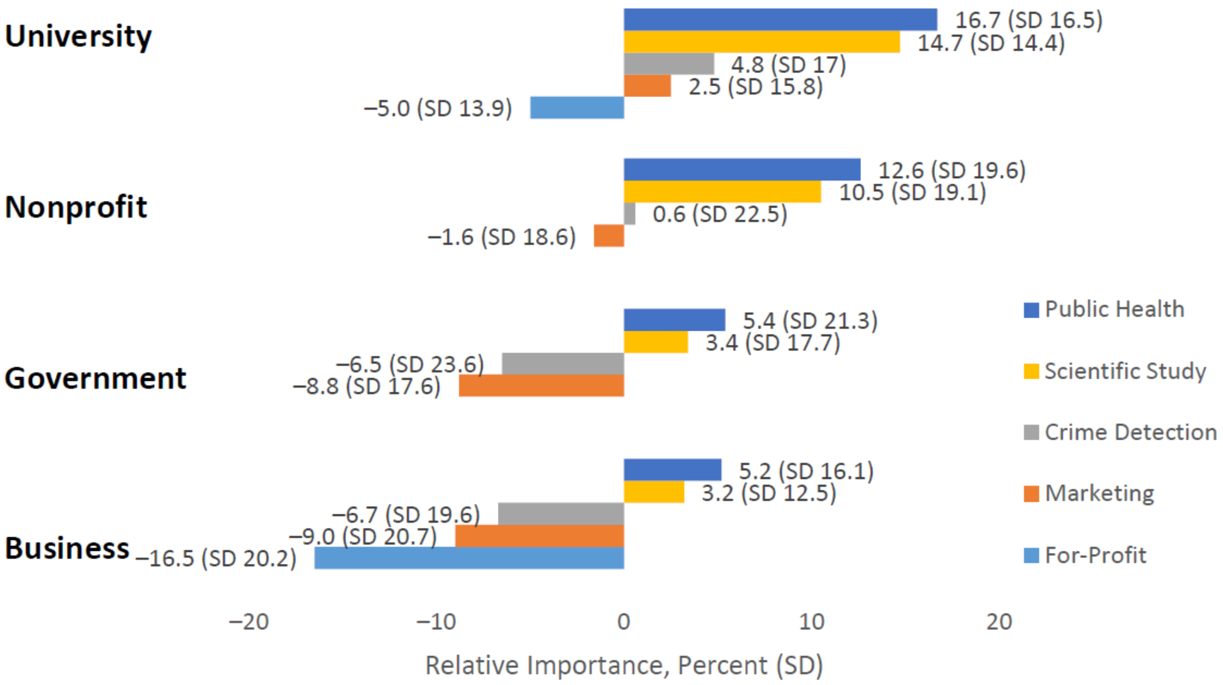
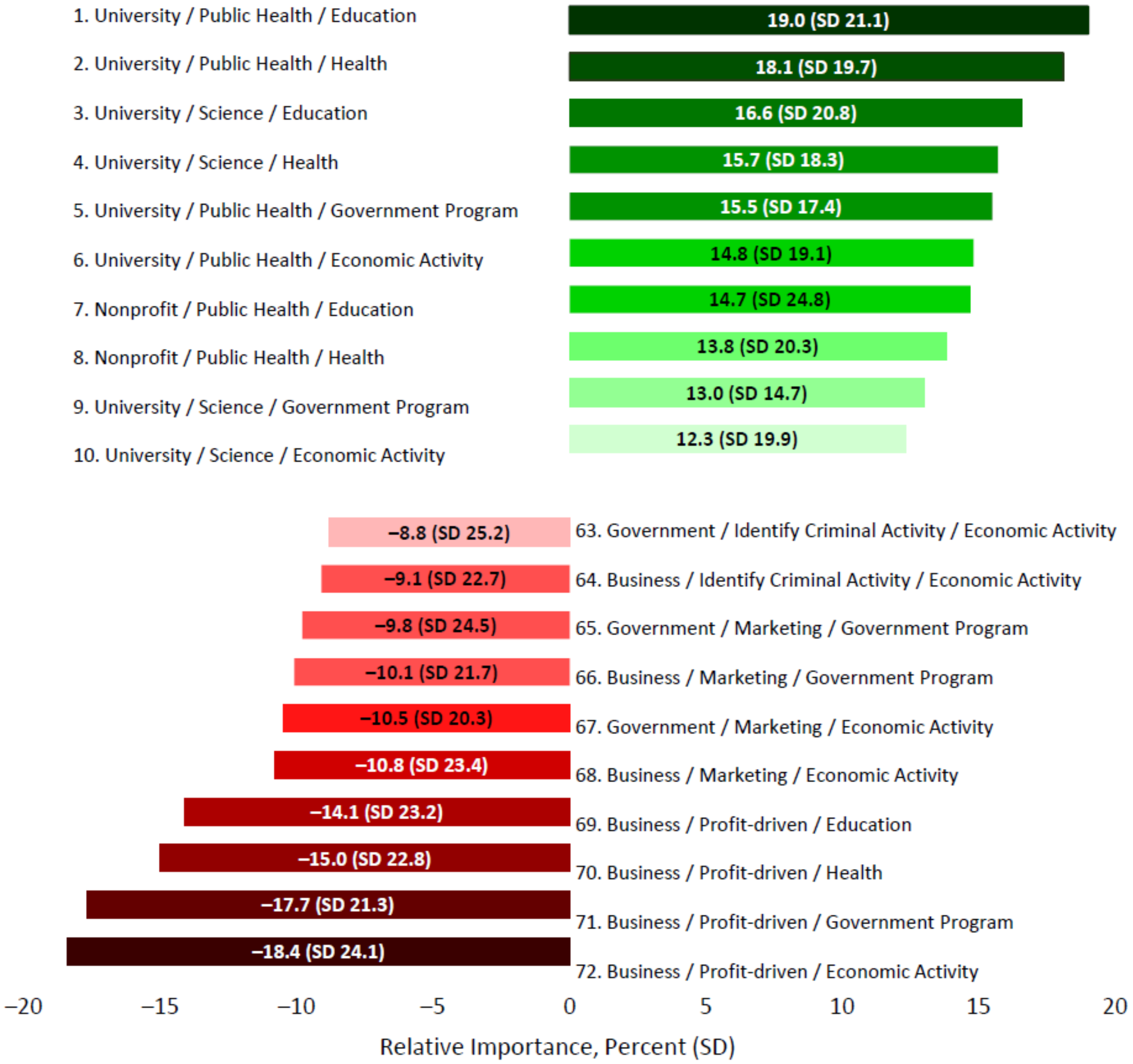


Figure 4. Top 10 and bottom 10 ranked data use scenarios derived from the sum of scenario attributes' relative values (who/use purpose/data source) in percentage (SD).



Discussion

Principal Findings

In contrast to federal and state laws, US residents make little distinctions across types of data. However, they express much more favorable preferences for uses by academic researchers and nonprofit organizations than by the government or the business community. Moreover, all types of users consistently preferred uses that focus on public health and scientific research rather than on crime detection, marketing, or for-profit activities. Our data demonstrate interesting inconsistencies between public preferences and US privacy laws. These inconsistencies are best exemplified by our participants' most preferred data reuse (researchers using education data to promote population health) and least preferred data reuse (businesses using consumer data for profit). Ironically, our data indicate that the US public's most preferred data reuse scenario is currently prohibited under the federal Family Educational Rights and Privacy Act of 1974 while the US public's least preferred data reuse is completely legal and ubiquitous under the permissive Federal Trade Commission Act [38,46]. A recent review of 22 federal data protection frameworks funded by the Network for Public Health Law indicates that few data protection laws have a general research exception and fewer have a specific exception for public health uses. However, these data uses were by far the most preferred options of those we presented to our participant and were consistently preferred, regardless of who was the data user. Yet, public health uses are treated differently under different US laws [6]. For example, the law protecting substance use treatment records has hamstrung the use of data in the present opioid epidemic response, while the laws covering cell phone location data have permitted public health officials to track entire populations in the current COVID-19 pandemic [47,48].

Our participants also strongly favored data uses by universities and nonprofit organizations. Both universities and nonprofit organizations received higher preference ratings for all data use activities when compared to those received by the government or businesses. In some cases, activities that participants viewed as heavily undesirable when conducted by the government or business (crime detection, marketing) were rated favorably when conducted by a nonprofit organization or university. In contrast, the least preferred scenarios involved data reused for profit-driven or marketing activities by businesses or government. Mistrust in government has been documented in other research on attitudes of research and is perhaps unsurprising in the present partisan political environment [21]. Negative preference ratings for businesses, profit-driven activities, and marketing are likely due to frequent stories of controversial data use, mismanagement, or breaches that are all too common in the news [49]. This finding is consistent with other research documenting strong public attitudes in favor of altruistic goals and skepticism of data uses that advance specific individual or private (ie, for-profit) interests [19,21].

We did find some preference differences for certain data types, but these differences were modest. Our data show that the public prefers the use of health or educational data (both heavily regulated under US laws) as compared to government data or economic data. Still, our data do not show any strong preferences. The public seems to view data as data. We noted that 4 of the 5 data use purposes we included in our study fall neatly into 2 broad categories: altruistic purposes and self-serving purposes. Public health and scientific purposes both ultimately contribute to the greater good, and our data suggest that these purposes are strongly preferred by the US public, regardless of who is doing the activity. In contrast, our respondents generally found those activities that are primarily self-serving (ie, profit-driven or marketing/recruitment activities) undesirable, regardless of who was doing the activity. The lone exception was marketing by universities, which received a modest positive relative importance score. Consequently, it could be that our participants based some of their preference decisions on whether they saw the data use as contributing to an altruistic or common good objective as opposed to primarily benefiting the data user's self-interests.

Identifying criminal activity was the one data use that does not neatly fit in the broad categories of altruistic or self-serving purposes. While law enforcement clearly has some social benefits (as do all the activities used in our study), identifying criminal activity implies punishment for some individuals. Consequently, it is not entirely altruistic and not entirely self-serving. Interestingly, participant preferences for identifying criminal activity seemed to vary depending on the data user. Universities and nonprofit organizations both received positive relative importance scores whereas governments and businesses received negative scores. Just as with other data uses, it could be that participants positively associate universities and nonprofit organizations with motivations more in line with social benefits rather than individual benefits.

Collectively, our results do not support the current patchwork of US data protection laws. Many US data protection laws focus primarily on the type of data (ie, health, education, governmental program data), but our respondents were fairly indifferent toward these distinctions. Instead, our findings suggest that the US public is much more interested in who is using the data and for what purposes the data are being used. In particular, our results suggest that the US public has a strong preference for data uses that promote the common good as opposed to individual or self-serving interests.

In fact, findings suggest that US preferences more closely align with a comprehensive data protection framework such as the General Data Protection Regulation enacted by the European Union where rules vary based on data use but are broadly applicable to all identifiable data [50]. For example, the General Data Protection Regulation has broad applications and express provisions permitting scientific research and activities in the public interest (eg, public health) [51,52]. Policy makers who revise laws to increase access to data for research and public health can support data protection through new security standards. A 2009 report by the Institute of Medicine, "Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health through Research" [53] argued for a different data protection approach to "enhance privacy protections through improved data security, increased transparency of activities and policies, and greater accountability." These good governance practices, as opposed to strict prohibitions on uses and disclosures (ie, for research or public health), provide a way to protect individuals while permitting big data applications (eg, linking data from different sources) with social benefits. These results provide strong public support for continued efforts to make data available for research and public health.

Limitations

There are 2 important limitations. We did not capture the universe of data use possibilities; therefore, the measured participants' preferences are relative to the 72 provided scenarios. Additionally, this design measured participants' preferences rather than acceptability, meaning that a participants' least preferred scenario could still be acceptable to them or the most preferred scenario might be unacceptable.

Conclusion

Importantly, these results support a close re-examination of the absence of public health and research data use exceptions in US laws. It is clear that the US public strongly prefers using data to promote population health (as compared to other legal data uses); yet, few laws allow this kind of exception. The Family Educational Rights and Privacy Act provides an excellent example, given that it does not have a public health exception (or a research exception that permits exploring health implications) despite being one of the most potent known social determinants of health. Moreover, the absence of these data use exceptions within the current patchwork of inconsistent US data protection laws persistently frustrates secondary database researchers and public health professionals, thereby delaying, impeding, or increasing the cost of data-intensive scientific discovery and public health practice [1,2,4,6]. These findings clearly show that there is poor alignment between the present US legal data protection framework and the preferences of the US population.

Acknowledgments

This study was funded in part by the Presidential Impact Fellow Award at Texas A&M University and the Population Informatics Lab in the School of Public Health at Texas A&M University.

Authors' Contributions

CS and HCK were responsible for conceptualization, study design and implementation, data analysis, interpretation, writing and revisions, and supervision. TG and MR were responsible for study implementation, data analysis, interpretation, writing, and revisions. QZ and MM were responsible for data analysis, interpretation, writing, and revisions.

Conflicts of Interest

None declared.

References

1. Hulkower R, Penn M, Schmit C. Privacy and confidentiality of public health information. In: Magnuson JA, Dixon BE, editors. Public Health Informatics and Information Systems. Cham, Switzerland: Springer Nature; 2020:147-166.
2. Schmit C, Kelly K, Bernstein J. Cross Sector Data Sharing: Necessity, Challenge, and Hope. J Law Med Ethics 2019 Jun;47(2_suppl):83-86. [doi: 10.1177/1073110519857325] [Medline: 31298141]
3. Begley EB, Ware JM, Hexem SA, Rapposelli K, Thompson K, Penn MS, et al. Personally Identifiable Information in State Laws: Use, Release, and Collaboration at Health Departments. Am J Public Health 2017 Aug;107(8):1272-1276. [doi: 10.2105/ajph.2017.303862]
4. O'Connor J, Matthews G. Informational Privacy, Public Health, and State Laws. Am J Public Health 2011 Oct;101(10):1845-1850. [doi: 10.2105/ajph.2011.300206]
5. Savage L. To bring health information privacy into the 21st century, look beyond HIPAA. Health Affairs Blog. 2018 Jul 05. URL: https://www.healthaffairs.org/do/10.1377/hblog20180702.168974/full/ [accessed 2020-07-15]
6. Wartenberg D, Thompson WD. Privacy Versus Public Health: The Impact of Current Confidentiality Rules. Am J Public Health 2010 Mar;100(3):407-412. [doi: 10.2105/ajph.2009.166249]
7. Kum HC, Ahalt S. Privacy-by-Design: Understanding Data Access Models for Secondary Data. AMIA Jt Summits Transl Sci Proc 2013;2013:126-130 [FREE Full text] [Medline: 24303251]
8. Elfin D. DNA testing? you might want to wait for more legal protection. Bloomberg Law. 2019 Jan 07. URL: https://news.bloomberglaw.com/pharma-and-life-sciences/dna-testing-you-might-want-to-wait-for-more-legal-protection [accessed 2020-05-25]
9. North T. Google, Fitbit, and the sale of our private health data. Health Law & Policy Brief. 2019 Nov 18. URL: http://www.healthlawpolicy.org/google-fitbit-and-the-sale-of-our-private-health-data/ [accessed 2020-05-25]
10. Federal Trade Commission Act. Federal Trade Commission. 2012. URL: https://www.ftc.gov/enforcement/statutes/federal-trade-commission-act [accessed 2021-06-14]
11. Granville K. Facebook and Cambridge Analytica: What you need to know as the fallout widens. The New York Times. 2018 Mar 19. URL: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html [accessed 2020-05-05]
12. California Consumer Privacy Act. State of California Department of Justice. 2018. URL: https://oag.ca.gov/privacy/ccpa [accessed 2021-06-14]
13. European Commission. General Data Protection Regulation. Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0249&qid=1623685068702 [accessed 2021-06-14]
14. Cameron D. FTC members unanimously press Congress for tough national privacy protections. Gizmodo. 2019 May 08. URL: https://gizmodo.com/ftc-members-unanimously-press-congress-for-tough-nation-1834624258 [accessed 2020-05-21]
15. Mulligan S, Freeman W, Linebaugh C. Data protection law: An overview. Congressional Research Service. URL: https://crsreports.congress.gov/product/pdf/R/R45631 [accessed 2020-05-06]
16. Dinur I, Nissim K. Revealing Information while Preserving Privacy. PODS '03: Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems 2003 Jun 09:202-210 [FREE Full text] [doi: 10.1145/773153.773173]
17. Ragan E, Kum H, Ilangovan G, Wang H. Balancing privacy and information disclosure in interactive record linkage with visual masking. 2018 Apr 20 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; April; New York p. 1-12. [doi: 10.1145/3173574.3173900]
18. Kum H, Krishnamurthy A, Machanavajjhala A, Ahalt SC. Social Genome: Putting Big Data to Work for Population Informatics. Computer 2014 Jan;47(1):56-63. [doi: 10.1109/mc.2013.405]
19. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]

20. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
21. Haga S, O'Daniel J. Public perspectives regarding data-sharing practices in genomics research. Public Health Genomics 2011;14(6):319-324 [FREE Full text] [doi: 10.1159/000324705] [Medline: 21430368]
22. O'Brien EC, Rodriguez AM, Kum H, Schanberg LE, Fitz-Randolph M, O'Brien SM, et al. Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire. Int J Med Inform 2019 Jul;127:9-17 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.04.003] [Medline: 31128837]
23. Willison D, Keshavjee K, Nair K, Goldsmith C, Holbrook AM, Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness investigators. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. BMJ 2003 Feb 15;326(7385):373 [FREE Full text] [doi: 10.1136/bmj.326.7385.373] [Medline: 12586673]
24. Siddiqui K. Heuristics for sample size determination in multivariate statistical techniques. World Applied Sciences Journal 2013;27(2):285-287. [doi: 10.1002/0471667196.ess2329]
25. Smith HJ, Milberg SJ, Burke SJ. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly 1996 Jun;20(2):167. [doi: 10.2307/249477]
26. Stewart KA, Segars AH. An Empirical Examination of the Concern for Information Privacy Instrument. Information Systems Research 2002 Mar;13(1):36-49. [doi: 10.1287/isre.13.1.36.97]
27. Ben-Akiva M, de Palma A, McFadden D, Abou-Zeid M, Chiappori P, de Lapparent M, et al. Process and context in choice models. Mark Lett 2012 May 17;23(2):439-456. [doi: 10.1007/s11002-012-9180-7]
28. Ben-Akiva M, McFadden D, Abe M. Modeling methods for discrete choice analysis. Marketing Letters. 1997. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/47225/11002_2004_Article_138116.pdf?sequence=1 [accessed 2021-06-14]
29. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
30. Ryan M, Farrar SJB. Using conjoint analysis to elicit preferences for health care. BMJ 2000 Jun 03;320(7248):1530-1533 [FREE Full text] [doi: 10.1136/bmj.320.7248.1530] [Medline: 10834905]
31. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26(8):661-677. [doi: 10.2165/00019053-200826080-00004] [Medline: 18620460]
32. Marshall D, Bridges JF, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008. Patient 2010 Dec 01;3(4):249-256. [doi: 10.2165/11539650-000000000-00000] [Medline: 22273432]
33. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics 2019 Feb;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
34. Zimmermann TM, Clouth J, Elosge M, Heurich M, Schneider E, Wilhelm S, et al. Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. J Affect Disord 2013 Jun;148(2-3):210-219. [doi: 10.1016/j.jad.2012.11.062] [Medline: 23290792]
35. 45 CFR Parts 160 and 164. HIPAA Privacy Rule. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/introduction/index.html [accessed 2021-06-14]
36. Lamdan S. Privacy act of 1974. The SAGE Encyclopedia of Surveillance, Security, and Privacy. 2018. URL: http://sk.sagepub.com/reference/the-sage-encyclopedia-of-surveillance-security-privacy/i9567.xml [accessed 2021-06-14]
37. Porter JP. The Federal Policy for the Protection of Human Subjects. IRB: Ethics and Human Research 1991 Sep;13(5):8. [doi: 10.2307/3564127]
38. Family educational rights and privacy act (FERPA). US Department of Education. 1988. URL: https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html [accessed 2021-06-14]
39. Frost N, Freeman J, Brixner D, Mort J, Clem J, Ngorsuraches S. Patients' preferences and willingness-to-pay for disease-modifying therapies. Mult Scler Relat Disord 2019 Oct;35:55-60. [doi: 10.1016/j.msard.2019.07.005] [Medline: 31323483]
40. Jaynes J, Wong W, Xu H. Using blocked fractional factorial designs to construct discrete choice experiments for healthcare studies. Stat Med 2016 Jul 10;35(15):2543-2560 [FREE Full text] [doi: 10.1002/sim.6882] [Medline: 26823156]
41. Allenby G, Rossi P. Hierarchical bayes models. In: Grover R, Vriens M, editors. The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Thousand Oaks, CA: Sage Publications, Inc; 2006.
42. Mohammadi T, Zhang W, Sou J, Langlois S, Munro S, Anis AH. A Hierarchical Bayes Approach to Modeling Heterogeneity in Discrete Choice Experiments: An Application to Public Preferences for Prenatal Screening. Patient 2020 Apr;13(2):211-223. [doi: 10.1007/s40271-019-00402-w] [Medline: 31814082]
43. Regier DA, Ryan M, Phimister E, Marra CA. Bayesian and classical estimation of mixed logit: An application to genetic testing. J Health Econ 2009 May;28(3):598-610. [doi: 10.1016/j.jhealeco.2008.11.003] [Medline: 19345433]

44. Sermas R, Colias J. ChoiceModelR. 2015. URL: https://cran.r-project.org/web/packages/ChoiceModelR/ChoiceModelR.pdf [accessed 2020-05-06]
45. Berchick E, Hood E, Barnett J. Health insurance coverage in the United States: 2018. US Census Bureau. 2019 Nov. URL: https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf [accessed 2021-06-14]
46. Solove DJ, Hartzog W. The FTC and the New Common Law of Privacy. Columbia Law Review 2014;114:583. [doi: 10.2139/ssrn.2312913]
47. Frakt AB, Bagley N. Protection or Harm? Suppressing Substance-Use Data. N Engl J Med 2015 May 14;372(20):1879-1881. [doi: 10.1056/nejmp1501362]
48. O'Sullivan D. How the cell phones of spring breakers who flouted coronavirus warnings were tracked? CNN. 2020 Apr 04. URL: https://www.cnn.com/2020/04/04/tech/location-tracking-florida-coronavirus/index.html [accessed 2020-05-25]
49. Equifax data breach settlement. Federal Trade Commission. 2020. URL: https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement [accessed 2020-05-06]
50. Bari L, O'Neill D. Rethinking patient data privacy in the era of digital health. Health Affairs Blog. 2019. URL: https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/ [accessed 2020-07-16]
51. European Parliament. GDPR article 6 (1)(e). Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2021-06-14]
52. European Parliament. GDPR articles 5 (1)(b), 89(1). Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2021-06-14]
53. Nass S, Levit L, Gostin L, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington DC: National Academies Press; 2009.

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by J Jose, P Swire; comments to author 22.11.20; revised version received 08.12.20; accepted 30.04.21; published 05.07.21

Please cite as:
Schmit C, Giannouchos T, Ramezani M, Zheng Q, Morrisey MA, Kum HC
US Privacy Laws Go Against Public Preferences and Impede Public Health and Research: Survey Study
J Med Internet Res 2021;23(7):e25266
URL: https://www.jmir.org/2021/7/e25266
doi: 10.2196/25266
PMID: