Proceedings of the 55th Hawaii International Conference on System Sciences | 2022

What Makes Health Data Privacy Calculus Unique? Separating Probability from Impact

Mark Keith
Brigham Young University
mark.keith@gmail.com

Autumn Clark
Brigham Young University
autumnpclark@gmail.com

Tamara Masters
University of Utah
tamara.masters@eccles.utah.edu

Curtis Wigington
Adobe
wigingto@adobe.com

Abstract

Patient health data is heavily regulated and sensitive. Patients will sometimes falsify data to avoid embarrassment resulting in misdiagnoses and even death. Existing research to explain this phenomenon is scarce with little more than attitudes and intents modeled. Similarly, health data disclosure research has only applied existing theories with additional constructs for the healthcare context. We argue that health data has a fundamentally different cost/benefit calculus than the non-health contexts of traditional privacy research. By separating the probability *of disclosure risks and benefits from the* impact *of that disclosure, it is easier to understand and interpret health data disclosure. In a study of 1590 patients disclosing health information electronically, we find that the benefits of disclosure are more difficult to conceptualize than the impact of the risk. We validate this using both a stated and objective (mouse tracking) measure of patient lying.*

1. Introduction

Information disclosed over the Internet via websites, mobile applications, and other Internet-connected devices is the source of incredible benefit and risk to consumers. The mobile application market alone is expected to reach over $366 billion by 2027 [1]. The valuable personal information disclosed through mobile apps (and many other Internet-connected sources) motivated at least 1,923 data breaches in 2020 including 37 billion records compromised [2].

As a result, regulators have become increasingly active in passing legislation to help consumers maintain the value derived from disclosing sensitive information while reducing the risk that it will be exposed. For example, the Health Insurance Portability and Accountability Act (HIPAA) is arguably one of the earliest and most detailed types of regulations regarding information privacy [3].

HIPAA was designed, in part, to prevent insurance providers from using certain patient medical history data to charge higher premiums. HIPAA is executed in the features provided by electronic health records (EHR) systems that store and transfer patient health data among clinicians who need it to determine accurate diagnoses and provide appropriate care. There is a real need for accurate health information disclosure since it is used quite literally in "life or death" situations. However, some have argued that health data is also the most sensitive type of personal information since it can lead to great embarrassment [4, 5], increased insurance costs [3], and even to patients being held "hostage" when life-giving medical devices such as pacemakers are hacked [6]. These risks cause many patients to lie to their healthcare providers about their health histories which can lead to misdiagnoses [4, 7]. Misdiagnoses are the most frequent type of medical error and are estimated to cause up to 80,000 deaths in the United States each year [8].

While some researchers have attempted to explain patient attitudes toward EHR systems that request their personal information [9], the adoption of medical devices that collect personal information [10], and willingness to disclose health data [11] and allow viral contact tracing [5], the theoretical contributions available from the general information privacy literature [12, 13] have not been well-applied to the healthcare domain. For example, existing research does not measure true patient health disclosure in the presence of actual perceived risks, but instead utilizes hypothetical scenarios or measures intentions or attitudes only. Similarly, patient health data disclosure in realistic scenarios has not been examined even though the importance of true disclosure data has been highlighted in general information privacy research [14, 15]. Therefore, it is more likely that existing research is biased towards outcomes that patients *believe* they would adopt rather than what they would *actually do* when their health is on the line. This bias is, perhaps, the greatest threat to valid information privacy research [13, 16, 17].

Another opportunity left by existing research is that the theories used to explain consumer information disclosure are often "dragged and dropped" into the patient health data context. Although researchers typically add new variables relevant to the healthcare

URI: https://hdl.handle.net/10125/79926
978-0-9981331-5-7

domain of interest, by and large, most have not explored the assumptions of these theories to posit how the fundamental relationships might change. To address this opportunity, we return to the "roots" of information disclosure theory which explain our progression from beliefs to attitudes to intentions to behaviors: the *theory of planned behavior* [18].

To demonstrate our theory, we developed and validated a revised scale measuring privacy calculus theory [19] in which the perceived probability of both disclosure risks and benefits is distinguished from their respective perceived impact. We test this scale in a unique methodology where participants are deceived (with IRB approval) to believe they are participating in a medical study about indicators of depression and anxiety in which they will need to provide certain data concerning their medical history. Unaware of the true purpose of the study, participants believe that real data is necessary and, thus, believe there is a legitimate level of risk that they could experience embarrassment if their data is misused. Therefore, we measured actual information disclosure as opposed to attitudes or intentions alone. Furthermore, in addition to a stated response, we used a novel approach to measure lying that has recently been validated in IS research based on mouse cursor movements [20].

In summary, we find that the likelihood of benefits of accurate information disclosure (e.g. improved mental health) are more difficult to conceptualize and perceive—causing them to have a lower effect on accurate information disclosure than traditionally found in non-health related contexts. As a result, the perceived impact of risks (e.g. embarrassment) plays a larger role in the healthcare context relative to non-health related contexts.

2. Theoretical Background

2.1. Privacy Calculus

To aid in the risk/reward battle, academic research has progressed with a variety of theories explaining why, or with whom, consumers decide to share their personal information. Each theory is useful in specific contexts but limited in others. In general, these theories can be separated into one of two categories. The first are those based on choice theory [21] which assumes that individuals are rational actors who prefer outcomes with greater net value. Their behaviors and choices involve a tradeoff (a.k.a. "calculus") between the expected benefit minus the cost of decision outcomes. *Privacy calculus* theory [19, 22] is the adaptation of rational choice theory in the Internet information privacy context. This has been the dominant theory used to explain consumer information disclosure decisions.

Figure 1 visualizes privacy calculus theory. There are three to five primary constructs included in most studies based on this theory. In its most basic form, perceived disclosure privacy risks (the costs) reduce intention to disclose accurate information while perceived disclosure benefits increase intention. Dinev and Hart [19] were among the first to develop this theory and included two other covariates: trust in the transaction partner and privacy concerns. The latter refers to the more general privacy concerns about all companies (often defined using [23]) while the former refers to the specific privacy risk associated with the transaction partner in context. Although trust is a unique construct, it is often omitted when using a brief measure of privacy calculus theory because of its high correlation with perceived disclosure risk.

Besides rational choice theory [21], privacy calculus theory is also based on the *theory of planned behavior* (TPB) [18] which posits that behavioral intentions lead to actual behavior which is omitted from the original model.

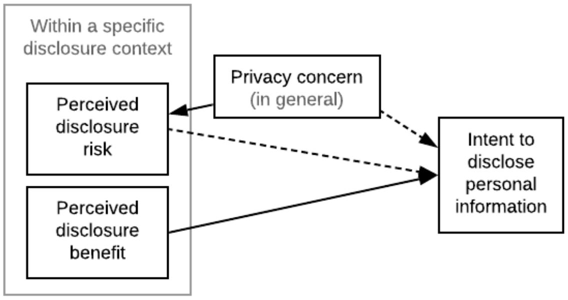
Within a specific disclosure context
Perceived disclosure risk
Perceived disclosure benefit
Privacy concern (in general)
Intent to disclose personal information

Figure 1. Core Privacy Calculus Theory

2.2. Alternatives to Privacy Calculus

Despite its usefulness and accuracy in many contexts, privacy calculus has also been criticized for not explaining why many consumers who claim to be very concerned about privacy still decide to disclose seemingly large amounts of personal data for very small rewards [16]. Termed the "privacy paradox", this criticism has led to the second general category of information privacy research: providing various explanations of "bounded" rational behavior. These studies [e.g., 14, 17, 24, 25-27] adapt various theories from behavioral economics intended to explain various phenomena of consumers behaving in ways that seem irrational.

For example, *prospect theory* [28, 29] is used to explain how information disclosure behavior is not consistent at all levels of disclosure risk and reward [14, 24, 27]. An extension of prospect theory, *hyperbolic discounting* [30] is used to explain how the temporal order of when information disclosure costs versus benefits are realized can impact a consumer's information disclosure rationality [14, 25]. Others have explained this paradox simply as an *illusion of control* [31] where consumers do make a rational decision based on what they believe to be true and their behavior is only irrational because they do not understand the extent of the risks. Lastly, of note, another explanation comes from theory on the *elaboration likelihood* model of persuasion [32] where illogical information disclosure decisions are made essentially because consumers are "lazy" information processors who are either too busy or distracted at the time of disclosure to pay attention or research the full extent of privacy risks [33, 34].

Each of these alternative theories have merit and clearly the information disclosure phenomenon is a complex and multi-faceted issue that is influenced at many levels including general concerns, the specific disclosure context, and psychological state of the discloser. However, the rational choice-based privacy calculus theory may be more efficacious than it seems by taking a closer look at the perceived disclosure risk and benefit constructs.

2.3. Taking a Closer Look at TPB

As mentioned above, we can learn more from a (re)visit of the TPB and its companion *theory of reasoned action* (TRA) [35]. These theories explain that an individual's attitude and perceptions, subject norms, and perceived behavioral control lead to intentions and actual behaviors. From this lens, the perceived risks and benefits of disclosure represent attitudes and perceptions referred to in TPB/TRA.

Ajzen [18] notes that our attitudes and perceptions regarding a specific task are largely based on our confidence that we can realize or accomplish that task. If we are confident that an intended behavior will yield the outcomes we are interested in, then that attitude will have a greater impact on our intentions and behavior. This theorizing was based on Bandura [36] concept of self-efficacy which explains how our confidence in achieving a task explains why we try harder (a.k.a. "cope") to achieve that task.

However, TPB/TRA is based on important boundary conditions. The one most relevant to our point: the effect of perceptions on intentions and behaviors depends on the degree of specificity of those perceptions [37]. In other words, the easier it is to conceptualize or perceive the specific coping path from the intended behavior to the intended outcome, the more likely that our perceptions will lead to intentions and behaviors.

To understand this assumption, consider a scenario where a consumer decides to download a mobile application to get restaurant recommendations. After installing the app, the consumer is prompted to disclose data that will be used to personalize the app and improve its value. This opportunity to transact with the provider offers both a risk/cost and a benefit. Based on our prior arguments, the easier it is for the consumer to conceptualize the specific risks and benefits of this transaction, the more likely those perceptions will accurately explain their intentions and behaviors. In other words, if the consumer cannot determine whether the app will provide her with good or bad restaurant recommendations, then her perceived disclosure benefits will have a weaker effect on her intention to disclose her data. Similarly, if she cannot accurately determine the level of privacy risk associated with the disclosure, then her perceived disclosure risk will have a weaker effect on her behavioral intention.

This effect is not hard to find in existing privacy calculus research. Consider the study by Keith, et al. [38] who examined whether subjects would disclose personal data to a mobile app designed to give personalized recommendations. The authors attempted to deceive participants into believing that they were being recruited to beta test an app being developed by a local software company rather than the researchers. Because of this uncertainty, subjects had a difficult time assessing the true nature of the privacy risk. On the other hand, it was much easier for subjects to assess whether the app would be beneficial because it provided clear screens demonstrating the app features. As a result, perceived disclosure risk had only a very small effect on actual information disclosure relative to perceived disclosure benefits.

A review of privacy calculus research generally supports this reasoning. The effect of perceived risks and benefits varies depending on their respective specificity. Mobile app disclosure studies often provide screenshots of the app but little information about the privacy risk—causing perceived risk to have a smaller effect.

2.4. Separating Probability from Impact

There are several ways to demonstrate the importance of perception specificity and health data provides a nice juxtaposition to more commonly studied disclosure contexts. To add specificity to the cost and benefit perceptions of information disclosure, we can break each evaluation into its two fundamental parts: probability and impact [28]. Current privacy calculus measurements combine these concepts in benefit and cost perceptions. Table 1 lists some sample measures used in prior research [15, 39]:

Table 1. Sample Risk and Benefit Measures

Perceived Disclosure Benefit
The [mobile app name] can provide me with personalized services tailored to my activity context.
The [mobile app name] can provide me with more relevant information tailored to my preferences or personal interests.
The [mobile app name] can provide me with the kind of information or service that I might like.
Perceived Disclosure Risks
Providing [mobile app provider] with my personal information would involve many unexpected problems.
It would be risky to disclose my personal information to [mobile app provider].
There would be high potential for loss in disclosing my personal information to [mobile app provider].

Based on economic risk theory, the perceived disclosure benefit is a combination of both the probability of that benefit being obtained and the impact of that benefit. For example, a mobile app that provides driving directions has a high probability, but low impact compared to a dating app which has a lower probability and higher impact of a benefit. By measuring probability and impact distinctly, it is possible to explain intentions and behaviors with greater accuracy, and therefore find fewer outcomes that appear to be "irrational." Therefore, we specify a privacy calculus model with these details as visualized in Figure 2.

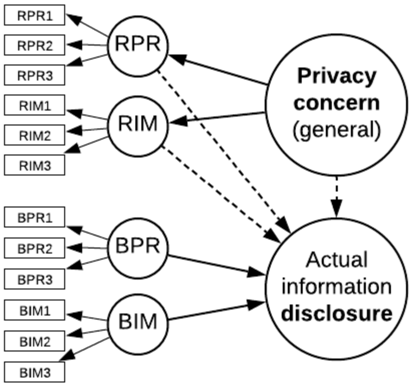

Figure 2. Expanded Privacy Calculus

Again, the key distinction of this model from traditional privacy calculus is that perceived disclosure privacy risks are formed by a combination of both risk probability (RPR) and risk impact (RIM). Perceived disclosure benefits are based on benefit probability (BPR) and benefit impact (BIM). Although we model risk and benefit as second-order formative constructs, the first-order sub-constructs are still reflective as in prior research [15, 19, 39]. Finally, based on prior research [15], we replace the intent to disclose information with actual information disclosure to eliminate the possibility of the privacy paradox confounding results [16]. We note that "actual" information disclosure scopes out the inclusion of false information disclosure which does not provide the same risks and benefits of factual data.

2.5. Health Data Hypotheses

The theoretical model in Figure 2 is intended to be more specific than traditional privacy calculus which has the added benefit of allowing us to infer how health data is unique from other contexts. The primary risks of disclosing personal health information include embarrassment, pain (of treatment), and financial loss (due to insurance premiums) [4, 7]. The probability of these risks may vary and still be difficult to conceptualize. However, whereas the mobile app context leaves the discloser with no idea of what risks she may face, in the healthcare context, a patient has a more concrete idea of whether they will face embarrassment or pain—making the perception more specific and more strongly related to actual disclosure.

Furthermore, by law, patients are informed of their HIPAA rights before every treatment making them more aware of the risk probabilities and thus, also increasing the specificity of the construct. In summary, we hypothesize the overall effect of perceived privacy risk based on privacy calculus theory and the distinct effects of probability versus impact based on the specificity assumption of TRA/TPB:

H1a: Perceived disclosure risk probability (RPR) decreases actual information disclosure

H1b: Perceived disclosure risk impact (RIM) decreases actual information disclosure

Conversely, relative to more general disclosure contexts, the perceived disclosure benefits of health information are likely less specific. For example, downloading an app that guides its user through a city or theme park would have benefits that are relatively easy to conceptualize and predict. On the other hand, disclosing health background data to a clinician is intended to help them provide better care and the discloser to eventually have improved health. However, that improvement is not as immediate as a park directions app and many other factors influence the likelihood that they will get better health like the competency of the clinician and willingness of the discloser to adhere to clinician recommendations.

Therefore, while the probability of receiving health benefits may be difficult to conceptualize, the impact of those benefits would still be similarly more difficult to conceptualize depending on the healthcare context. Similar to H1, based on privacy calculus theory, we hypothesize:

H2a: Perceived disclosure benefit probability (BPR) increases actual information disclosure

H2b: Perceived disclosure benefit impact (BIM) increases actual information disclosure

In summary, we model and hypothesize a more specified version of privacy calculus and explore how the healthcare context may differ from traditional contexts, such as mobile apps, where the effect of perceived disclosure benefits tend to outweigh perceived disclosure risks [15, 39].

3. Methodology

To test our theoretical model, we needed to create a realistic scenario where participants would believe that they needed to disclose factual health data in the presence of real risk. Hypothetical scenarios or disclosure intentions are prone to bias and more generally lacking in information privacy research [13] and health privacy research in particular. Therefore, with IRB approval, we modified consent to deceive participants into believing we were a team of clinicians and medical researchers studying mental health and that we wanted to measure their depression and anxiety to understand how their personal health history was related.

Data for our experiment was collected through a survey instrument posted on Mechanical Turk (MTurk), which received nearly 1600 responses. Survey participants were compensated $1.50 for a completed survey. The downside of this design is that participants did not have a personal interaction (allowing the potential for direct embarrassment) with any clinician. However, this was deemed acceptable to prevent the participants from knowing that they were actually participating in an academic study about information privacy. Submitting health history in an electronic format is commonplace in today's EHR systems [40].

The survey was created in Qualtrics to feel like a medical intake form as opposed to an academic research study by removing any reference to the sponsoring university and eliminating the consent form.

3.1. Measurement

Respondents initially answered 16 questions from a commonly used depression/anxiety scale [41] including measures from the past two weeks, such as restless sleep, talking less than usual, or feeling fearful. By using a common mental health scale used by many primary care physicians, participants may have been more likely to believe the false scenario. Attention check questions were inserted to help identify invalid responses.

Next, participants were asked questions about their personal health history and wellness behavior including height, weight, the frequency of alcohol consumed, illegal drug usage, prescription or nonprescription abuse, smoking, sexual activity, physical exercise and sleep. These data points were selected to provide a range of sensitivity to allow most participants to have at least some level of sensitivity to provide this information.

Lying, or the accuracy of their disclosure was measured in two ways. First, we used a novel mouse-tracking methodology validated in recent IS research [20] based on the distance, speed, and change of directions of the mouse used to complete the data entry form. Mouse tracking has been demonstrated to indicate a level of cognitive distress [20] interpreted as lying [42]. Individuals wishing to duplicate this study should ensure that participants have JavaScript enabled in order to capture as much usable mouse tracking data as possible. Second, subjects were then shown a statement informing them of the true purpose of the study to investigate the human tendency to misrepresent sensitive health information. Subjects were assured that they would receive full credit for participation regardless of whether they exaggerated or misrepresented prior responses. They were also assured that their responses were anonymous and impossible to trace back to them. The participants then indicated, on a scale spanning -5 to 5, how much they had over or understated their responses in the previous section, with -5 being a gross understatement, 0 being exactly accurate, and 5 being a gross overstatement. Their original answer to each question was also shown to help them recall and more accurately rate their original accuracy. When modeling, we use the absolute value of this response to control for lying in either direction.

Next, participants responded to a Likert-style scale that was loosely based on existing privacy calculus research [15, 19, 27, 39] but generated from the ground up for this study. The process by which these questions were designed and validated is outlined in section 3.1.1.

Social desirability was also measured using an existing scale [43] because it is a known covariate explaining whether respondents would act in a way that is deemed socially acceptable such as responding truthfully on a medical intake form. Finally, participants were asked for demographic information including income, gender, age, ethnicity, and education level, and debriefed on the true purpose of the study.

3.1.1. Content Adequacy Tests: Validly measuring the probability of a benefit/risk distinctly from the impact is difficult. These factors are highly correlated making it difficult to achieve discriminant validity. Our intent was to create a scale that would generalize to most health data contexts including mental health. Therefore, to establish content validity [44], we began by meeting with three clinicians (a doctor, physician's assistant, and nurse) with experience measuring depression and anxiety to ask them how they would characterize the impacts versus probability of the risks and benefits of disclosing health data.

To verify content validity, we used a technique known as a content adequacy test (CAT) [see complete details in 44]. This test involved four rounds of data collection and scale revisions. These data collections were based on 100-180 MTurk master workers each with no restrictions by any demographic. They were simply invited to participate in a mental health study, given construct definitions of RIM, RPR, BIM, and BPR, and asked to indicate how much each scale item seemed to measure each of the four constructs. The process is complete when a repeated measures analysis of variance test with contrasts indicates that each scale item measures its intended construct significantly more than every other construct. The final scale items are summarized in Table 2. The scale for general privacy concern was drawn from prior research [39].

Table 2. Final Scale Items for RIM, RPR, BIM, BPR

BIM1: Diagnoses made from the information I just shared could have a positive impact on my health.
BIM2: Diagnoses made from the information I just shared could help me receive better healthcare.
BIM3: Diagnoses made from the information I just shared could help me manage my health.
BPR1: It is likely that the information I just shared will help a healthcare provider make an accurate diagnosis.
BPR2: It is likely that the information I just shared will help a healthcare provider understand my symptoms.
BPR3: It is likely that the information I just shared will help a healthcare provider accurately classify my conditions.
RIM1: If an unethical person accesses the information I just disclosed, it could cause me embarrassment.
RIM2: If an unethical person accesses the information I just disclosed, it could cost me a lot of money.
RIM3: If an unethical person accesses the information I just disclosed, it could require a lot of effort to resolve the problems that could arise.
RIM4: If an unethical person accesses the information I just disclosed, the consequences to me could be high.
RPR1: It is likely that an unethical person will obtain access to the information I just disclosed.
RPR2: It is likely that the information I just disclosed will not be kept safe.
RPR3: It is likely that the information I just disclosed will be shared with third parties I did not anticipate.

3.1.2. Mouse Tracking Data: As mentioned above, participants' mouse movements were tracked by proprietary software designed to detect and flag fraudulent behavior. There were multiple variables recorded, such as the distance traveled by the mouse tracker, the speed at which it moved, and the total time a respondent took to answer the question. Z-scores were then computed for each of these variables in order to standardize the scale within each participant. Thus, responses that included mouse movements that were larger, took longer, or included more changes in direction, indicate greater cognitive distress and possible lying. Because of technical incompatibilities (e.g. people who took the survey over a mobile device or had JavaScript disabled), mouse tracking data was available for only 591 of the 1590 responses. Participants were not informed that their mouse movement would be tracked.

4. Results

4.1. Demographics and Descriptives

The sample included 39.91% women, 6.31% Hispanic, 6.84% Asian, 11.75% African or African American, and 1.66% other non-Caucasian participants, and were an average age of 36.0. Figure 3 provides a descriptive visualization of the average response and level of lying about each data type requested. As a reminder, we use an absolute value of lying to control for those lying in either direction.

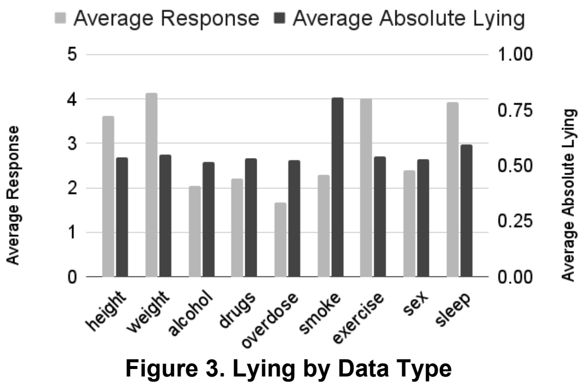

Figure 3. Lying by Data Type

4.2. Measurement Model

We estimated the reliability, convergent and discriminant validity, covariance, and common methods bias of the final sample. Table 3 summarizes the results and indicates that each scale had sufficient reliability (α > 0.7) [45]. Convergent validity is sufficient when composite reliability (CR) is over 0.7 and the average variance extracted (AVE) is over 0.5 for each scale [46]. All criteria were met.

Covariance was tested by calculating the variance inflation factor (VIF) for each exogenous construct. Every VIF score was below the recommended cutoff of 10.0 [47].

Table 3. Reliability, Validity, and Covariance

	BIM	BPR	PC	RIM	RPR	SD	**α**	**C.R.**	**VIF**
BIM	**0.69**	0.78	-0.04	-0.17	0.04	-0.04	0.79	0.87	2.556
BPR	0.60	**0.70**	-0.02	-0.14	0.04	-0.03	0.79	0.87	2.506
PC	0.00	0.00	**0.68**	0.57	0.72	-0.12	0.77	0.87	2.249
RIM	0.03	0.02	0.32	**0.81**	0.58	-0.13	0.77	0.89	1.712
RPR	0.00	0.00	0.52	0.33	**0.73**	-0.16	0.81	0.89	2.347
SD	0.00	0.00	0.01	0.02	0.03	**0.63**	0.75	0.84	1.031

Notes: AVEs along the diagonal, correlations above, squared correlations below

Discriminant validity is sufficient when the AVE for each reflective construct is greater than that construct's squared correlation with every other factor. This criterion was also met as each number in the diagonal (bolded and underlined) is greater than each of the values below it. In summary, we conclude that the data exhibited sufficient measurement model quality.

4.3. Hypothesis Testing

Hypothesis testing was performed using a partial least squares (PLS) based structural equation model (SEM) in SmartPLS 3.0 [48]. This was appropriate because our measure of accurate disclosure is formative representing the combined accuracy across all data types (height, weight, drugs, prescriptions, smoking, alcohol, sex, sleep, exercise) and subjects may be willing to lie more or less depending on the sensitivity of each data type to them personally. Our model was tested first using their stated responses indicating how much they lied after learning of the true nature of the study and second using the mouse tracking data. Both measures were reversed to frame the endogenous variable as the accuracy of disclosure rather than the extent of lying.

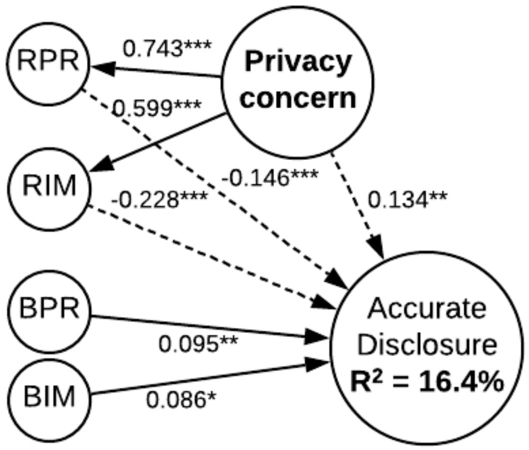

Figure 4. Coefficients & P-values for Stated Lying

The stated lying/accuracy model indicates that all relationships were significant. However, the effect of perceived risk probability (β = -0.146, $p < 0.001$) and impact (β = -0.228, $p < 0.001$) was much higher than those of perceived benefits probability (β = 0.095, $p < 0.01$) and impact (β = 0.086, $p < 0.05$). Despite these differences, all hypotheses are supported by this stated accuracy model. The overall model explained 16.4 percent of the variance in accurate disclosure.

To estimate our model using mouse tracking data, three measures were combined in a formative construct representing the accuracy of disclosure: mouse distance, slower speed, and total time to answer [20]. A separate model was estimated for each type of health data since different people are more or less likely to lie about different types of data. These models also include the demographic variables and two other control variables: social desirability (SD) and their initial response to each data type. However, for simplicity, we only show the relationships of the four sub-constructs of interest—RIM, RPR, BIM, and BPR—on mouse-tracked accuracy in Table 4.

Table 4. Coefficients for Mouse Tracking Models

	BIM	BPR	RIM	RPR	R^2
Alcohol	**-0.156†**	**0.134†**	**-0.078†**	**-0.114†**	15.2%
Drugs	0.053	0.033	-0.028	**-0.129***	12.3%
Exercise	-0.071	0.016	-0.065	**-0.166****	11.0%
Height	-0.170	**0.158***	**-0.106***	**-0.142***	6.2%
Overdose	**-0.182***	**0.127†**	-0.003	-0.092	6.8%
Sex	-0.130	**0.140†**	**-0.094***	-0.044	11.6%
Sleep	**-0.119†**	0.073	**-0.079†**	**-0.182****	8.8%
Smoke	-0.092	0.003	-0.046	**-0.083†**	11.1%
Weight	-0.201	**0.157***	**-0.087†**	-0.044	5.9%

Across the nine types of health data collected, BIM had a significant (or partially significant) effect on three of them (alcohol, prescription drug overdoses, and sleep), BPR on five, RIM on five, and RPR on six. These results agree with the stated measure of lying visualized in Figure 4 with one very interesting exception. For the significant relationships, those of BIM, RIM, and RPR significantly reduced accurate disclosure while those of BPR increased lying. In summary, H1a, H2b.

Although not included in Table 4, the demographics and controls also exhibited some significant relationships. Generally speaking across the data types, younger participants, non-Caucasian ethnicities, and men were more likely to disclose accurate health information based on the mouse tracking data. Income and education had no effect.

5. Discussion

Our study contributes two primary findings for both 1) the healthcare data domain, and 2) privacy calculus theory. First, both the stated response measures of accuracy and the mouse tracking measures suggest that perceived disclosure risks are more correlated with accurate health data disclosure than perceived disclosure benefits. Based on the TPB/TRA assumption of specificity, we interpret this to mean that the impact and probability of perceived privacy risks are easier to conceptualize for patients than are the benefits of disclosure. This is likely to be a continual issue for health data disclosure because patients typically do not know what diagnoses, treatment, or expected success will be at the time they are asked to disclose their health data. Therefore, it is difficult for them to accurately conceptualize the benefits of disclosing accurate information, causing it to have a weaker effect on actual disclosure behavior.

Second, we found that the perceived impacts versus probabilities of both risks and benefits do not always agree as hypothesized. For example, based on the mouse tracking data, the perceived BIM of disclosing alcohol, prescription overdoses, and sleep data significantly decreased the accuracy of disclosed information rather than increasing it as expected in H2b.

We believe this is a phenomenon that is unique to the healthcare context. Unlike the mobile app context referenced above, it is quite difficult to estimate and conceptualize exactly how a specific disclosure will result in a specific health benefit. For example, an alcoholic in relapse may not want to hear that he or she will need to give up their alcohol. They may believe that if the clinician knows the extent of their alcohol consumption, they are more likely to make an accurate diagnosis that is reflective of their alcohol problem. Therefore, the more they believe that accurate alcohol disclosure will lead to an accurate and impactful diagnosis, the more likely they are to lie about their true consumption.

Clearly, our suggested implication needs further research to be validated. However, if true, our study reveals a critical need to update privacy calculus theory. Its current formulation does not sufficiently specify risks and benefits by breaking them into their component parts of impact and probability, which is requisite for explaining much of seemingly 'irrational' consumer behavior. Also, if true, the implication for healthcare practice is that interventions may be needed to encourage more accurate health data disclosure which could end up saving lives.

5.2. Limitations and Future Research

Our study and results offer many opportunities for future research. Concerning our research design, although we improved the external validity of our results by deceiving participants, they never had to disclose their name or identity in the initial survey. Future designs should make the disclosure risks greater to see if our results are consistent.

Given that patients are more likely to lie if they believe that benefits will be impactful, it would be very useful to test various treatments in an experimental design to see if this effect can be reversed to encourage greater accuracy.

6. References

[1] Gvr, https://www.grandviewresearch.com/industry-analysis/mobile-application-market, 2020

[2] Govtech, https://www.govtech.com/blogs/lohrmann-on-cybersecurity/2020-data-breaches-point-to-cybersecurity-trends-for-2021.html, June 14th, 2021

[3] Annas, G.J., "Hipaa Regulations-a New Era of Medical-Record Privacy?", New England Journal of Medicine, 348(15), 2003, pp. 1486-1490.

[4] Medicareadvantage.Com, https://www.medicareadvantage.com/patient-doctor-lies-survey, August 2018, 2018

[5] Hassandoust, F., Akhlaghpour, S., and Johnston, A.C., "Individuals' Privacy Concerns and Adoption of Contact Tracing Mobile Applications in a Pandemic: A Situational Privacy Calculus Perspective", Journal of the American Medical Informatics Association, 28(3), 2021, pp. 463-471.

[6] Newman, L.H., "Medical Devices Are the Next Security Nightmare", WIRED, Mar, 2017,

[7] Palmieri, J.J., and Stern, T.A., "Lies in the Doctor-Patient Relationship", Primary care companion to the Journal of clinical psychiatry, 11(4), 2009, pp. 163.

[8] https://www.medicalnewstoday.com/articles/325811

[9] Dinev, T., Albano, V., Xu, H., D'atri, A., and Hart, P., "Individuals' Attitudes Towards Electronic Health Records: A Privacy Calculus Perspective": Advances in Healthcare Informatics and Analytics, Springer, 2016, pp. 19-50.

[10] Li, H., Wu, J., Gao, Y., and Shi, Y., "Examining Individuals' Adoption of Healthcare Wearable Devices: An Empirical Study from Privacy Calculus Perspective", International journal of medical informatics, 88(2016, pp. 8-17.

[11] Bol, N., Dienlin, T., Kruikemeier, S., Sax, M., Boerman, S.C., Strycharz, J., Helberger, N., and De Vreese, C.H., "Understanding the Effects of Personalization as a Privacy Calculus: Analyzing Self-Disclosure across Health, News, and Commerce Contexts", Journal of Computer-Mediated Communication, 23(6), 2018, pp. 370-388.

[12] Smith, H.J., Dinev, T., and Xu, H., "Information Privacy Research: An Interdisciplinary Review", MIS Quarterly, 35(4), 2011, pp. 989-1015.

[13] Bélanger, F., and Crossler, R.E., "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems", MIS Quarterly, 35(4), 2011, pp. 1017-1041.

[14] Keith, M.J., Babb, J., and Lowry, P.B., "A Longitudinal Study of Information Privacy on Mobile Devices": Book A Longitudinal Study of Information Privacy on Mobile Devices, 2013, pp. 6-9.

[15] Keith, M.J., Thompson, S.C., Hale, J., Lowry, P.B., and Greer, C., "Information Disclosure on Mobile Devices: Re-Examining Privacy Calculus with Actual User Behavior", International Journal of Human-Computer Studies, 71(12), 2013, pp. 1163–1173.

[16] Norberg, P.A., Horne, D.R., and Horne, D.A., "The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors", Journal of consumer affairs, 41(1), 2007, pp. 100-126.

[17] Kokolakis, S., "Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon", Computers & security, 64(2017, pp. 122-134.

[18] Ajzen, I., "The Theory of Planned Behavior", Organizational Behavior and Human Decision Processes, 50(2), 1991, pp. 179-211.

[19] Dinev, T., and Hart, P., "An Extended Privacy Calculus Model for E-Commerce Transactions", Information systems research, 17(1), 2006, pp. 61-80.

[20] Jenkins, J.L., Proudfoot, J., Valacich, J., Grimes, G.M., and Nunamaker Jr, J.F., "Sleight of Hand: Identifying Concealed Information by Monitoring Mouse-Cursor Movements", Journal of the Association for Information Systems, 20(1), 2019, pp. 3.

[21] Glasser, W., Choice Theory: A New Psychology of Personal Freedom, HarperPerennial, 1999.

[22] Laufer, R.S., and Wolfe, M., "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory", Journal of Social Issues, 33(3), 1977, pp. 22-42.

[23] Smith, H.J., Milberg, S.J., and Burke, S.J., "Information Privacy: Measuring Individual's Concerns About Organizational Practices", MIS Quarterly, 20(2), 1996, pp. 167-196.

[24] Acquisti, A., and Grossklags, J., "Privacy and Rationality in Individual Decision Making", IEEE Security & Privacy, 3(1), 2005, pp. 26-33.

[25] Acquisti, A., and Grossklags, J., "Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior", UC Berkeley 2nd Annual Workshop on "Economics and Information Security", 2003

[26] Acquisti, A., Taylor, C., and Wagman, L., "The Economics of Privacy", Journal of economic Literature, 54(2), 2016, pp. 442-492.

[27] Keith, M.J., Thompson, S.C., Hale, J., and Greer, C., "Examining the Rationality of Information Disclosure through Mobile Devices": Book Examining the Rationality of Information Disclosure through Mobile Devices, Orlando, FL, 2012

[28] Kahneman, D., and Tversky, A., "Prospect Theory: An Analysis of Decision under Risk": Handbook of the Fundamentals of Financial Decision Making: Part I, World Scientific, 2013, pp. 99-127.

[29] Tversky, A., and Kahneman, D., "Advances in Prospect Theory: Cumulative Representation of Uncertainty", Journal of Risk and uncertainty, 5(4), 1992, pp. 297-323.

[30] Laibson, D., "Golden Eggs and Hyperbolic Discounting", The Quarterly Journal of Economics, 112(2), 1997, pp. 443-478.

[31] Brandimarte, L., Acquisti, A., Loewenstein, G., and Babcock, L., "Privacy Concerns and Information Disclosure: An Illusion of Control Hypothesis", 2009,

[32] Petty, R.E., and Cacioppo, J.T., "The Elaboration Likelihood Model of Persuasion", Advances in experimental social psychology, 19(1986, pp. 123-205.

[33] Angst, C.M., and Agarwal, R., "Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion", MIS Quarterly, 2009, pp. 339-370.

[34] Lowry, P.B., Moody, G., Vance, A., Jensen, M., Jenkins, J., and Wells, T., "Using an Elaboration Likelihood Approach to Better Understand the Persuasiveness of Website Privacy Assurance Cues for Online Consumers", Journal of the American Society for Information Science and Technology, 63(4), 2012, pp. 755-776.

[35] Fishbein, M., "A Theory of Reasoned Action: Some Applications and Implications", 1979,

[36] Bandura, A., "Self-Efficacy: Toward a Unifying Theory of Behavioral Change", Psychological Review, 84(2), 1977, pp. 191-215.

[37] Madden, T.J., Ellen, P.S., and Ajzen, I., "A Comparison of the Theory of Planned Behavior and the Theory of

Reasoned Action", Personality and social psychology bulletin, 18(1), 1992, pp. 3-9.

[38] Keith, M.J., Thompson, S.C., Hale, J., Benjamin Lowry, P., and Greer, C., "Information Disclosure on Mobile Devices: Re-Examining Privacy Calculus with Actual User Behavior", International Journal of Human-Computer Studies, 71(12), 2013, pp. 1163–1173.

[39] Xu, H., Teo, H.H., Tan, B.C.Y., and Agarwal, R., "The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services", Journal of Management Information Systems, 26(3), 2010, pp. 135-174.

[40] Payne, T.H., Corley, S., Cullen, T.A., Gandhi, T.K., Harrington, L., Kuperman, G.J., Mattison, J.E., Mccallie, D.P., Mcdonald, C.J., and Tang, P.C., "Report of the Amia Ehr-2020 Task Force on the Status and Future Direction of Ehrs", Journal of the American Medical Informatics Association, 22(5), 2015, pp. 1102-1110.

[41] Zigmond, A.S., and Snaith, R.P., "The Hospital Anxiety and Depression Scale", Acta psychiatrica scandinavica, 67(6), 1983, pp. 361-370.

[42] Mazza, C., Monaro, M., Burla, F., Colasanti, M., Orrù, G., Ferracuti, S., and Roma, P., "Use of Mouse-Tracking Software to Detect Faking-Good Behavior on Personality Questionnaires: An Explorative Study", Scientific reports, 10(1), 2020, pp. 1-13.

[43] Fischer, D.G., and Fick, C., "Measuring Social Desirability: Short Forms of the Marlowe-Crowne Social Desirability Scale", Educational and psychological measurement, 53(2), 1993, pp. 417-424.

[44] Mackenzie, S.B., Podsakoff, P.M., and Podsakoff, N.P., "Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques", MIS Quarterly, 35(2), 2011, pp. 293-334.

[45] Santos, J.R.A., "Cronbach's Alpha: A Tool for Assessing the Reliability of Scales", Journal of extension, 37(2), 1999, pp. 1-5.

[46] Gefen, D., and Straub, D.W., "A Practical Guide to Factorial Validity Using Pls-Graph: Tutorial and Annotated Example", Communications of the AIS, 16(5), 2005, pp. 91-109.

[47] Salmerón, R., García, C., and García, J., "Variance Inflation Factor and Condition Number in Multiple Linear Regression", Journal of Statistical Computation and Simulation, 88(12), 2018, pp. 2365-2384.

[48] Ringle, C.M., Wende, S., and Becer, J.-M., "Smartpls 3", Boenningstedt: SmartPLS GmbH, http://www.smartpls.com, 2015,