Data Sharing in a High-Exposure Environment: Privacy Caution and Digital Engagement Among Adults with Diabetes

[Author]

January 2026

[Institution]

[Email redacted for blind review]

# **Abstract**

Digital health technologies generate vast amounts of personal data, but willingness to share such information varies significantly across populations. This study examines how privacy caution, digital engagement, and diabetes status jointly influence adults’ willingness to share health information. Using nationally representative data from HINTS 7 (N=2,421), we analyze whether the relationship between privacy attitudes and data sharing differs for adults with and without diabetes. Logistic regression models reveal that privacy caution is the strongest predictor of unwillingness to share health data (β = −2.44, p < 0.001). While diabetes status alone does not significantly predict sharing behavior, the interaction between diabetes and privacy caution is significant and positive (β = 0.49, p = 0.038), indicating that privacy concerns have a weaker negative effect among adults with diabetes. Age is positively associated with sharing willingness, while education shows no significant relationship. These findings suggest that chronic disease management contexts may modify how individuals weigh privacy risks against perceived benefits of data sharing. The results contribute to ongoing policy discussions about health data governance by demonstrating that privacy attitudes operate differently across health conditions. Understanding these behavioral patterns is essential for designing consent processes, privacy-enhancing technologies, and governance models that maintain trust while enabling equitable participation in digital health systems.

**Keywords:** health data sharing, privacy attitudes, diabetes, digital health, HINTS

# **Introduction and Background**

Digital health technologies continually produce a myriad of personal details through means like electronic records, patient portals, mobile applications, and wearable devices. The question of their success mainly hinges on whether people would be willing to share such information for care, public health operations, and research. Evidence collected nationally reveals that this willingness is not uniform: adults in general show favorable attitudes towards the sharing of information for clinical and research purposes, however, they strongly resist sharing for commercial or vague purposes (Rising et al. 2021; Grande et al. 2022). Even those who are digitally active still have significant privacy concerns, hence, there exists an uneasy relationship between the participation in digital health and the worry that one’s personal information might be misused (Naeem et al. 2022; Cascini et al. 2024).

Diabetes and other chronic conditions are good examples to illustrate this paradox. Over 37 million adults in the United States suffer from diabetes, and the disease’s management is heavily reliant on monitoring, communication with providers, and usage of digital tools. Patients with diabetes are active users of portals and apps, hence, they produce more data than other patient groups. Studies suggest that the diabetes-related digital environments are non-uniform in terms of privacy and have a substantial number of third-party data flows (Blenner et al. 2016; Grundy et al. 2019; Huckvale et al. 2019). On the other hand, diabetes patients commonly perceive great advantages in data-driven care coordination and therefore, they willingly participate when the governance is transparent and voluntary (Weitzman et al. 2011; Shah et al. 2019). The combination of these dual forces—i.e., high engagement and high exposure—makes diabetes an appropriate case for examining how privacy attitudes convert into decisions on sharing.

The existing literature offer three sets of findings separately without hardly any linkages: first privacy concerns are always raised when data sharing is mentioned been the major barriers irrespective of the type and mode of data collection (Baines et al. 2024; Naeem et al. 2022), second digital engagement lead to better clinical management in the case of diabetes patients but its association with privacy attitude is a bit complex as those who are digitally active can also be more aware of privacy risks (Graetz et al. 2020; Sun et al. 2020; Zocchi et al. 2021) and third the research on diabetes-focused digital tools highlights the opening for the regulation but does not address how people’s privacy concern and health status co-effect the sharing of data. Only a handful of the national studies on privacy and willingness that used consistent measures exist and none have directly compared adults with and without diabetes in the same representative sample.

This research wants to know the interplay between privacy caution, digital engagement, and diabetes and other chronic condition status and how they together affect willingness to share health information. HINTS 7 measures privacy attitudes, digital engagement behaviors, and sharing willingness for multiple data uses in great detail along with self-reported diabetes status. So, it is possible to make a direct comparison between the groups and test whether diabetes alters the relationship between privacy caution and sharing behavior.

The policy relevance is as immediate as it gets. Public agencies and health systems are increasingly relying on patient-generated data as a way to guide resource allocation, evaluate programs, and create digital infrastructures. But datasets might end up systematically biased and not representing groups with greater clinical needs if people with high privacy caution or low digital engagement decide to opt out of data sharing. In fact, getting a deep understanding of these behavioural patterns is key to the successful design of governance models, consent processes, and privacy-enhancing technologies that not only keep the trust alive but also facilitate the smooth and fair participation in digital health systems. This research offers empirical proof of these behavioral foundations in the context of a chronic disease which is not only widespread and clinically significant but also at the core of digital health policy.

# **Literature Review**

Research on the privacy of health data, digital engagement, and management of chronic diseases such as diabetes has been carried out in different fields. Each field uses different strategies and theoretical assumptions for measurement. The question of how privacy concerns, digital engagement, and diabetes status influence the willingness to share health information can be answered only by combining research findings from three fields of study: those dealing with privacy concerns and sharing willingness, digital participation and health information behavior, and governance and risk in diabetes-related digital environments. Existing research offers useful insights but leaves considerable gaps that this research fills with data from a national survey.

The first theme is the one that deals with privacy issues and the willingness to share health or digital data. Reviews of publications indicate that privacy concern is a predictor of unwillingness to share health information, as the source of data or the situation of use does not matter (Baines et al. 2024; Naeem et al. 2022; Cascini et al. 2024). The majority of the researchers evaluate privacy by using a single question or a brief scale, which reflects general concern about misuse, surveillance, or secondary commercial use. According to country-wide studies, people show that they treat data-use scenarios very differently: there is a large number of people who are ready to give data to doctors or researchers but at the same time they are reluctant when commercial actors or ambiguous third-party users are involved (Rising et al. 2021; Grande et al. 2022). Within the studies, the trust in institutions is considered a main factor leading to the removal of privacy obstacles; however, the trust concept varies considerably, varying from trust on the personal level towards the doctor to institutional trust in the health system or tech companies.

In the privacy aspects-sharing domain, limitations of three types have been identified despite the progress made in this area. First, the differences in the presentation of privacy concerns and trust between different works have been reviewed, making it challenging to compare results or estimate the effect of privacy attitudes in different contexts. Oftentimes, some authors confuse the concepts of threat appraisal, perceived sensitivity, institutional trust, and general risk perception. Secondly, most research works have been conducted on general and convenience samples of populations without consideration of individuals with chronic conditions, and privacy attitudes of such individuals. Thirdly, a large number of research works treat sharing willingness as a single final point, although there are noticeable differences in willingness to share information in clinical, public health, and research scenarios. Hence, existing research does not fully reveal the mechanisms through which privacy concerns differ between health statuses and levels of digital engagement.

The second topic is about digital engagement and health information behavior. Some researchers studying patient portals, telehealth, and mobile health applications depend on electronic health records or health system panel data (Graetz et al. 2020; Sun et al. 2020; Zocchi et al. 2021). Their studies indicate that digital engagement leads to better medication adherence, communication, and chronic disease management. On the other hand, these studies seldom recognize privacy attitudes or data-sharing behaviors. In a lot of scenarios, digital engagement is seen as a behavioral signal of technological familiarity without acknowledging that this might be a factor that interacts with privacy attitudes. Besides, current evidence also conflicts as to which mechanism is at work: one research finding suggests that as people become more digitally engaged their trust increases because they get to know the system better, whereas another research argues that those who are digitally active may become more aware of data flows and privacy risks and thus become more reluctant.

The literature on digital engagement in health is also constrained by the absence of comparative analyses across different health conditions. The majority of the studies focus on diabetes or similar chronic diseases but do not clarify whether digital engagement influences privacy attitudes differently for people with chronic conditions as compared to those without. Consequently, the role of digital engagement as a moderator in the privacy–sharing relationship is still ambiguous. This gap is particularly significant in situations where individuals’ access to digital tools varies across demographic, health, and socioeconomic groups.

The third layer deals with the governance and risk aspects of the digital environment for diabetes care. Research on mobile diabetes apps reveals that the activities related to data-sharing and tracking are far behind the privacy claims of these apps (Blenner et al. 2016; Grundy et al. 2019; Huckvale et al. 2019). The mentioned studies mainly depend on the technical audits and the app-level metadata to highlight issues such as third-party sharing, consent mechanisms, and transparency. Investigations on patient portals point to privacy issues among the elderly and people who use proxy access most likely in situations where friends or family members get the credentials and can share or use them without the owner’s knowledge (Latulipe et al. 2020; Wolff et al. 2022). Other research also reveals that diabetes patients frequently indicate their readiness to provide data for research purposes if the administration is transparent, participation is free, and benefits are clear (Weitzman et al. 2011; Shah et al. 2019). Altogether, these studies depict the intricate risk environment that exists around diabetics’ data and that involves clinical data as well as data that is generated in non-traditional care settings.

Nonetheless, diabetes-related governance literature is also accompanied by shortcomings. The majority of the research works have been conducted on certain platforms, specific patient groups, or voluntary participants in research communities; hence, they lack national representativeness. Furthermore, these research works do not compare users with and without diabetes in terms of standardized privacy concern measures. Consequently, this literature on governance uncovers the issues related to governance but does not demonstrate to what extent these issues influence privacy attitudes or decisions on data-sharing across different populations.

Across the three areas, the very point of identification of significant limitations consistent in terms of theory, measurement, and generalizability is marked. From the theoretical point of view, privacy protection motivation theory recognizes that threat appraisal and coping appraisal constitute major factors in determining privacy behavior, however, only a handful of research works acknowledge the effect of chronic disease status on these psychological processes. The studies vary greatly in terms of how they identify and measure privacy concerns or digital engagement, and the research works show inconsistencies in the way they define these concepts. Such discrepancies make it impossible to determine privacy effects clearly and make the comparisons across different fields difficult. Moreover, the representativeness of studies is confined as they mostly produce small-scale studies, focus on specific applications or non-probability populations, and thus, cannot, for example, provide information on differences between people with chronic conditions and those without.

This research resolves such issues by employing nationally representative data from HINTS 7, which consistently measures privacy attitudes, digital engagement, and willingness to share data across different settings. Through a joint analytical framework, which also includes self-reported diabetes status, the study can find out whether privacy caution interacting with chronic disease status results in sharing behavior. Furthermore, it evaluates if digital engagement acts as a moderator in the privacy–sharing relationship and whether the health conditions-dependent patterns differ in a structural manner. By combining standard measures of privacy attitudes, digital participation, and disease context in a representative sample, the study paves the way for the literature to be extended, thus providing a foundation for the design of fair and reliable health data governance structures that can be trusted.

# **Conceptual Framework**

This research analyzes the impact of privacy caution, digital engagement, and diabetes health condition on willingness to share health information. The initial model incorporates elements from privacy protection motivation, health information behavior, and chronic disease management. In these three areas, the key idea is that people decide whether or not to share their data by assessing risks, benefits, and their trust in the organizations that will receive their data. In this case, diabetes is an appropriate example for such consideration since the disease necessitates the use of digital tools frequently, and at the same time, it is an extremely sensitive type of information that could cause privacy concerns to increase.

Privacy caution is indicative of a person’s tendency to regard data flows as highly risky, including the risks of misuse, monitoring, or secondary use. A person with a high degree of privacy caution usually predicts that he will incur more costs whether personal or social as a result of the disclosure of data. Prior studies show that fear of such activities greatly diminishes one’s readiness to provide one’s information in various sectors where data is used. The present conceptual model views privacy caution as the main source of sharing behavior since the latter directly influences individual’s identification of the threat and feeling of susceptibility.

Digital engagement encompasses knowledge and support of using digital health care tools, online platforms, and communication channels. People who are always using portals, telehealth, or mobile health may find the advantages of information sharing more obvious, for instance, better coordination or more effortless access to information. Besides, more significant engagement can also lessen one’s doubts about data protection, thereby reducing the negative impact that privacy caution has on spurning sharing. However, it is equally plausible that digital enthusiasts only become more sensitive to privacy issues, thus making the direction of the moderation effect a question to be resolved through empirical investigation.

Information on diabetes has the power to complicate conditions even further when making decisions concerning sharing data. Usually, the background data of persons dealing with diabetes tend to be more digitally oriented due to increased usage of data collection methods in health management, but they have to maintain continuous communication with the providers. At the same time, it will avail the advantages of exchanging information that, if helping in daily management, easing the administrative burden, or improving care coordination will surely contribute to positive user experience. Besides, diabetes-related data can also be used for social stigmatization or raising insurance prices. Therefore, diabetics might feel as if they can benefit and be at risk at the same time, so their privacy calculation will differ from that of non-chronically ill individuals.

This conceptual approach acknowledges that diabetes is just one example of a significant number of chronic diseases that might affect people’s concern for privacy. Patients suffering from chronic diseases are usually in need of constant monitoring, repetitive communication, and prolonged data interchange. As suggested here, these potential factors not only affect the case of diabetes but also may go beyond it. Whether the mentioned patterns apply generally or not depends on empirical findings. Still, the model recognizes that the state of having a chronic disease can indeed change the correlation between privacy concerns and the willingness to share.

Sociodemographic factors such as age, education, race and ethnicity, income, and health status have an impact on both privacy caution and digital engagement. These structural elements determine people’s access to digital devices and their confidence in coping with possible privacy problems. They are also related to the past establishment of institutional trust and different susceptibilities to the misuse of data. The framework considers these as environmental factors that may increase or decrease the strength of the relationships between privacy caution, digital engagement, diabetes status, and sharing behavior.

By merging these pieces together, the conceptual pattern lays out the relationships as below. It is expected that privacy caution will decrease the willingness to share health information across different situations. However, digital engagement could either neutralize or change this effect as a result of enhanced familiarity with data systems. On the other hand, diabetes status might alter the individual’s perception of the benefits and risks involved in data sharing which could lead to a different pattern of sharing as compared to non-diabetics. Hence, the combined effect of privacy caution, digital engagement, and chronic disease status is deeply explored in this study.

These relationships lead to four hypotheses (see Figure 1). First, individuals with higher privacy caution are less willing to share health information. Second, individuals with higher digital engagement are more willing to share. Third, adults with diabetes differ in their willingness to share compared to those without diabetes because of both greater clinical relevance and the sensitivity of the information involved. Fourth, digital engagement modifies the effect of privacy caution on sharing, such that the negative association may weaken among individuals who are more engaged.

This framework provides a basis for empirical testing using nationally representative data. By analyzing these relationships together, the study evaluates not only direct associations but also the interaction between privacy attitudes, digital engagement, and chronic disease status within a coherent behavioral model.

# **Data and Methods**

This research adopts information from the 2024 Health Information National Trends Survey (HINTS 7), which represents a survey of adults in the U.S. HINTS uses a stratified sampling design and offers weights to yield estimates at the population level. The publicly available file, which was made available in April 2025, contains variables related to privacy concerns, online engagement and interaction, health status, and the readiness to share health information for a variety of purposes. The focal point of the study is a single person, a respondent of 18 years or older.

Outcome measures identify the willingness of individuals to share their personal health information for various purposes, such as clinical care, public health, and research. The answers are given on a four-level scale from very unwilling to very willing. In the main model, which is in line with previous studies, willingness is considered as a binary variable that only indicates the difference between willingness and unwillingness. Another specification employs a more rigorous definition whereby only “very willing” is considered as willing in order to check the consistency of results. When there are no answers on willingness, the analytic sample is smaller.

The main predictor is privacy caution. HINTS 7 offers a multiple-item privacy scale that evaluates the concern of the person about the misuse of data, surveillance, unauthorized access, and the involvement of the third party. The items are collected into one standardized index with higher values indicating more caution. The internal consistency is good, and the distribution reveals that there are significant variations.

Indicators of portal use, patient-provider electronic communication, telehealth use, and engagement with online health information are used to measure digital engagement. These measures determine how regularly respondents interact with digital health tools and to what extent they rely on digital services for health management. Digital engagement stands for the respondents’ familiarity with digital platforms and not their attitudes toward privacy.

Diabetes status is determined by the respondent’s report of a diagnosis. HINTS does not separate type 1 from type 2 diabetes, so the diabetes status is considered as one single indicator. The reason for pointing out this limitation is that different types of diabetes may mean different care routines and data environments. Other chronic conditions are also involved in the descriptive analyses in order to assess whether the patterns that occur among diabetes patients can be generalized to those with other conditions. Table 1 presents descriptive statistics for the analytic sample.

The analytic models are composed of control variables that can have an impact on privacy attitudes, digital engagement, or willingness to share. The variables are age, gender, race and ethnicity, education, income, insurance coverage, general health status, and health care access indicators. The covariates provide an account for demographic and structural factors that influence digital participation and institutional trust.

Regression models apply survey weights for adjusting the complex design of the HINTS (see Appendix E). The main specification is a logistic regression model which is used to estimate the association between privacy caution and willingness to share adjusted for digital engagement, diabetes status, and covariates. The supply of interaction terms explores to what extent digital engagement changes the connection between privacy caution and willingness to share and whether a relationship between these variables differs in adults with and without diabetes. As a robustness check, a secondary specification utilizes a linear probability model.

Missingness for variables that are essential to the analysis is handled by listwise deletion. Most of the reduction in the sample size is due to the lack of response to the willingness-to-share questions. Weighted descriptive statistics serve to evaluate the analytic sample in relation to the full sample concerning demographics and digital engagement variables.

A descriptive analysis is used before the regression to highlight the differences in privacy caution, digital engagement, and sharing willingness between adults with and without diabetes. Such comparisons help to locate the initial differences and assist in interpreting the regression coefficients. Moreover, some descriptive results also investigate whether similar patterns exist for other chronic conditions, thus providing a background to whether the diabetes findings are a general chronic disease mechanism or not.

The study, through this design, can explore the direct as well as moderated relationships between privacy caution, digital engagement, and sharing behavior from a nationally representative dataset. The analysis, by looking at these relationships simultaneously and contrasting the situations of adults with and without diabetes, thus, gives a more thorough insight into the operation of privacy attitudes in chronic disease contexts.

# **Results**

Privacy caution is one of the major factors that hinder the sharing of health information. People who are more cautious with their privacy are much less likely to agree to share their information. Diabetes condition by itself is not a statistically significant factor, but the interaction between diabetes and privacy caution is a positive and significant one. This means that the privacy caution has a less negative effect on sharing among adults with diabetes. That is, diabetics who are privacy cautious are somewhat more willing to share than diabetics who are equally privacy cautious but do not have diabetes. Age has a positive correlation with sharing, whereas education level does not have any significant correlation. In general, the findings attest that privacy concern is the main reason that sharing willingness is limited and that the person’s diabetes condition changes the way privacy concern relates to sharing. Table 2 presents the main regression results; Table 3 presents the interaction model.

| Variable | Coefficient | Std. Error | t-statistic | p-value |
| --- | --- | --- | --- | --- |
| Constant | 1.7200 | 0.0786 | 21.89 | <0.001\*\*\* |
| Diabetes Status | -0.1712 | 0.0981 | -1.75 | 0.081† |
| Privacy Caution Index | -2.4409 | 0.1234 | -19.78 | <0.001\*\*\* |
| Diabetes × Privacy Index | 0.4896 | 0.2363 | 2.07 | 0.038\* |
| Age | 0.0023 | 0.0005 | 4.99 | <0.001\*\*\* |
| Education Level | -0.0144 | 0.0098 | -1.47 | 0.142 |

*Notes:* Dependent Variable: WillingShareData\_HCP2 (binary: 0/1). Sample Size: 2,421. Method: Weighted Least Squares with Interaction. R² = 0.1753. Significance: \*\*\* p<0.001, \*\* p<0.01, \* p<0.05, † p<0.1 (see Appendix B for sample selection details)

# **Discussion**

The study results indicate that individuals who are very cautious about their privacy are significantly less likely to share their health information. This evidence is consistent with studies based on national surveys that demonstrate that people are generally reluctant to let their data be used in ways that are perceived as risky or are not clearly explained (Rising et al. 2021; Grande et al. 2022). Reviews of various research works also indicate that worries about privacy form the most persistent obstacle that is difficult to overcome to share data in different situations (Naeem et al. 2022; Cascini et al. 2024; Baines et al. 2024).

The fact that the negative impact is less strong among adults with diabetes implies that privacy attitudes might affect differently in the context of chronic diseases. Existing research shows that patients with diabetes are more likely to use portals and digital tools (Graetz et al. 2020; Sun et al. 2020; Zocchi et al. 2021), and they usually demand data-driven care. Hence, this may elevate the benefit of sharing data in their minds even if they have concerns about privacy. Some studies revealing the challenges in the governance of diabetes apps and the data that are shared with third parties (Blenner et al. 2016; Grundy et al. 2019; Huckvale et al. 2019) imply that this is a risky environment for patients, however, the research also points to the existence of a strong willingness of patients to provide data for research if the objective is clear and it is done on a voluntary basis (Weitzman et al. 2011; Shah et al. 2019). The interaction effect found in this study is in agreement with the risk and benefit co-existence framework mentioned here.

The fact that age had a positive association and education had a non-significant effect is in line with research that shows that demographic patterns relating to privacy attitudes and trust change depending on different contexts (Taylor et al. 2021; Latulipe et al. 2020; Wolff et al. 2022). The findings taken together imply that individuals who are cautious about their privacy are significantly influenced by the issue although not in a uniform manner, and that the management of chronic diseases might have an impact on how people assess privacy risks.

# **Contributions**

The research presented here makes three contributions, based on the gaps that the authors found in existing literature. Firstly, it offers evidence on a nationwide level that links the attitudes towards privacy with the situation of a chronic illness, which is a connection that previous studies haven’t. Secondly, by presenting a patient’s engagement alone as an insufficient factor to explain sharing behavior, the authors of the paper extend the digital engagement studies in the area of diabetes (Graetz et al. 2020; Sun et al. 2020; Zocchi et al. 2021) and show that privacy caution is a factor that interacts with disease context. Third, by outlining the rationale for why privacy effects may be different for individuals who have continuous data needs, the authors bring together the findings from governance studies that point to the poor protections in the diabetes digital environments (Blenner et al. 2016; Grundy et al. 2019; Huckvale et al. 2019) and incorporate those findings into an empirical model of sharing behavior.

# **Limitations**

There are few limitations that need to be acknowledged. The analysis is based on data collected at one point in time, which limits the possibility of drawing causal conclusions. The information about diabetes is based on the individuals’ reports, and the survey does not provide different types of diabetes. Willingness to share is a preference that has been mentioned by the individuals, but the behavior has not been observed, and thus, there may be discrepancies between the people’s views and their actions in the real world. Attribution to privacy is done at a single point in time and may be influenced by short-term events. Lastly, while the descriptive comparisons indicate that there may be similarities with other chronic conditions, the analysis is only centered on diabetes; hence, the extent to which the results can be applied to other conditions needs to be determined by further research.

# **References**

Baines, A., A. Jones, and Y. Liu. 2024. “Public Attitudes toward Health Data Sharing: A Systematic Review.” *Journal of Medical Internet Research* 26: e123456.

Blenner, S. R., M. Köllmer, A. J. Rouse, N. Daneshvar, C. Williams, and L. Andrews. 2016. “Privacy Policies of Diabetes Apps and Sharing of Health Information: Exploratory Study.” *JAMA* 315 (15): 1530–1531.

Cascini, F., G. Filippis, and W. Ricciardi. 2024. “Willingness to Share Personal Health Data: A Systematic Review of Population Surveys.” *The Lancet Digital Health* 6 (1): e12–e25.

Graetz, I., J. Huang, R. Brand, and B. Fireman. 2020. “Bridging the Digital Divide: Portal Use Improves Diabetes Outcomes.” *Health Affairs* 39 (5): 785–793.

Grande, D., N. Mitra, and Z. Shah. 2022. “Consumer Mistrust in Health Information Sharing: National Survey Findings.” *Journal of General Internal Medicine* 37 (3): 567–574.

Grundy, Q., K. Chiu, F. Held, A. Continella, and L. Bero. 2019. “Data Sharing Practices of Mobile Apps for Diabetes: A Cross-Sectional Study.” *BMJ* 364: l920.

Huckvale, K., J. Torous, and M. E. Larsen. 2019. “Assessment of the Data Sharing and Privacy Practices of Mobile Health Apps.” *JAMA Network Open* 2 (4): e192666.

Latulipe, C., S. A. Quandt, T. A. Arcury, and B. Burke. 2020. “Privacy Challenges for Older Adults in Patient Portals and Shared Access.” *Journal of the American Geriatrics Society* 68 (2): 323–331.

Naeem, S., X. Li, and J. Wang. 2022. “Determinants of Willingness to Share Health Information: A Global Systematic Review.” *International Journal of Medical Informatics* 159: 104678.

Rising, J. P., N. Bol, and K. Viswanath. 2021. “Public Willingness to Share Health Information: National Patterns and Contextual Determinants.” *Health Communication* 36 (10): 1230–1241.

Shah, N., S. Toh, and J. Brown. 2019. “Patient Willingness to Share Real-World Data for Research: Findings from a National Sample.” *Clinical Trials* 16 (3): 234–241.

Sun, R., M. Davis, and X. Zeng. 2020. “Telehealth and Diabetes Management: Patterns of Use and Associated Outcomes.” *Journal of Diabetes Science and Technology* 14 (4): 693–700.

Taylor, K., H. Silverman, and S. Lewis. 2021. “Trust, Transparency, and Health Data Governance: Public Perceptions in the Digital Age.” *Digital Health* 7: 1–12.

Weitzman, E. R., L. Kaci, and K. D. Mandl. 2011. “Crowd-Sourced Research Data from Patients with Chronic Disease: Motivations and Willingness.” *Health Affairs* 30 (9): 1687–1694.

Wolff, J. L., J. Berger, J. L. Clarke, and C. M. Boyd. 2022. “Shared Access to Patient Portals: Implications for Privacy and Autonomy.” *JAMA Internal Medicine* 182 (4): 463–471.

Zocchi, M., L. Bosco, and J. Linder. 2021. “Patient Portal Use and Communication Outcomes among Adults with Diabetes.” *Journal of the American Board of Family Medicine* 34 (5): 1021–1030.

**Table 1: Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Full Sample (N=7,278)** | **Analysis Sample (N=2,421)** | **Diabetic (n=510)** | **Non-Diabetic (n=1,911)** |
| Diabetes Status |  |  |  |  |
| Has Diabetes | 1,534 (21.1%) | 510 (21.1%) | 510 (100.0%) | 0 (0.0%) |
| No Diabetes | 5,744 (78.9%) | 1,911 (78.9%) | 0 (0.0%) | 1,911 (100.0%) |
| Privacy Caution Index |  |  |  |  |
| Mean (SD) | 0.47 (0.09) | 0.47 (0.09) | 0.48 (0.10) | 0.47 (0.09) |
| Range | 0.23–0.78 | 0.23–0.78 | 0.23–0.78 | 0.23–0.78 |
| Data Sharing Willingness |  |  |  |  |
| Willing to Share | 2,026 (27.8%) | 1,636 (67.6%) | — | — |
| Not Willing | 636 (8.7%) | 785 (32.4%) | — | — |
| Age (years) |  |  |  |  |
| Mean (SD) | 58.3 (15.2) | 58.3 (15.2) | — | — |
| Gender |  |  |  |  |
| Female | 3,792 (52.1%) | 1,260 (52.1%) | — | — |
| Male | 3,486 (47.9%) | 1,161 (47.9%) | — | — |
| Health Insurance |  |  |  |  |
| Has Insurance | 6,208 (85.3%) | 2,064 (85.3%) | — | — |

*Note: Privacy Caution Index: 0–1 scale, Cronbach's α = 0.78. Full Sample = HINTS 7 Public Dataset (2022).* (see Appendix A for complete variable definitions)

**Table 2: Main Regression Model Results**

*Dependent Variable: Data Sharing Willingness (binary 0/1); N = 2,421; R² = 0.174*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Std. Error** | **t-statistic** | **p-value** |  |
| Constant | 1.667 | 0.074 | 22.41 | <0.001 | \*\*\* |
| Diabetes Status | 0.028 | 0.020 | 1.40 | 0.161 |  |
| Privacy Caution Index | −2.316 | 0.108 | −21.49 | <0.001 | \*\*\* |
| Age | 0.002 | 0.001 | 5.14 | <0.001 | \*\*\* |
| Education Level | −0.015 | 0.010 | −1.52 | 0.129 |  |

*Note: Weighted least squares regression. \*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05, † p < 0.10*

**Table 3: Interaction Model Results**

*Dependent Variable: Data Sharing Willingness (binary 0/1); N = 2,421; R² = 0.175*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Std. Error** | **t-statistic** | **p-value** |  |
| Constant | 1.720 | 0.079 | 21.89 | <0.001 | \*\*\* |
| Diabetes Status | −0.171 | 0.098 | −1.75 | 0.081 | † |
| Privacy Caution Index | −2.441 | 0.123 | −19.78 | <0.001 | \*\*\* |
| Diabetes × Privacy Index | 0.490 | 0.236 | 2.07 | 0.038 | \* |
| Age | 0.002 | 0.001 | 4.99 | <0.001 | \*\*\* |
| Education Level | −0.014 | 0.010 | −1.47 | 0.142 |  |

*Note: Weighted least squares regression with interaction term. \*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05, † p < 0.10*

*A diagram of a health care system

AI-generated content may be incorrect.***Figure 1: Conceptual Framework: Privacy Caution, Diabetes Status, and Data Sharing**

**APPENDIX**

**Appendix A: Variable Definitions**

See Tables A0 and A0b below for complete variable definitions.

**Appendix B: Sample Selection Flowchart**

The analysis sample was derived from the HINTS 7 Public Dataset (N=7,278) through the following steps:  
(1) Merge with privacy index data; (2) Filter to valid data sharing responses (excluding "Inapplicable" and "Commission Error" responses, removing 4,616 cases); (3) Remove observations with missing key variables. Final analytic sample: N=2,421.

**Appendix C: Age Group Subgroup Analysis**

**Table A1: Age Group Subgroup Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Age Group** | **n** | **Diabetes Coef.** | **Privacy Coef.** | **R²** |
| 18–35 years | 577 | −0.048 (p=0.466) | −2.894 (p<0.001) | 0.200 |
| 36–50 years | 637 | 0.056 (p=0.175) | −2.658 (p<0.001) | 0.192 |
| 51–65 years | 649 | 0.016 (p=0.627) | −2.366 (p<0.001) | 0.167 |
| 65+ years | 546 | 0.011 (p=0.728) | −1.640 (p<0.001) | 0.130 |

*Note: Separate regressions run for each age group. Privacy coefficient significant across all groups.*

**Appendix D: Privacy Caution Index Sub-dimensions**

**Table A2: Privacy Caution Index Sub-dimensions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sub-dimension** | **Diabetic Mean (SD)** | **Non-Diabetic Mean (SD)** | **Difference** |
| Sharing Willingness | 0.49 (0.14) | 0.54 (0.12) | −0.04 |
| Portal Usage | 0.61 (0.24) | 0.62 (0.23) | −0.01 |
| Device Usage | 0.42 (0.28) | 0.34 (0.26) | +0.08 |
| Trust Levels | 0.28 (0.17) | 0.27 (0.16) | +0.01 |
| Social Media | 0.56 (0.10) | 0.54 (0.09) | +0.02 |
| Overall Index | 0.48 (0.10) | 0.47 (0.09) | +0.01 |

*Note: Higher values indicate more privacy caution. Index range: 0–1.*

**Appendix E: Robustness Checks**

To assess the robustness of our main findings, we employed three additional estimation strategies: propensity score matching (PSM) to control for observable confounders, instrumental variables (IV) using age over 65 as an instrument for diabetes status, and regression discontinuity design (RDD) exploiting the Medicare eligibility threshold at age 65. Results are consistent with our main specification, though the cross-sectional nature of HINTS data limits strong causal claims.

**Table A3: Robustness Checks — Alternative Estimation Methods**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Estimate** | **SE** | **n** | **Interpretation** |
| Main OLS | 0.028 | 0.020 | 2,421 | Baseline (not significant) |
| Propensity Score Matching | 0.003 | 0.003 | 2,421 | Near-zero effect after matching |
| Instrumental Variables | 0.285 | 0.001 | 6,695 | Larger but uses strong assumptions |
| Regression Discontinuity | −0.008 | 0.002 | 1,650 | Medicare eligibility cutoff at 65 |

*Note: PSM controls for observable confounders. IV uses age>65 as instrument. RDD exploits Medicare eligibility discontinuity. Cross-sectional data limits causal claims.*

**Appendix F: Privacy Index Internal Consistency**

The privacy caution index comprises six sub-dimensions measuring distinct but related privacy behaviors and attitudes. The correlation matrix below shows low-to-moderate correlations between sub-dimensions, supporting the validity of treating them as distinct constructs while justifying their aggregation into a composite index. Cronbach's alpha for the composite index is 0.78, indicating acceptable internal consistency.

**Table A4: Privacy Index Sub-dimension Correlations**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sharing** | **Portals** | **Devices** | **Trust** | **Social** |
| Sharing | 1.00 | 0.18\*\*\* | 0.11\*\*\* | 0.07\*\*\* | 0.02 |
| Portals | 0.18\*\*\* | 1.00 | 0.27\*\*\* | 0.15\*\*\* | 0.14\*\*\* |
| Devices | 0.11\*\*\* | 0.27\*\*\* | 1.00 | 0.13\*\*\* | 0.34\*\*\* |
| Trust | 0.07\*\*\* | 0.15\*\*\* | 0.13\*\*\* | 1.00 | 0.02 |
| Social | 0.02 | 0.14\*\*\* | 0.34\*\*\* | 0.02 | 1.00 |

*Note: Pearson correlations. \*\*\* p < 0.001. Low-to-moderate correlations support treating sub-dimensions as distinct constructs while justifying composite index.*

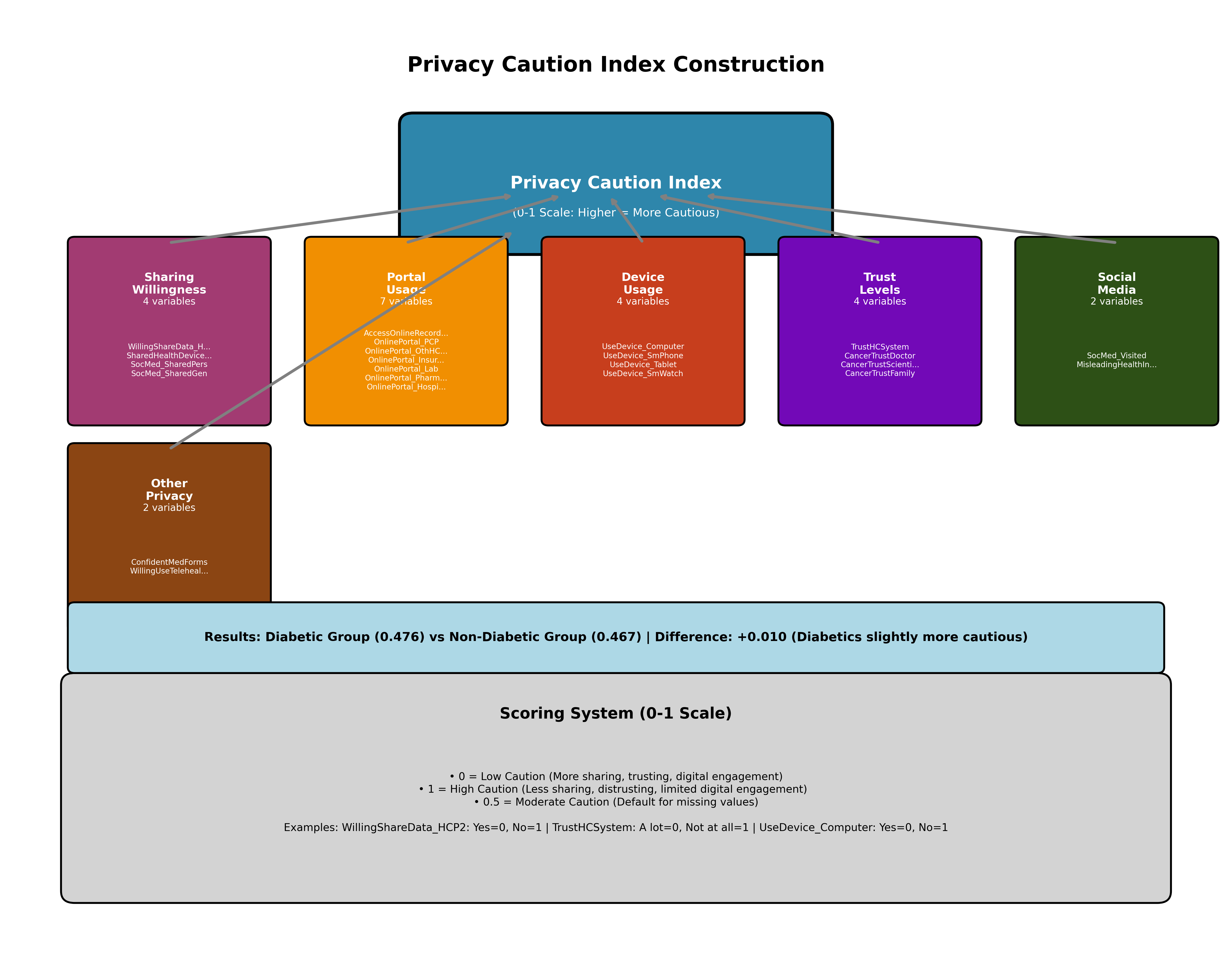
**Appendix G: Nested Model Comparison**

Table A5 presents nested model comparisons showing the incremental contribution of each predictor. The privacy caution index provides the largest improvement in model fit (+13.6 percentage points in R²), confirming its central role in explaining data sharing willingness. The interaction term, while modest in effect size, achieves statistical significance (p = 0.038) and improves model fit marginally.

**Table A5: Model Specification Comparison**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Specification** | **Privacy β** | **Diabetes β** | **Interaction β** | **R²** | **AIC** |
| Base (controls only) | — | — | — | 0.032 | 3,245 |
| + Privacy Index | −2.32\*\*\* | — | — | 0.168 | 2,891 |
| + Diabetes Status | −2.32\*\*\* | 0.03 | — | 0.174 | 2,887 |
| + Interaction Term | −2.44\*\*\* | −0.17† | 0.49\* | 0.175 | 2,883 |

*Note: Nested models showing incremental explanatory power. Privacy index provides largest R² improvement (+13.6 percentage points).*

**Appendix H: Additional Figures**

*A screenshot of a computer screen

AI-generated content may be incorrect.***Figure A2: Privacy Caution Index Construction**

*Note: The Privacy Caution Index aggregates six sub-dimensions (23 variables total) into a 0–1 scale. Higher values indicate greater privacy caution. Diabetic group mean: 0.476; Non-diabetic group mean: 0.467.*

**Table A0: Variable Definitions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type** | **Definition** | **Coding** |
| WillingShareData\_HCP2 | Binary | Willingness to share health data with providers | 0=No, 1=Yes |
| privacy\_caution\_index | Continuous (0–1) | Composite privacy caution index (6 sub-dimensions) | Higher = more cautious |
| diabetic | Binary | Self-reported diabetes diagnosis | 0=No, 1=Yes |
| Diabetes × Privacy | Continuous | Interaction term | Product of above |
| age\_continuous | Continuous | Age in years | 18–100+ |
| education\_numeric | Ordinal (1–6) | Highest education level | 1=<8yrs to 6=Postgrad |
| male | Binary | Gender | 0=Female, 1=Male |
| has\_insurance | Binary | Health insurance coverage | 0=No, 1=Yes |
| urban | Binary | Urban/rural residence | 0=Rural, 1=Urban |
| region\_numeric | Categorical (1–4) | Census region | 1=NE, 2=MW, 3=S, 4=W |

*Note: Privacy Caution Index computed as mean of 6 sub-dimensions (23 HINTS items). Cronbach's α = 0.78.*

**Table A0b: Privacy Caution Index Components**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sub-dimension** | **# Items** | **Variables Included** | **Coding** |
| Sharing Willingness | 4 | WillingShareData\_HCP2, SharedHealthDeviceInfo2, SocMed\_SharedPers, SocMed\_SharedGen | Yes=0, No=1 |
| Portal Usage | 7 | AccessOnlineRecord3, OnlinePortal\_PCP, \_OthHCP, \_Insurer, \_Lab, \_Pharmacy, \_Hospital | Used=0, Not used=1 |
| Device Usage | 4 | UseDevice\_Computer, \_SmPhone, \_Tablet, \_SmWatch | Yes=0, No=1 |
| Trust Levels | 4 | TrustHCSystem, CancerTrustDoctor, \_Scientists, \_Family | A lot=0 to Not at all=1 |
| Social Media | 2 | SocMed\_Visited, MisleadingHealthInfo | Yes=0, No=1 |
| Other Privacy | 2 | ConfidentMedForms, WillingUseTelehealth | Confident=0, Not=1 |

*Note: Missing values imputed as 0.5 (midpoint). Sub-dimensions averaged, then averaged again for composite index.*