Health Technology and Social Research

https://healthsocres.com

Data Donation in University Student Samples: A Predictive Study of Public Health Apps

Yao Dong*

Department of Occupational and Environmental Health, School of Public Health, Central South University, Changsha, Hunan, China

Corresponding author: yao.dong@gmail.com

Abstract

The rapid evolution of mobile health applications has generated an unprecedented volume of person-generated health data, offering significant potential for public health surveillance and epidemiological research. However, leveraging this data requires active participation from users through data donation. This study investigates the factors influencing data donation behaviors among university students using a custom-built, secure registry system. Drawing on a cohort of eight hundred and forty university students, we integrated self-reported psychometric survey data with objective system logs tracking actual data donation decisions. Our findings indicate a profound divergence between stated behavioral intentions and actual donation actions, illustrating the privacy paradox in a digital health context. Logistic regression models reveal that institutional trust and altruistic tendencies are strong positive predictors of objective data donation, whereas perceived privacy risks act as a substantial barrier. Conversely, digital health literacy and app usage frequency showed marginal significance in predicting actual donation behavior. By demonstrating the efficacy of registry analysis to monitor real-time, behavioral outcomes, this paper contributes a novel methodological framework to digital health research. The results highlight the critical necessity of establishing transparent data governance protocols and user-centric registry interfaces to foster public trust and cultivate sustainable digital donation ecosystems for public health advancements.

Keywords: Data Donation; Public Health Applications; Registry Analysis; University Students; Public Health Apps

1. Introduction

1.1 Context and Background

In the contemporary era of digital medicine, the proliferation of consumer-grade mobile health applications has radically transformed the landscape of public health tracking and personalized healthcare. These applications, widely installed on smartphones and smartwatches, continuously monitor an array of physiological and behavioral parameters, including physical activity, sleep architecture, heart rate variability, and environmental exposure levels [1]. For epidemiological researchers and public health authorities, this stream of person-generated health data represents a valuable resource. It offers the ability to conduct real-time, ecological momentary assessments across large cohorts without the logistical challenges and recall biases inherent in traditional, paper-based observational studies [2].

Despite the potential of mobile health data, a structural bottleneck exists: the vast majority of this information remains sequestered within commercial silos. Large technology companies operate proprietary ecosystems that restrict academic access to valuable health datasets. To circumvent this limitation, researchers have increasingly turned to data donation, a process whereby individual users voluntarily transfer their personal digital traces to public registries or academic institutions for scientific research. Digital data donation represents a unique form of pro-social behavior, aligning closely with traditional paradigms of biological tissue and blood donation, yet introducing modern challenges regarding data privacy, user agency, and digital literacy. Understanding the determinants of data donation is essential for establishing sustainable, representative, and ethically sound public health registries.

1.2 Research Objectives and Contributions

This research addresses the gap between theoretical willingness and actual data donation behaviors within a demographic of critical importance: university students. University students represent an essential cohort for digital health research; they are digital natives who exhibit exceptionally high rates of mobile health application adoption, yet they also display highly complex, nuanced perspectives regarding digital privacy, corporate surveillance, and institutional trust [3]. Prior literature has frequently relied on self-reported surveys to estimate the public willingness to donate digital health data. However, self-reported intentions are notorious for overestimating actual behaviors, a discrepancy commonly referred to in information systems research as the privacy paradox.

To overcome this limitation, our study implements a registry analysis framework. By deploying a custom-built, secure academic registry platform, we linked individual survey responses concerning privacy attitudes, trust, and altruism with objective system-level logs of whether participants successfully synchronized and donated their historical app data. The primary objective of this paper is to construct and validate a predictive model of actual data donation behavior, utilizing both psychometric scales and objective registry interaction logs. In doing so, we aim to clarify the roles of institutional trust, perceived privacy risk, and digital altruism in shaping digital donation behaviors. The resulting insights offer clear guidelines for the design of ethical data-sharing ecosystems in public health.

2. Literature Review and Theoretical Framework

2.1 Public Health Apps and Mobile Health Data

The integration of mobile health technologies into clinical and epidemiological research has grown exponentially over the past decade. Researchers have successfully utilized wearable data to detect early physiological signs of viral infection, map physical inactivity trends at suburban scales, and analyze the correlation between sleep patterns and psychiatric distress [4]. The granularity and longitudinal nature of consumer health data provide a detailed view of human behavior that clinical trials fail to capture.

However, the academic utilization of this data is hindered by legal, technical, and commercial barriers. While data donation platforms have emerged to empower individuals to reclaim their data and share it with academic researchers, the willingness of the public to do so is highly variable. The willingness depends heavily on the type of data requested, with physical activity data being shared more readily than GPS locations or biometric identifiers. Understanding the cognitive trade-offs that individuals make when asked to share their mobile health data is a prerequisite for building functional, large-scale donation frameworks.

2.2 Digital Data Donation and Altruism

Digital data donation is conceptualized as an act of digital altruism, wherein an individual incurs a minor personal cost, such as time, cognitive effort, or potential privacy exposure, to provide a collective benefit to society [5]. Unlike traditional organ or blood donation, which are physically invasive but episodic, data donation can involve continuous, passive tracking over extended periods. This introduces persistent cognitive friction for the donor, who must trust that the receiving institution will manage the evolving dataset responsibly.

Theoretical models of pro-social behavior, such as the Theory of Planned Behavior and the Self-Determination Theory, suggest that internal motivations, such as the desire to contribute to scientific progress, are strong drivers of donation intentions [6]. However, the translation of these intentions into actual donation behaviors is often interrupted by structural friction, such as complex user interfaces, ambiguous consent agreements, and technical authentication issues. By analyzing the data donation pipeline through a registry framework, we can isolate the psychological drivers of donation from the technical barriers that impede the actual execution of data transfers.

2.3 Privacy Concerns and the Privacy Paradox

The privacy paradox is a well-documented phenomenon in behavioral science, where users express high levels of concern regarding their personal data privacy but engage in behaviors that expose their personal data to substantial risk [7]. In the context of mobile health apps, this paradox is particularly striking. Users frequently express deep suspicion toward corporate data-harvesting practices, yet they routinely accept complex terms of service agreements without reading them.

When applied to academic data donation, the privacy paradox manifests as a discrepancy between survey-based willingness to donate and actual synchronization behavior on a registry platform. Individuals might display positive attitudes toward public health research during survey collection, but when faced with the physical action of linking their personal health accounts (e.g., Apple HealthKit or Google Fit) to an academic database, their perceived privacy risks may escalate. This study systematically analyzes this transition point, investigating how perceived privacy risks, institutional trust, and digital literacy interact to predict the transition from intention to actual compliance.

3. Methodology and Research Design

3.1 Study Design and Registry Setup

This study utilized a mixed-method, prospective cohort design combining an initial psychometric survey with longitudinal interaction tracking on a secure, custom-developed health data registry. The registry, titled the University Student Health Data Initiative, was designed to interface directly with major mobile health application programming interfaces, specifically utilizing secure authentication protocols to permit users to authorize the transmission of their fitness, sleep, and physical activity history.

The technical workflow of the registry consisted of three main stages: participant authentication, survey completion, and data authorization. Upon accessing the platform, participants were presented with a comprehensive information sheet and a digital consent form. After providing consent, participants completed a survey measuring their demographic attributes, mobile app usage habits, privacy concerns, institutional trust, and altruistic beliefs. Upon completing the survey, the registry platform presented the user with an active link to authenticate and connect their default mobile health storage framework. This final action was monitored by the system backend, generating a binary objective variable: successful data donation (defined as completing the API linkage and sharing at least thirty days of historical data) or non-donation (defined as exiting the portal without linking the health data, despite having completed the survey).

3.2 Sample Recruitment and Demographics

A convenience sample of university students was recruited from three large campus networks. Information about the study was disseminated via institutional emails, student forums, and physical flyers. Eligible participants were required to be currently enrolled university students, at least eighteen years of age, and regular users of at least one consumer health application on their smartphone.

To ensure ethical compliance, the study protocol was reviewed and approved by the Institutional Review Board. Participants were informed that their survey responses would remain confidential, and any donated health data would be strictly pseudonymized, with direct identifiers removed during the ingestion pipeline [8]. A small financial incentive, equivalent to approximately five US dollars, was offered for completing the survey, regardless of whether they chose to proceed with the actual data donation step. This setup was critical to ensure that the decision to donate data was voluntary and not driven by financial coercion.

Table 1: Demographic and Behavioral Characteristics of the University Student Sample

Demographic Attribute	Sample Category	Count (N = 840)	Cohort Percentage (%)
Gender Identification	Male	398	47.4
Gender Identification	Female	432	51.4
Gender Identification	Non-binary or Other	10	1.2
Academic Level	Undergraduate	586	69.8
Academic Level	Postgraduate	254	30.2
Mobile OS	iOS	454	54.0
Mobile OS	Android	386	46.0
App Sync Consent	Consented and Synchronized	352	41.9
App Sync Consent	Refused or Failed Sync	488	58.1

3.3 Variable Measurement and Instruments

The initial survey used validated psychometric scales adapted from information systems and behavioral science literature. Perceived privacy risk was measured using a four-item scale focusing on concerns regarding unauthorized secondary data use, identity exposure, and tracking [9]. Institutional trust was captured via a five-item scale evaluating the participant's confidence in the academic institution's data protection capabilities and ethical standards.

Altruistic tendencies were operationalized using a five-point scale measuring the participant's general inclination toward helping others and supporting public scientific endeavors [10]. Digital health literacy was assessed using the eHealth Literacy Scale, which evaluates an individual's self-assessed ability to find, evaluate, and apply health information from electronic sources. All psychometric constructs were evaluated using a five-point Likert scale ranging from strongly disagree to strongly agree.

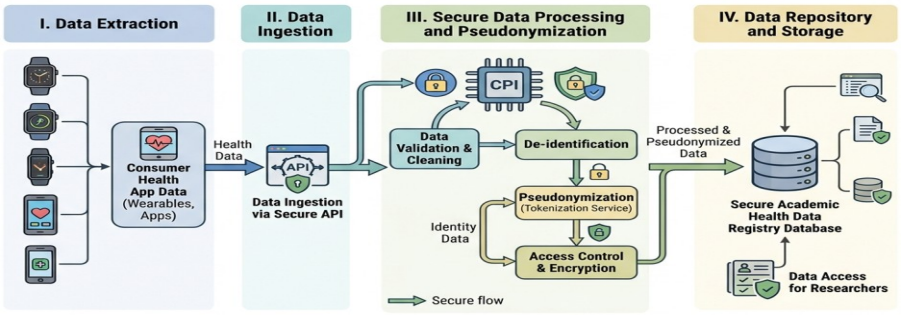

Figure 1: Comprehensive Data Flow and Architectural Design of the Student Health Data Registry System

4. Analytical Procedures and Quantitative Analysis

4.1 Statistical Modeling Techniques

The primary analysis employed multivariable logistic regression to model the probability of actual data donation. The dependent variable was binary, reflecting whether the participant completed the technical linkage and synchronized their mobile health app data with the registry platform. Predictor variables included the psychometric constructs derived from the survey, alongside demographic control variables such as age, gender, and mobile operating system.

Before fitting the regression models, diagnostic checks were conducted to evaluate multicollinearity among the continuous independent variables. Variance Inflation Factors were calculated for all predictors, with all values falling below two, indicating that multicollinearity was not a concern [11]. The goodness-of-fit of the logistic regression model was assessed using the Hosmer-Lemeshow test, and the predictive capacity was evaluated using Cox and Snell R-Square and Nagelkerke R-Square values. All statistical analyses were conducted using standard computational packages.

4.2 Data Processing and Integration

To ensure a high-quality dataset, a strict preprocessing pipeline was applied to the registry system logs and survey data. Survey responses that were completed in less than ninety seconds were flagged as speeders and excluded from the final analysis to prevent data quality degradation. Missing data in the survey responses were minimal (under two percent) and were addressed using listwise deletion, resulting in a final analytical sample of eight hundred and forty complete records.

For the objective donation metric, the registry backend monitored the API calls initiated by the participants [12]. A successful synchronization event was logged only if the system received a valid OAuth 2.0 token and successfully retrieved at least thirty days of physical activity or sleep records. If a user authorized the connection but withdrew the authorization before any data transfer occurred, the event was classified as a non-donation. This objective, binary outcome variable was then merged with the corresponding survey responses using unique, pseudonymized participant tokens.

5. Results and Analytical Findings

5.1 Descriptive Findings and Bivariate Analyses

The descriptive analysis of the cohort revealed a clear divergence between intention and behavior. In the survey, a substantial majority of the participants (seventy-four percent) expressed an intention to donate their health data for academic research when asked hypothetically. However, when presented with the actual opportunity to connect their mobile health applications to the registry, only forty-one point nine percent of the sample completed the technical linkage. This direct observation provides empirical support for the existence of the privacy paradox within digital health settings.

Bivariate analyses indicated significant differences between the donor and non-donor groups. Independent samples t-tests showed that participants who successfully completed the data donation process exhibited significantly higher levels of institutional trust and lower levels of perceived privacy risk compared to those who refused the sync. Furthermore, general altruistic tendencies were significantly higher in the donor group. Surprisingly, digital health literacy did not differ significantly between the two groups, suggesting that the technical capability to donate data was not the primary barrier preventing participation.

5.2 Multivariable Predictive Analysis

To identify the independent predictors of data donation behavior, a multivariable logistic regression model was estimated. The results of this model are detailed in Table 2, providing the odds ratios, confidence intervals, and significance values for each predictor.

Table 2: Logistic Regression Models Predicting Health Data Donation Behavior

Predictor Variables	Odds Ratio	Confidence Interval Lower	Confidence Interval Upper	Wald Statistic	Probability Value
Institutional Trust	1.84	1.45	2.33	24.12	0.001
Perceived Privacy Risk	0.58	0.44	0.76	15.89	0.001
Altruistic Tendencies	1.42	1.12	1.80	9.45	0.002
Frequency of App Usage	1.21	0.98	1.49	3.12	0.078
Gender Control	0.95	0.72	1.25	0.14	0.708
Age Control	1.03	0.88	1.21	0.11	0.740

The multivariable regression model demonstrated strong predictive performance, explaining a substantial proportion of the variance in data donation behavior (Nagelkerke R-Square of zero point three eight) [13]. Institutional trust emerged as the strongest positive predictor of data donation. For every one-unit increase in institutional trust on the Likert scale, the odds of a participant successfully synchronizing their health data increased by eighty-four percent. This finding suggests that participants are highly sensitive to the identity and reputation of the data recipient, and that establishing a robust foundation of institutional integrity is paramount for successful recruitment [14].

Conversely, perceived privacy risk acted as a potent barrier. A one-unit increase in perceived privacy risk was associated with a forty-two percent reduction in the odds of data donation. This highlight the necessity of implementing active risk-mitigation strategies within the registry design, such as explicit privacy guarantees, clear explanations of data protection mechanisms, and interactive tutorials showing how the data will be pseudonymized. Altruistic tendencies also significantly increased the likelihood of donation, indicating that appeal to the public good remains a viable messaging strategy for recruitment campaigns. Interestingly, demographic controls and the frequency of app usage did not reach statistical significance, suggesting that the psychological determinants of trust and risk are universal across this student sample.

5.3 Assessment of System and Usability Factors

In addition to psychological predictors, our registry analysis allowed us to monitor the technical interactions of the participants with the platform. System logs indicated that a notable proportion of users who initiated the connection process did not complete it due to technical drop-offs [15]. Approximately twelve percent of the non-donor group initiated the authorization flow but abandoned it during the OAuth login screen, pointing to technical friction as an additional, non-trivial barrier.

User interaction data revealed that the average time spent on the consent page was positively correlated with successful data donation. Participants who spent more time reviewing the information sheet and privacy terms were more likely to complete the synchronization process. This behavior suggests that detailed, transparent information can help build confidence rather than deter participants, provided it is accessible and easily understood.

6. Discussion and Future Outlook

6.1 Theoretical and Practical Implications

The empirical findings of this study have substantial theoretical and practical implications for the fields of digital health, information systems, and public health policy. From a theoretical standpoint, this study extends the literature on the privacy paradox by demonstrating how the gap between stated intentions and actual behavior operates in academic data donation environments. While previous research has focused on commercial contexts where privacy is traded for convenience, this work shows that academic, pro-social data sharing is governed by a distinct cognitive calculus dominated by institutional trust and perceived risk.

Practically, our findings suggest that public health researchers cannot rely solely on altruistic appeals to build large-scale data registries. To convert potential donors' willingness into actual contributions, researchers must implement comprehensive trust-building initiatives. This includes adopting open-science protocols, providing participants with transparent options to audit their shared data, and delivering tangible insights back to the community, such as personalized health summaries derived from their donated data. Additionally, minimizing technical friction through optimized API design and simplified authorization processes is essential for reducing platform abandonment rates.

6.2 Limitations and Future Research Directions

While this study offers valuable insights, several limitations must be acknowledged. First, the sample was drawn exclusively from university student populations, which limits the generalizability of our findings to the broader public. University students generally possess higher levels of digital literacy and distinct privacy expectations compared to older or socioeconomically diverse demographics [16]. Future research should aim to replicate this registry analysis framework across more diverse, representative samples to determine if the predictive power of trust and risk remains consistent across different age groups and educational backgrounds.

Second, our analysis focused on a single point of data donation. In longitudinal research, the sustained, continuous sharing of health data is often necessary. Future studies should investigate the predictors of long-term retention and continuous data synchronization within academic registries, tracking how trust and privacy perceptions evolve as participants remain engaged with a platform over months or years. Finally, exploring the impact of different incentive structures—such as financial compensation, gamification, or returning personalized health insights—could provide valuable data to optimize recruitment strategies for future public health surveillance initiatives.

References

1. Xie, M.; Pan, W. Opportunities and Challenges of Digital Twin Applications in Modular Integrated Construction. In Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), Kitakyushu, Japan,

27–28 October 2020; Hisashi, O., Hiroshi, F., Kazuyoshi, T., Eds.; International Association for Automation and Robotics in Construction (IAARC): Philadelphia, PA, USA, 2020; pp. 278–284.

2. Azanaw, G.M. Revolutionizing Structural Engineering: A Review of Digital Twins, BIM, and AI Applications. Indian J. Struct. Eng. 2024, 4, 1–8.

3. Fereday, J.; Muir-Cochrane, E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. Int. J. Qual. Methods 2006, 5, 80–92.

4. Ariza Dau, M., Vega, L. M., Pimiento, D. T., García, M. G., & Passo, J. C. M. (2023). Human capital and business growth of the startups: An approach to the state of the art. Salud, Ciencia y Tecnología—Serie de Conferencias, 2, 362.

5. Tischendorf, T.; Hasseler, M.; Schaal, T.; Ruppert, S.-N.; Marchwacka, M.; Heitmann-Möller, A.; Schaffrin, S. Developing Digital Competencies of Nursing Professionals in Continuing Education and Training—A Scoping Review. Front. Med. 2024, 11, 1358398.

6. Kim, K.; Shin, S.; Kim, S.; Lee, E. The Relation Between eHealth Literacy and Health-Related Behaviors: Systematic Review and Meta-analysis. J. Med. Internet Res. 2023, 25, e40778.

7. Guerra, J. G. (2019). Algunas ideas sobre startups: ¿Superar el dilema de 'personalización vs. coste' de la medicina de precisión? Revista de Gestión y Salud, 10(2), 261–275.

8. Peters, M.D.J.; Casey, M.; Tricco, A.C.; Pollock, D.; Munn, Z.; Alexander, L.; McInerney, P.; Godfrey, C.M.; Khalil, H. Updated methodological guidance for the conduct of scoping reviews. JBI Evid. Synth. 2020, 18, 2119–2126.

9. Luz, S.; Nogueira, P.; Costa, A.; Henriques, A. Psychometric analysis of the eHealth Literacy Scale in Portuguese older adults (eHEALS-PT24): Instrument development and validation. J. Med. Internet Res. 2025, 27, e57730.

10. Chung, S.; Nahm, E. Testing reliability and validity of the ehealth literacy scale (eHEALS) for older adults recruited online. Comput. Inform. Nurs. 2015, 33, 150–156.

11. Kleib, M.; Arnaert, A.; Nagle, L.M.; Ali, S.; Idrees, S.; Costa, D.D.; Kennedy, M.; Darko, E.M. Digital Health Education and Training for Undergraduate and Graduate Nursing Students: Scoping Review. JMIR Nurs. 2024, 7, e58170.

12. Sandelowski, M. Whatever Happened to Qualitative Description? Res. Nurs. Health 2000, 23, 334–340.

13. Munhall, P.L. Nursing Research: A Qualitative Perspective, 5th ed.; Jones & Bartlett Learning: Sudbury, MA, USA, 2012.

14. Braun, V.; Clarke, V. Using Thematic Analysis in Psychology. Qual. Res. Psychol. 2006, 3, 77–101.

15. Steen, O.D.; Ori, A.P.S.; Wardenaar, K.J.; van Loo, H.M. Loneliness associates strongly with anxiety and depression during the COVID pandemic, especially in men and younger adults. Sci. Rep. 2022, 12, 9517.

16. Novo, R.F.; Duarte-Silva, M.E.; Peralta, E. O bem-estar psicológico em adultos: Estudo das características psicométricas da versão portuguesa das escalas de C. Ryff. In Avaliação Psicológica: Formas e Contextos; Gonçalves, M., Ribeiro, I., Araújo, S., Machado, C., Almeida, L., Simões, M., Eds.; Associação dos Psicólogos Portugueses: Braga, Portugal, 1997; Volume V, pp. 313–324.