Privacy Concerns and Digital Engagement: Evidence from a National Panel Survey

Sueyoul Kim*

May 28, 2026

Abstract

Growing privacy concerns have provided important background for the introduction of personal data protection laws, such as the General Data Protection Regulation (GDPR) in the EU. However, empirical evidence on the relationship between privacy concerns and individual behavior remains limited. To fill this gap, this paper uses national-level panel survey data that directly measure individuals' privacy concerns along with a wide range of digital behaviors. Two empirical approaches—fixed-effects regressions and a difference-in-differences–style analysis exploiting differential exposure to a major data breach—yield consistent patterns: both higher privacy concerns and greater exposure to the data breach are associated with lower engagement in internet communities, reduced online content contribution, and lower digital consumption such as app purchases. These results suggest potential welfare gains from data protection laws if they successfully mitigate privacy concerns.

*Korea Development Institute. Email: sueyoul.econ@gmail.com. This paper was previously titled "The Impact of Privacy Concern on Consumer Behaviors". I am grateful for the thoughtful comments from the two editors, Luis Aguiar and Ulrich Kaiser, and the two anonymous referees, which significantly improved the paper. I am deeply indebted to my advisor, Ginger Jin, for her constant guidance and support. I appreciate the suggestions from Mario Leccese, Andrew Sweeting, Chenyu Yang, and Hyungjin Kim. Thanks go to Jaehong Choi, Keaton Ellis, Seho Kim, Michael Navarrete, Rachel Nesbit, and Ece Yagane for their encouragement. All errors are my own. This paper is the final working-paper version. The published version is available at https://doi.org/10.1016/j.infoecopol.2026.101164.

1 Introduction

Growing concerns about privacy—defined as "control or protection of personal information" (Acquisti et al., 2016)—have become a crucial issue in the digital era, as numerous companies collect detailed user information at an unprecedented scale. While platforms claim they use the data to provide better services, privacy violations have repeatedly led to significant financial penalties and public distrust. For example, TikTok, a social platform with more than one billion monthly active users, reached a USD 92 million class-action settlement for sharing users' biometric data with third-party companies without consent in August 2022.[1] Reflecting these concerns, 64% of Americans express worry about how TikTok uses user data, including nearly half of TikTok's own users (Pew Research, 2023).

Heightened privacy concerns have provided important background for the introduction of major laws that strengthen personal data protection. A representative example is the EU's General Data Protection Regulation (GDPR), adopted in 2016. EU Commissioner for Justice and Commission Vice-President Viviane Reding emphasized: "The protection of personal data is a fundamental right for all Europeans, but citizens do not always feel in full control of their personal data."[2] The core principles of the GDPR require firms to collect, process, and retain only data that is necessary for specific purposes, and only for as long as necessary.[3] Together, these principles and Reding's statement reflect the view that concerns about losing control over personal data—that is, privacy concerns—have been central to the motivation for modern data protection regulation, including the GDPR.

This paper informs the policy debate on data protection laws by empirically studying the relationship between privacy concerns and digital engagement behaviors. I document three main findings. First, higher privacy concerns are associated with lower levels of digital engagement, including reduced posting frequency on internet communities (e.g., Reddit) and knowledge-sharing platforms (e.g., Wikipedia). Second, privacy concerns are similarly linked to lower usage of email and cloud services, as well as reduced digital consumption such as paid app purchases. Finally, social networking site (SNS) usage shows no systematic association with privacy concerns in the fixed-effects specifications, but exhibits a statistically significant negative association with exposure to the Facebook–Cambridge Analytica scandal.

These findings imply potential benefits of data protection laws. In particular, effective data protection may help maintain user engagement and preserve the value of the internet as a platform for community building and knowledge sharing. User engagement is a fundamental driver of value for numerous websites, ranging from crowdsourced problem solving (e.g., receiving assistance with coding on Stack Exchange), to community building (e.g., finding local activity partners through Facebook groups), and to

[1] https://www.nbcchicago.com/news/local/judge-approves-92-million-tiktok-settlement-with-illinois-claimants-receiving-biggest-share/2921881/

[2] https://ec.europa.eu/commission/presscorner/detail/en/ip_12_46

[3] See Article 5, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

knowledge dissemination (e.g., restaurant reviews on Yelp). Given that billions of users benefit from these platforms, the costs of reduced engagement could be substantial. If effective data protection mitigates privacy concerns, it may help preserve the overall value of the internet. Moreover, the reduction in paid app purchases suggests that privacy concerns may cause consumers to hesitate in making purchasing decisions that involve the disclosure of personal information, representing another dimension along which data protection laws could be valuable.

To document these patterns, this paper uses a nationally representative panel survey from South Korea. A unique strength of the dataset is that it directly measures both privacy concerns and digital behaviors at the individual level over multiple years. Trained interviewers visit households and ask questions about a wide range of digital engagement and media usage, including how often panelists post to online communities, which social media platforms (e.g., Facebook) they use, their paid app purchases, and, importantly, their privacy concerns. This rich dataset provides a valuable opportunity to examine the relationship between privacy concerns and a broad set of digital behaviors and consumption.

While the data were collected in South Korea, the findings of this paper are likely to be broadly applicable for two reasons. First, South Korea has a digital environment similar to that of many developed countries. Pew Research (2022) shows that South Korea has a high internet penetration rate (99%), comparable to other advanced economies such as Canada (95%), Germany (93%), and the UK (93%). Smartphone ownership among adults in South Korea is also high (94%), similar to rates in the US (85%), Netherlands (90%), and Japan (84%). Finally, global online platforms such as Google, YouTube, and Facebook are widely used in Korea alongside local competitors such as Naver (search engine) and Kakao (messaging app), mirroring patterns observed in many other regions.

Second, privacy concerns and related data protection laws are global phenomena. For example, Pew Research (2019) reveals that a majority of Americans (79%) report being concerned about how companies use their data. Similarly, an EU survey found that 55% of internet users are concerned that criminals or fraudsters might access their shared personal information without their knowledge.[4] Data protection laws are being discussed and implemented widely, including not only the previously mentioned GDPR but also the California Consumer Privacy Act (CCPA) introduced in January 2018 and amendments to Japan's Act on the Protection of Personal Information (APPI) enacted in June 2020. These widespread concerns and regulatory responses make the findings from this paper informative beyond Korea.

For empirical analysis, two key confounding factors complicate identifying the causal effects of privacy concerns: reverse causality and unobserved individual heterogeneity. For example, posting personal experiences on Reddit may heighten privacy concerns, whereas this paper aims to study the opposite relationship—how privacy concerns affect engagement in internet communities. To address these issues, I take two complementary

[4] https://eupinions.eu/de/blog/more-than-half-of-all-eu-citizens-are-concerned-about-their-online-data

approaches. First, I include individual fixed effects and use one-year-lagged privacy concerns as the independent variable. Second, I leverage a plausibly exogenous event—a major Facebook data breach disclosure in 2018—to estimate difference-in-differences–style regressions exploiting differential exposure between Facebook users and non-users before and after the disclosure. While these approaches are not a panacea, they help mitigate confounding factors and reveal patterns that go beyond simple correlations.

While the two approaches differ substantially, the findings are largely consistent across them. Both (lagged) measured privacy concerns and exposure to the personal data breach are negatively associated with engagement in internet communities and online content provision. Email and cloud service usage, and paid app purchases are also negatively influenced. The expected time for replacing the current smartphone increased, potentially due to concerns about personal information leakage from discarded devices. Finally, the coefficient on SNS usage shows a significant negative sign only for the Facebook scandal rather than for general privacy concerns measured.

This paper proceeds as follows. The rest of the introduction reviews the related literature. Section 2 describes the dataset. Section 3 presents the conceptual framework and Section 4 the empirical strategy. Section 5 presents results and robustness checks. Section 6 discusses implications and concludes.

1.1 Related Literature

This paper contributes to three strands of literature. The first relates to the valuation and influence of privacy concerns in economics and marketing. For example, Prince and Wallsten (2022) directly investigates the valuation of privacy using discrete choice surveys and documents substantial heterogeneity across countries and data types. Prince and Wallsten (2025) studies whether people have a preference for data localization—i.e., prohibiting information from being sent outside the country—and finds that respondents in the seven surveyed countries place little to no value on such restrictions. A large body of work in theory (Acemoglu et al., 2022; Ichihashi, 2023; Miklós-Thal et al., 2024), experiments (Lin, 2022; Lee and Weber, 2025), and empirical analyses using consumer behavior data (Acquisti et al., 2013; Beresford et al., 2012; Tsai et al., 2011; Goldfarb and Tucker, 2012; Tucker, 2014) has contributed to this literature. Acquisti et al. (2016, 2015) provide comprehensive overviews of the economics of privacy.

With respect to this first strand of the literature, this paper provides novel empirical evidence on the links between privacy concerns and consumer-side behaviors, including online posting and app purchases. In particular, it sheds light on the previously understudied relationship between privacy concerns and app purchases. The direction of this relationship is ambiguous: privacy concerns may increase app spending because privacy-protecting apps tend to be more expensive (Kummer and Schulte, 2019; Cecere et al., 2022; Kesler, 2023), but they may also decrease app spending by making consumers more hesitant to submit personal or payment information (e.g., credit card information). This paper finds a negative association between privacy concerns and app purchases, suggesting that the latter mechanism dominates.

In addition, the paper provides empirical support for consumer behavior patterns

suggested in theoretical work. For example, Miklós-Thal et al. (2024) develop a theoretical model showing that privacy concerns may induce some users to share no data ("digital hermits"). Consistent with this prediction, the paper documents a negative association between privacy concerns and online posting frequency.

The second strand of literature examines personal data protection laws such as the GDPR and the CCPA. Research in this area primarily studies firms' responses and the (un)intended side effects of these laws, such as negative impacts on venture investment and increased market concentration (Goldberg et al., 2024; Jia et al., 2021; Aridor et al., 2023; Johnson et al., 2023). Johnson (2024) provides a comprehensive review of the economic consequences of the GDPR. Notably, existing work tends to focus on the supply side of the economy—that is, on firms and the potential negative effects arising from regulation.

This paper contributes to these debates by providing suggestive evidence on the *benefits* of data protection laws. The empirical analysis indicates that mitigating privacy concerns may increase engagement in online communities and knowledge-sharing platforms. Thus, if data protection laws are effective in reducing privacy concerns—as intended by policymakers–they could generate substantial welfare gains by supporting active information dissemination, advising, and networking in online spaces.[5] Given the scale of these platforms (e.g., Quora has over 400M active users), even small increases in participation could yield sizable social benefits.

The third strand of literature focuses on the "privacy paradox"—the phenomenon that stated privacy concerns often do *not* align with actual behavior. For example, although MIT students reported that they cared about privacy, they disclosed the email addresses of close friends in exchange for pizza (Athey et al., 2017). This paradox has been widely studied in economics, marketing, and legal scholarship (Barnes, 2006; Chen et al., 2021; Stutzman et al., 2013; Liao et al., 2024). Within this stream of work, this paper provides suggestive evidence that privacy concerns do translate into behavior in certain digital contexts (e.g., reduced online knowledge sharing), indicating that the paradox may not hold universally.

2 Data

2.1 Korea Media Panel Survey (KMPS)

The main dataset of this paper is the 2017–2022 sample of the Korean Media Panel Survey (KMPS). KMPS is an annual panel survey conducted by a Korean government-affiliated institute called the Korea Information Society Development Institute (KISDI).[6] The aim of KMPS is to provide a comprehensive and in-depth understanding of media usage trends. Examples of variables include possession and use status of televisions,

[5]Evidence on GDPR's effectiveness in reducing privacy concerns remains limited and warrants further investigation (Presthus and Sørum, 2021).

[6]The data is publicly available here: `https://stat.kisdi.re.kr/kor/contents/ContentsList.html?subject=&sub_div=E`. An example of a research paper that leveraged survey data from KISDI is Lee (2018) on quantifying consumer surplus from smartphone adoption.

smartphones, and wearable devices; broadcast communication service subscription and expenditure; restrictions on household media use; and media usage behavior related to over-the-top (OTT) services, e-mail, and cloud services.

Three features of KMPS make it attractive for studying the links between privacy concerns and behavior. First, KMPS measures privacy concerns by asking panelists eight privacy-related questions, and offers a wide range of variables about online engagement behaviors at the individual-year level. Second, KMPS has a panel structure. These first two features allow researchers to directly study how privacy concerns affect online behaviors while controlling for unobserved heterogeneity through fixed effects. Third, KMPS is a national-level, high-quality survey. Therefore, findings from KMPS are likely to be more credible and generalizable than findings from small, selected groups of respondents such as Amazon Mechanical Turk survey participants.[7]

2.2 The Privacy Concerns Measure

Privacy concerns are measured for all participating panelists each June, when the survey is administered. Panelists respond to eight privacy-related statements by selecting the option (e.g., "agree") that best reflects their views. The statements are as follows:

(1) I worry that someone I do not know might obtain my personal information from my online activity.
(2) I worry that my private information might be stored on the devices (PC, smartphone) that I use.
(3) I worry that my private information might be online without my knowledge.
(4) I worry that websites require too much information from me when I register.
(5) I worry about online identity theft.
(6) In general, I worry about my privacy when I use the internet.
(7) People online who do not clarify their identity are suspicious.
(8) I worry about identity theft involving my personal information, such as my profile picture and name.

To conduct quantitative analysis, I construct an individual-year measure of privacy concerns as follows:

$$PrivacyConcern_{it} = \frac{1}{8}\sum_{j=1}^{8} Q_{itj} \tag{1}$$

where Q_{itj} denotes panelist i's response in year t to privacy-related question j. Responses are coded numerically as follows: "strongly agree" = 4, "agree" = 3, "neutral"

[7]Specifically, Nielsen Korea contracts with KISDI to hire, educate, and monitor interviewers who help respondents answer accurately. All answers are cross-validated by a separate inspector. Lastly, KISDI allocates sample weights to account for non-response rates and make the survey nationally representative.

= 2, "disagree" = 1, and "strongly disagree" = 0. Higher values of this variable indicate greater privacy concerns.

Do the questions in the KMPS survey measure privacy concerns in a sensible way? I compare the questions with responses from Pew Research (2019), and find that the KMPS questions and the responses are well-aligned, even though the two surveys were conducted in different countries (Korea and the US). This finding adds to the credibility and potential general applicability of the data.

Specifically, Pew Research (2019) asked U.S. adults the open-ended question, "What does privacy mean to you?" The most commonly mentioned themes were: "Other people and organizations not being able to access their possessions or private life" (28%); "Control over information, possessions, self; deciding what's accessible to others" (26%); and "Themselves, their personal information and possessions, the desire to keep things to themselves" (15%). Taken together, these responses suggest that a central aspect of privacy is maintaining control over one's personal information. Indeed, "Privacy is not the opposite of sharing—rather, it is control over sharing." (Acquisti, Taylor and Wagman, 2016)

The way the KMPS measures privacy concerns appears consistent with this definition. For example, Questions (1)–(3) clearly align with the popular definition used in the Pew survey, and Question (6) directly refers to privacy concerns. For robustness checks in Section 5.6, I use an alternative measure of privacy concerns based on Questions (1) and (6), which more directly capture privacy concerns. I confirm that the main findings of this paper remain unchanged.

Privacy Concerns by Demographic Characteristics

The privacy concerns measure shows substantial variation across demographic groups. Figure 1 illustrates these patterns. Overall, privacy concerns tend to increase with income, consistent with the possibility that higher-income individuals perceive greater potential losses from personal-information–related crimes (e.g., phone scams). Women also report slightly higher levels of privacy concerns than men.

Trends over time are less clear; however, privacy concerns consistently display an inverted U-shaped relationship with age, which may partly reflect lower levels of digital engagement among older individuals.

These relationships between privacy concerns and demographic characteristics are further confirmed in Table 1, which reports linear regression estimates of the privacy concerns measure on demographic covariates. The estimates indicate that individuals with higher income tend to exhibit greater privacy concerns and that privacy concerns display an inverted-U relationship with age. These patterns remain robust to the inclusion of year fixed effects and to the use of an alternative privacy concerns measure constructed from a subset of two questions that are more directly related to privacy.

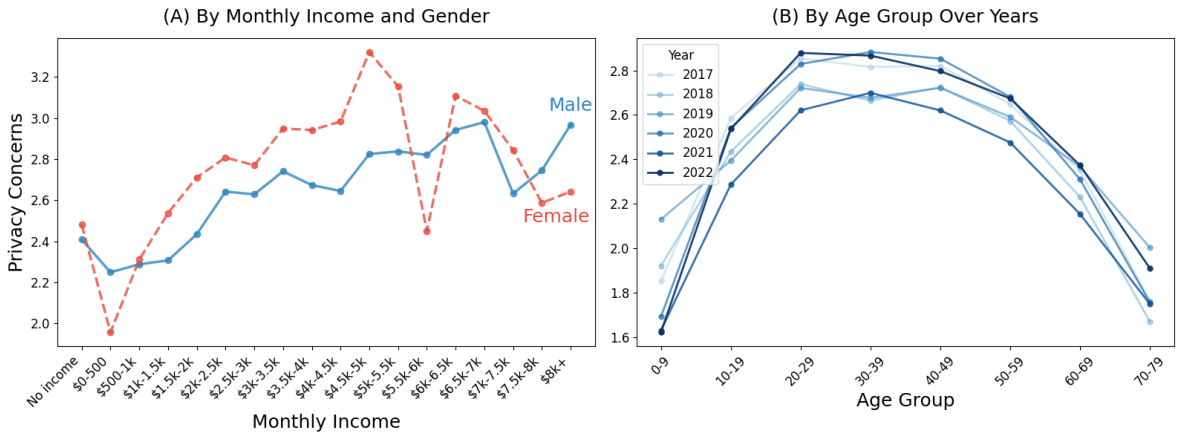

Figure 1: Privacy concerns across demographics

Notes: Income is converted at 1,000 KRW = 1 USD. Averages computed using population weights.

Table 1: Demographic Correlates of Privacy Concerns

Dep. Var:	Privacy Concerns Measure		
	(1)	(2)	(3)
School years	0.0441***	0.0442***	0.0421***
	(0.0020)	(0.0020)	(0.0030)
Income (1,000 USD/month)	0.0266***	0.0267***	0.0246***
	(0.0040)	(0.0040)	(0.0040)
Married (dummy)	0.0512***	0.0502***	0.0511***
	(0.0170)	(0.0170)	(0.0180)
Religious (dummy)	0.0803***	0.0832***	0.0874***
	(0.0120)	(0.0120)	(0.0130)
Age (years)	0.0205***	0.0205***	0.0193***
	(0.0020)	(0.0020)	(0.0020)
Age^2	-0.0003***	-0.0003***	-0.0003***
	(0.00002)	(0.00002)	(0.00002)
Privacy measure	Full (8 items)	Full (8 items)	2 items
Year FE		✓	✓
Observations	52,720	52,720	52,720
R^2	0.100	0.106	0.087

Notes: The dependent variable is a privacy concerns measure on a 0–4 scale. Columns (1) and (2) use the full measure constructed from all eight survey questions, while column (3) uses an alternative measure based on two directly privacy-related questions (Questions (1) and (6); see Section 2.2). All specifications are estimated using linear regression with population weights. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

2.3 Dependent Variables

KMPS investigates various media-related and online activities. Typically, KMPS first asks if a panelist is engaging in a specific activity (e.g., posting on knowledge contribution platforms). Then KMPS asks about the frequency of the activity, and for some items, which specific service (e.g., among SNS options: Facebook, Twitter, or Instagram) the panelist uses.

Among the variables available in the KMPS, I select outcome measures that could plausibly be influenced by privacy concerns. Because privacy can be compromised through any content uploaded online, the first set of outcomes covers a broad range of online engagement behaviors: (1) whether a panelist participates in any online communities and, if so, how frequently they upload posts or leave comments; and (2) whether a panelist contributes user-generated content on platforms such as Stack Exchange and, if so, how often.

In addition, personal information can be leaked through the use of online services. Accordingly, I examine (3) email and cloud-service usage. Privacy concerns may also increase reluctance to submit credit-card or address information, so I analyze (4) app purchases and associated spending. I further study (5) the smartphone replacement cycle, as discarded devices can be a source of personal-data leakage (Roberts et al., 2023), a risk that is salient to Korean consumers (Park, 2023). Finally, as an additional form of online engagement, I examine (6) social networking service (SNS) usage (e.g., Instagram).

I elaborate on the dependent variables of interest in the remainder of this section.

Internet Community Engagement

An internet community is an online platform where people with similar interests (e.g., fishing, stamp collecting) gather and interact. Examples include Reddit, 9GAG, Blind, and topic-specific Facebook groups. In Korea, Naver and Daum's (popular portal sites) cafe tabs and DCInside would be relevant examples. Compared to SNS, these websites are more interest-based and allow more pseudonymous participation.

KMPS asks if a panelist is a member of an internet community and how often they engage in reading, commenting, or sharing posts. For frequency questions, panelists choose one option from choices such as "rarely", "1-3 times a week", ... "Almost every day". I convert these responses to numeric values for quantitative analysis. For example, "1-3 times a week" is converted to 8 ($= 2 \times 4$), as my unit is times per month. Appendix B presents coding details for converting responses to numeric values.

Online Content Contribution

KMPS assesses a panelist's contribution to online content by how often they engage in the following activities: asking and answering questions on knowledge contribution platforms (e.g., Wikipedia, Stack Exchange), voting in online surveys or polls, and making recommendations online (e.g., restaurant reviews on Yelp). I convert responses to these items to numeric values and use them as dependent variables.

Digital Service Usage and App Spending

KMPS records whether a panelist uses an email service or a cloud service (e.g., Dropbox) as dummy variables. In addition, for panelists who made any paid app purchases, the survey provides the annual number of paid apps purchased and the total amount spent on these purchases. More specifically, app spending refers to the amount of money spent on app purchases, including both paid apps and paid items within apps (e.g., in-game purchases). The time horizon for this spending is from July of the previous year to June of the survey year, as KMPS is administered every June. All spending amounts are reported in Korean won (KRW).

Smartphone Replacement

Panelists report the expected remaining time before replacing their current smartphone, which serves as an indicator of their willingness to replace the device sooner or later. Specifically, they choose from categorical time intervals (e.g., "within 6 months," "6 months–1 year") to indicate when they expect to replace their smartphone.

Social Networking Sites (SNS)

Regarding SNS, KMPS investigates the following: whether a panelist has an SNS account and the up to three specific services that a panelist uses (e.g., Facebook, Instagram). From this information, I construct dummy variables for general SNS usage and specific services. For example, if a panelist reports that they use SNS, and it is only Instagram, the SNS and Instagram dummies equal 1 for this individual-year observation, while dummies for other SNSs like Twitter become zero.

Data Summary

Table 2 summarizes variable definitions and descriptive statistics. The main dataset is an unbalanced individual–year panel covering the period 2017–2022. It contains 12,763 unique individuals and 52,720 individual–year observations. The number of observed panelists varies across years, ranging from 7,836 in 2017 to 9,484 in 2022. This variation reflects panel attrition, the recruitment of new participants, and sample restrictions. In particular, respondents who reported that they do not engage in online activities were excluded, as their privacy concerns cannot be measured. As shown in Figure A1, excluded respondents are predominantly under age 10 or over age 70.

Table 2: Variable Descriptions and Descriptive Statistics

Variable	**Description**	**Mean**	**Std**
(Independent Variable)			
PrivacyConcern	The numeric measure of privacy concerns	2.56	0.99
(Dependent Variables)			
$\mathbb{1}$\{UsesIC\}	Dummy variable for using any internet communities (IC)	0.18	0.39
Posting	Panelist posts on ICs	0.64	2.73
Sharing	Panelist shares posts on ICs	0.56	2.61
Commenting	Panelist comments on ICs	1.06	3.93
Question posting	Panelist posts questions online	0.21	1.51
Answering	Panelist answers questions online	0.19	1.50
Voting	Panelist votes in online polls or shares opinions	0.33	1.92
Recommendation	Panelist makes recommendations online	0.46	2.30
$\mathbb{1}$\{Email\}	Dummy variable for using email	0.70	0.46
$\mathbb{1}$\{CloudSvc\}	Dummy variable for using cloud services	0.19	0.40
$\mathbb{1}$\{Purchase\}	Dummy for purchasing any mobile apps (past 12 months)	0.05	0.22
# of Apps	Number of mobile apps purchased	0.12	0.92
Spending	Spending on mobile apps (USD)	1.78	10.87
Replacement Time	Expected time until smartphone replacement (months)	24.34	12.84
$\mathbb{1}$\{SNS\}	Dummy variable for using any SNS	0.56	0.50
$\mathbb{1}$\{Facebook\}	Dummy variable for using Facebook	0.28	0.45
$\mathbb{1}$\{Twitter\}	Dummy variable for using Twitter	0.10	0.30
$\mathbb{1}$\{Instagram\}	Dummy variable for using Instagram	0.26	0.44
$\mathbb{1}$\{KakaoStory\}	Dummy variable for using KakaoStory	0.22	0.41
$\mathbb{1}$\{NaverBand\}	Dummy variable for using NaverBand	0.15	0.36

Notes: The unit for activity variables (e.g., posting) is times per month. The unit of replacement time is months. The observation level is individual-year. For income and app purchase spending, the approximation 1000 KRW $\approx$ 1 USD was used. Population weights were incorporated when computing means and standard deviations.

3 Conceptual Framework

This section presents a conceptual framework illustrating that the relationship between privacy concerns and digital engagement behaviors is theoretically ambiguous. This ambiguity motivates the need for empirical analysis. Consider the following generic consumer utility function that incorporates privacy concerns:

$$u(e) = b(e) - c(e) \quad (b'(e) > 0, b''(e) < 0) \tag{2}$$

where $e > 0$ measures any engagement behavior that improves utility but imposes privacy costs (e.g., uploading daily life content and getting comments on Facebook). $b(e)$ represents the utility benefit and $c(e)$ represents the privacy cost from this behavior. Note that while I use the term "engagement" for simplicity, this framework applies to a broader range of digital behaviors, such as using a cloud service. The optimal engagement of the consumer e^* is determined at the point where $b'(e) = c'(e)$.

This general model generates starkly different predictions depending on the specific functional form of $c(e)$ for how heightened privacy concerns affect the agent's engagement decisions.[8]

Case 1: $c(e) = e^2$; under heightened privacy concerns, $c(e) \to c^{new}(e) = (e+1)^2$.
In this case, heightened privacy concerns cause a *decrease* in engagement, e.g., uploading posts on one's blog less often.

Case 2: $c(e) = \sqrt{e}$; under heightened privacy concerns, $c(e) \to c^{new}(e) = \sqrt{e+1}$.
In this case, heightened privacy concerns cause an *increase* in engagement, e.g., uploading posts on one's blog more often.

The takeaway is that theoretical models may not provide clear or general predictions about how privacy concerns affect engagement decisions. This limitation arises from two factors. First, as the examples show, concave or convex specifications of $c(e)$ yield very different predictions, yet there is no consensus on the shape of this function. The shape depends on numerous factors, such as the psychological costs to consumers, firms' ability to infer consumer types, and the risk of personal information misuse (e.g., credit card cloning). This fundamental difficulty is well-recognized in the literature. As Acquisti, Taylor and Wagman (2016) observe, "characterizing a single unifying economic theory of privacy is hard, because privacy issues of economic relevance arise in widely diverse contexts."

Second, it remains unclear how privacy concerns measured in the real-world data should be mathematically interpreted in theoretical models. For example, the existing literature often decomposes $c(e)$ into two elements, $c(e) = v \cdot p(e)$, where $p(e)$ represents a more objective cost of privacy (e.g., how accurately a firm can infer a consumer's type), and v represents how much the agent worries about it (Ichihashi, 2023; Miklós-Thal et al., 2024). However, how should the model interpret the privacy concerns measure

[8] I use the term "heightened privacy concerns" to refer to both increases in directly measured privacy concerns and external events (e.g., major data breaches) that are likely to worsen such concerns or increase sensitivity to them.

in a survey? The measure could reflect v (e.g., the psychological cost for disclosed personal information), or $p(e)$ (e.g., firms may infer a consumer's type more accurately, and consumers recognize it).

This paper does not attempt to identify the right theoretical framework but focuses on empirical links between privacy concerns and a wide range of digital behaviors. Such empirical patterns could prove informative when more specific privacy models emerge and generate conflicting predictions about privacy concerns and digital behaviors.

4 Empirical Strategy

The goal of this paper is to explore the relationship between privacy concerns and online engagement behaviors. Establishing clean causal effects is challenging because credible exogenous variation in privacy concerns is sparse. I take two approaches to address endogeneity issues: fixed effects regressions using lagged privacy concerns, and a difference-in-differences–style approach that exploits differential exposure to the Facebook–Cambridge Analytica data breach disclosure in March 2018. While these two approaches are not ideal experiments, they help reveal patterns beyond naive correlations. Moreover, the key findings of the paper (e.g., privacy concerns are negatively associated with online information sharing) are consistent across both approaches, which enhances the credibility of the results.

The threat of endogeneity likely comes through two channels: 1) reverse causality and 2) unobserved individual heterogeneity. First, panelists who upload more posts about their daily lives may become more concerned about their privacy. This reverse causality creates potential biases that obscure the causal effects of privacy concerns. Second, some panelists may inherently care more about both positive and negative responses from others. Such tendencies could lead them to post more on online communities to receive encouraging comments while also worrying more about the risks of their private or undesirable activities being revealed—thus exhibiting higher privacy concerns. This cross-individual variation, however, is not informative for evaluating the effects of privacy concerns because it reflects unobserved personal characteristics rather than changes in privacy concerns themselves.

4.1 Individual Fixed-Effects Approach

I employ the following regression specification to address two plausible confounding factors: reverse causality and unobserved individual heterogeneity.

$$Y_{it} = \beta^{PrivConc} PrivacyConcern_{it-1} + \lambda_i + \delta_t + \varepsilon_{it} \tag{3}$$

where subscripts i and t index individuals and years, respectively; $PrivacyConcern$ is the numeric measure constructed in (1); and λ_i and δ_t denote individual and year fixed effects.

where subscripts i and t represent an individual panelist and year, respectively; $PrivacyConcern$ is the numeric measure constructed in (1); λ_i and δ_t are panelist and

year fixed effects.

This specification incorporates two features to address potential sources of endogeneity. First, I use lagged privacy concerns as the independent variable to mitigate reverse causality. Second, I include individual fixed effects to control for unobserved time-invariant heterogeneity across individuals.[9]

4.2 Difference-in-Differences–Style Analysis

Additionally, I leverage a major personal information breach disclosure—the Facebook–Cambridge Analytica scandal—to implement a difference-in-differences–style analysis that exploits differential exposure to the shock across individuals. In March 2018, a whistleblower disclosed that a political consulting firm, Cambridge Analytica, had been collecting personal data from up to 87 million Facebook users without their consent since 2014. Cambridge Analytica used this data to create psychological profiles of users and target political advertisements, particularly during the 2016 U.S. presidential election. As a result, Facebook was later fined $5 billion by the Federal Trade Commission for its privacy violations.

This scandal also attracted significant public attention in South Korea. Figure 2 shows that, when the scandal was disclosed, search volumes for privacy-related keywords, including "personal information breach" and "privacy" increased sharply on Naver, the largest search engine in Korea.

Motivated by this observation, I adopt a difference-in-differences–style approach that exploits the Facebook–Cambridge Analytica scandal as a common shock with heterogeneous exposure across individuals. In particular, I compare Facebook users as of 2017 and non-users before and after the disclosure. In practice, I estimate the following regression:

$$Y_{it} = \beta^{DD} \left(\mathbb{1}\{\text{2017 FB user}\}_i \times \mathbb{1}\{\text{post-2018}\}_t \right) + \lambda_i + \delta_t + \varepsilon_{it} \tag{4}$$

where the two indicator variables represent whether the panelist was using Facebook prior to the disclosure of the scandal in March 2018 (KMPS is conducted every June, so I rely on the 2017 survey to identify this), and whether the observation is from 2018 or later, respectively.

Two caveats are worth noting. First, rather than relying on clearly defined treatment and control groups as in a canonical difference-in-differences design, this approach captures differential exposure across groups that are more or less affected by the shock. Second, because only one pre-event period is available, it is difficult to assess the parallel trends assumption. Accordingly, the estimates should be interpreted as reflecting heterogeneous responses to the event, rather than as causal effects under a standard difference-in-differences design.

[9] One could additionally include time-varying demographic characteristics such as income and marital status. However, these variables exhibit limited within-individual variation conditional on individual fixed effects; thus, I exclude them from the baseline specification for parsimony. Their inclusion does not meaningfully affect the estimates.

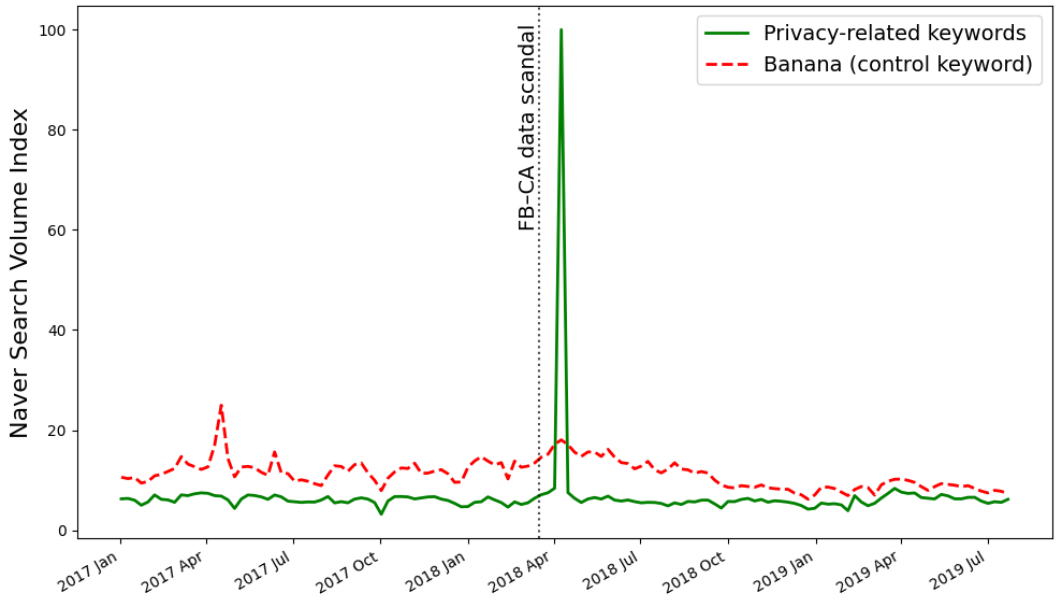

Figure 2: Search volume trends for privacy-related keywords on Naver, a Korean search engine

Notes: The vertical line represents March 17, 2018, when The Guardian published an interview with Christopher Wylie, a former Cambridge Analytica employee and whistleblower. Privacy-related keywords include: personal data breach, personal information, and privacy. Banana keywords (in Korean and English) were included to give a sense of the scale of the scandal's effect.

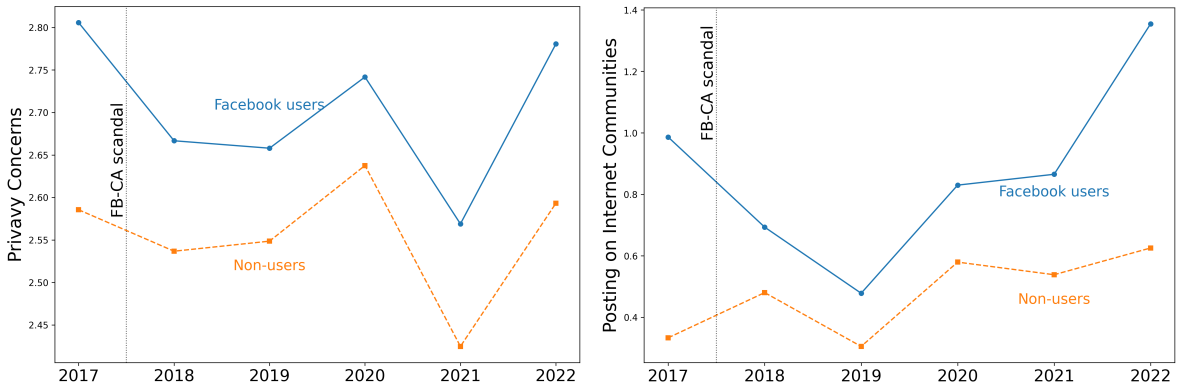

Figure 3: Privacy Concerns and Posting on Internet Communities Trends

Notes: Facebook users are panelists who were using Facebook as of June 2017, and non-users are those who were not using Facebook as of that date. The sample includes 1,966 Facebook users and 5,870 non-users, respectively. This classification was made before the major personal data breach (Facebook-Cambridge Analytica scandal) was disclosed in March 2018. The unit of Internet community posting is times per month. Population weights were incorporated when calculating averages across panelists within each group.

The idea behind this approach is that Facebook users in 2017 were more exposed to privacy-related shocks following the scandal disclosure than non-users in 2017. This expectation is natural, but the data show more nuanced patterns. Figure 3 shows that the Facebook user group did not exhibit increased privacy concerns between 2017 and 2018. Instead, partly due to a general trend, their privacy concerns decreased. However, Facebook users reduced their online posting significantly, which is consistent with the expected behavior when individuals react more sensitively to privacy risks.

These patterns suggest that the effect of the major data breach may operate through how *sensitive* individuals are to a given level of privacy concerns. Even when both Facebook users and non-users report similar levels of stated privacy concerns (e.g., both answer "I somewhat feel concerned about my privacy online"), Facebook users may be more likely to respond through behavioral adjustments, such as posting less online.

Therefore, results from this approach can be viewed as complementary to those from the lagged privacy concerns approach. The latter examines the potential consequences of the level of privacy concerns, while this approach captures differences in the intensity of behavioral responses to a given level of privacy concerns. Both effects are relevant for a comprehensive evaluation of the potential consequences of privacy concerns.

5 Results

5.1 Internet Community Engagement

Table 3 shows that both lagged privacy concerns and data breach experience are significantly and negatively associated with internet community engagement, including using, posting, sharing posts, and commenting activities. This provides suggestive evidence that people with greater privacy concerns may become less active in internet communities such as Reddit or interest-based LinkedIn groups.

Specifically, column (1) of Table 3 shows that a one–standard deviation (1 SD) increase in privacy concerns is associated with a 0.7 percentage points (hereafter, pp) lower probability of using any internet community. This effect is straightforward to compute because the standard deviation of privacy concerns is 0.99 (Table 2). Similarly, columns (2)–(4) show that a 1 SD increase in privacy concerns is associated with lower levels of posting, sharing (others' posts), and commenting activity by 0.06, 0.08, and 0.06 times per month, respectively.

Being a Facebook user when the Facebook–Cambridge Analytica scandal was disclosed is also associated with a 9.9 pp lower probability of using any internet community. The corresponding associations for specific activities are larger: posting, sharing (others' posts), and commenting are lower by 0.24, 0.15, and 0.15 times per month, respectively.

This result has two potential implications. First, privacy concerns may diminish the social value of internet communities, such as information dissemination and community building. For example, job seekers may find it more difficult to obtain helpful advice for

Table 3: Regression results of internet community engagement.

	Internet Community (IC) dummy and activities			
	(1)	(2)	(3)	(4)
Dep. Var.	$\mathbb{1}\{\text{UsesIC}\}$	Posting	Sharing (a post)	Commenting
Panel A: Fixed-Effects Specification				
$\beta^{PrivConc}$	-0.007**	-0.059**	-0.082***	-0.062
	(0.004)	(0.027)	(0.030)	(0.041)
Observations	35,987	35,987	27,416	35,987
R^2	0.515	0.461	0.489	0.496
Panel B: Difference-in-Differences–Style Specification				
FB user × Post-2018	-0.099***	-0.241**	-0.146	-0.148
	(0.015)	(0.104)	(0.094)	(0.145)
Observations	36,604	36,604	31,033	36,604
R^2	0.493	0.368	0.370	0.417
Indiv FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: Internet community refers to preference-based online communities, such as Naver Cafe (Korea), subreddits, or Facebook groups. The unit for activities is times per month. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

building suitable human capital for their careers, and board game enthusiasts may find it more challenging to locate local friends who share their interests. These activities depend crucially on how active internet community members are.

Second, from the perspective of internet community managers, heightened privacy concerns may have negative consequences for their revenue. This is because internet communities typically generate revenue from advertisements (e.g., Google ad banners), and this advertising revenue depends on the number of "eyeballs"—that is, active users in the communities. If users spend less time and engage less in internet communities, advertising revenues are likely to decline.

5.2 Online Content Contribution

Table 4: Regression results of online content contribution

	Online Content Contribution			
	(1)	(2)	(3)	(4)
Dep. Var.	Question posting	Answering	Voting	Recommendation
Panel A: Fixed-Effects Specification				
$\beta^{PrivConc}$	-0.069***	-0.036*	-0.046**	-0.005
	(0.021)	(0.021)	(0.023)	(0.024)
Observations	27,416	27,416	27,416	51,328
R^2	0.431	0.428	0.488	0.360
Panel B: Difference-in-Differences–Style Specification				
FB user × Post-2018	-0.061	-0.061	-0.304***	-0.254**
	(0.060)	(0.062)	(0.102)	(0.110)
Observations	31,033	31,033	31,033	36,604
R^2	0.318	0.300	0.364	0.351
Indiv FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: Question posting and answering refer to activities on knowledge-sharing platforms like Quora. The unit for activities is times per month. Voting refers to participating in online activities such as opinion polls (e.g., "Rome or London for travel destination?"). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 shows similar negative coefficients for online content contribution, specifically posting questions, answers, voting on polls, or making recommendations. Examples where such activities are widespread include Q&A platforms (e.g., Quora, Stack Exchange) and review platforms (e.g., Yelp, TripAdvisor).

Specifically, column (1) of Table 4 shows that a one–standard deviation increase in privacy concerns (1 SD = 0.99) is associated with 0.07 fewer questions posted per month. Columns (2) and (3) show similar patterns: a 1 SD increase in privacy concerns

is associated with 0.04 fewer answers and 0.05 fewer votes per month, respectively. In column (4), the coefficient is close to zero and statistically insignificant, suggesting no clear association between privacy concerns and recommendation activity.

In Panel B, being a Facebook user at the time of the Facebook–Cambridge Analytica scandal disclosure is associated with significantly lower levels of voting and recommendation activity, by 0.30 and 0.25 times per month, respectively. The estimated associations for question posting and answering are modest and statistically insignificant.

Notably, the effects associated with exposure to the Facebook–Cambridge Analytica scandal are more pronounced for voting and recommendations. This pattern is intuitive, as these activities are more common on Facebook (e.g., Facebook Poll features or local restaurant recommendations) than more explicit question-and-answer formats.

The implications of this result are similar to those for internet communities in the previous subsection. Heightened privacy concerns might disrupt the intrinsic value of online knowledge-sharing platforms by discouraging user activities. Additionally, these platforms may suffer from reduced advertising revenue due to decreased user engagement.

5.3 Email and Cloud Service

Table 5: Regression results of email and cloud service usage.

	(1)	(2)
Dep. Var.	$\mathbb{1}\{\text{Email}\}$	$\mathbb{1}\{\text{Cloud Service}\}$
Panel A: Fixed-Effects Specification		
$\beta^{PrivConc}$	-0.009***	-0.006*
	(0.004)	(0.004)
Observations	35,987	35,987
R^2	0.661	0.560
Panel B: Difference-in-Differences–Style Specification		
FB user × Post-2018	-0.064***	-0.021
	(0.009)	(0.013)
Observations	36,604	36,604
R^2	0.640	0.523
Indiv FE	✓	✓
Year FE	✓	✓

Notes: Dependent variables are dummies for email and cloud service usage at the individual-year level. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 5 shows small but significant negative associations between privacy concerns

and email and cloud service usage. This suggests the possibility that privacy concerns could reduce usage of services through which personal information could be leaked.

Specifically, column (1) of Table 5 shows that a one–standard deviation increase in privacy concerns (1 SD = 0.99) is associated with a 0.9 pp lower probability of using email services. Column (2) shows a similar pattern for cloud services: a 1 SD increase in privacy concerns is associated with a 0.6 pp lower probability of using cloud storage services.

In Panel B, being a Facebook user at the time of the Facebook–Cambridge Analytica scandal disclosure is associated with a 6.4 pp lower probability of using email services and a 2.1 pp lower probability of using cloud services. Although the magnitudes for cloud services are smaller, both estimates suggest that individuals exposed to the scandal reduced engagement with services that require storing or transmitting personal data.

This result may be surprising, since email is often perceived as an essential communication tool and was widely used during the data period (2017–2022). However, email can also be substituted for other services due to rapidly evolving information technologies. The email usage rate is around 70 percent according to KMPS data and external statistics.[10] In contrast, the penetration rate of messenger apps is much higher and continues growing, reaching 92 percent as of 2024. Due to high smartphone penetration rates, important social notifications such as government tax notices, fines, and credit card billing are all available through messenger apps in Korea.

Regarding privacy-related substitution between email (and cloud services) versus messenger apps (e.g., WhatsApp, KakaoTalk), one speculative but possible explanation is media framing. It is common to highlight public figures in data breach cases involving email/cloud services, such as "Jennifer Lawrence's iCloud hacked," while the size of an anonymous crowd is more emphasized in messaging app breach cases, such as "WhatsApp data leaked: 500 million user records for sale online." [11] This contrast may create a biased impression that email and cloud services are riskier options for privacy than messaging apps.

5.4 App Purchases and Smartphone Replacement

Table 6 shows negative associations between privacy concerns and mobile app purchases, as well as expected remaining time for replacing the current smartphone. This pattern is consistent with the possibility that heightened privacy concerns may reduce consumers' willingness to provide personal information required for app purchases, such as billing addresses or credit card details.

Specifically, columns (1)–(3) of Table 6 show that a one–standard deviation increase in privacy concerns (1 SD = 0.99) is associated with lower mobile app purchasing activity. A 1 SD increase in privacy concerns is associated with a 0.6 pp lower probability

[10] Statistics from Seoul (capital of Korea) reveal that email usage rates were 65-68% during the same period; see https://data.seoul.go.kr/dataList/10908/S/2/datasetView.do

[11] For examples, see https://www.bbc.com/news/newsbeat-29008876 about iCloud breach and https://cybernews.com/news/whatsapp-data-leak/ for WhatsApp case.

Table 6: Regression results of mobile app purchases and smartphone replacement time

	Mobile App Purchases			Smartphone
	(1)	(2)	(3)	(4)
Dep. Var.	$\mathbb{1}\{\text{Purchase}\}$	# of apps	Spending	Replacement time
Panel A: Fixed-Effects Specification				
$\beta^{PrivConc}$	-0.006***	-0.021***	-0.348***	0.473**
	(0.002)	(0.008)	(0.125)	(0.210)
Observations	35,987	35,987	35,987	16,815
R^2	0.408	0.358	0.353	0.491
Panel B: Difference-in-Differences–Style Specification				
FB user × Post-2018	-0.022***	-0.017	-0.495	1.179**
	(0.009)	(0.035)	(0.414)	(0.459)
Observations	36,604	36,604	36,604	23,986
R^2	0.373	0.293	0.302	0.416
Indiv FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: $\mathbb{1}\{\text{purchase}\}$, # of apps, and spending represent, respectively, a dummy for any paid app purchase, the number of paid apps purchased, and the amount spent on app purchases (unit: 1,000 KRW $\approx$ 1 USD) in the last 12 months. Replacement time refers to the expected remaining time for replacing the current smartphone (unit: months). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of making any paid app purchase (column 1), purchasing 0.02 fewer paid apps (column 2), and spending 348 KRW less on mobile apps (column 3). In contrast, column (4) shows that higher privacy concerns are associated with a longer expected smartphone replacement time: a 1 SD increase corresponds to a 0.47-month longer replacement horizon.

In Panel B, being a Facebook user when the Facebook–Cambridge Analytica scandal was disclosed is associated with a 2.2 pp lower probability of making any paid app purchase. The associations for the number of apps purchased and app spending are negative but statistically imprecise. By contrast, the association with smartphone replacement time is sizable: individuals exposed to the scandal report an expected replacement horizon that is 1.18 months longer. This pattern is consistent with the idea that privacy-related shocks may lead individuals to delay upgrading devices or purchasing new apps that require additional data permissions.

These patterns suggest three takeaways. First, they contrast with the prediction that privacy-sensitive consumers may spend more on paid apps, as free apps are often more privacy-invasive (Kesler, 2023; Cecere et al., 2022). While such consumers may have an incentive to purchase paid apps to reduce privacy exposure, an alternative response is to avoid app purchases altogether. The results in Table 6 suggest that the latter behavior may be more prevalent on the demand side.

Second, the decline in app purchases can also be interpreted as reflecting lower overall digital engagement, consistent with the reduced engagement in online communities and knowledge-sharing platforms documented earlier. One possible explanation is that privacy-sensitive consumers may be reluctant to share personal information with less well-known app developers, while transactions mediated by established platforms (e.g., Google or Apple) are perceived as involving lower privacy risk than sharing payment information directly with individual app providers.

Finally, the results also suggest that privacy concerns may be related to keeping the current smartphone for a longer period. This interpretation is consistent with broader evidence that discarded smartphones often contain recoverable personal information. In South Korea, several cases have involved the recovery of deleted data from used devices, raising public concern about personal-information leakage.[12] Survey evidence further indicates that privacy is a major factor in disposal decisions: Park (2023) report that concerns about personal-information leakage are the most common reason for not selling used smartphones. Similar issues have been documented in international forensic studies of secondhand devices (e.g., Roberts et al., 2023). Taken together, these findings provide suggestive support for a link between privacy concerns and smartphone replacement or disposal decisions.

5.5 Social Networking Sites

Table 7 shows that general privacy concerns are not strongly negatively associated with SNS usage, whereas exposure to the Facebook–Cambridge Analytica scandal is.

[12]For example, see https://imnews.imbc.com/replay/2021/nwdesk/article/6302495_34936.html (in Korean).

Table 7: Regression results of SNS usage

Dep. Var.	(1) SNS	(2) Facebook	(3) Twitter	(4) Instagram	(5) KStory	(6) NBand
Panel A: Fixed-Effects Specification						
$\beta^{PrivConc}$	0.003	0.004	-0.002	0.007*	-0.003	-0.001
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	35,987	35,987	35,987	35,987	35,987	35,987
R^2	0.659	0.585	0.484	0.623	0.509	0.503
Panel B: Difference-in-Differences–Style Specification						
FB user × Post-2018	-0.325***	-0.630***	-0.101***	0.051***	-0.038**	-0.057***
	(0.011)	(0.012)	(0.013)	(0.014)	(0.015)	(0.013)
Observations	36,604	36,604	36,604	36,604	36,604	36,604
R^2	0.592	0.586	0.451	0.551	0.442	0.431
Indiv FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes: Dependent variables are dummies for using each SNS at the individual-year level. SNS refers to using any social networking sites. KStory and NBand refer to KakaoStory and NaverBand, popular Korean social platforms. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Because Facebook is one of the most widely used SNS platforms, it is plausible that a shock from the scandal is more strongly associated with SNS usage (Panel B) than general privacy concerns are (Panel A).

Specifically, Panel A of Table 7 shows that privacy concerns are not strongly associated with overall SNS usage or with the use of specific platforms such as Facebook, Twitter, KakaoStory, or NaverBand. The coefficients are small in magnitude and statistically insignificant across most platforms. The one exception is Instagram: a one–standard deviation increase in privacy concerns (1 SD = 0.99) is associated with a 0.7 pp higher probability of using Instagram (column 4), although this association is modest in size.

By contrast, the difference-in-differences–style analysis in Panel B shows pronounced changes in platform usage following the Facebook–Cambridge Analytica scandal. Being a Facebook user at the time of the scandal disclosure is associated with substantially lower probabilities of using SNSs in general (32.5 pp), Facebook specifically (63.0 pp), Twitter (10.1 pp), KakaoStory (3.8 pp), and NaverBand (5.7 pp). Instagram is the only platform that moves in the opposite direction: the estimated association corresponds to a 5.1 pp higher probability of use among treated individuals. This pattern suggests a shift away from Facebook and other platforms toward Instagram, although the absolute magnitudes differ—the positive spillover to Instagram is about 13 times smaller.

This heterogeneity could be explained by the content similarities between SNS platforms, as displayed in Figure 4. All SNS platforms that experienced decreased usage,

including KakaoStory and Naver Band, are primarily text-focused. The mainstream content on such platforms consists of text, such as daily diaries or travelogues. Therefore, consumers may consider these text-based platforms more similar to each other than to Instagram, where photos and images dominate. For example, a consumer may feel unsafe uploading detailed events of their daily life on Facebook and switch to Instagram, where less privacy-revealing snapshots (e.g., a taco for dinner) are mainstream.

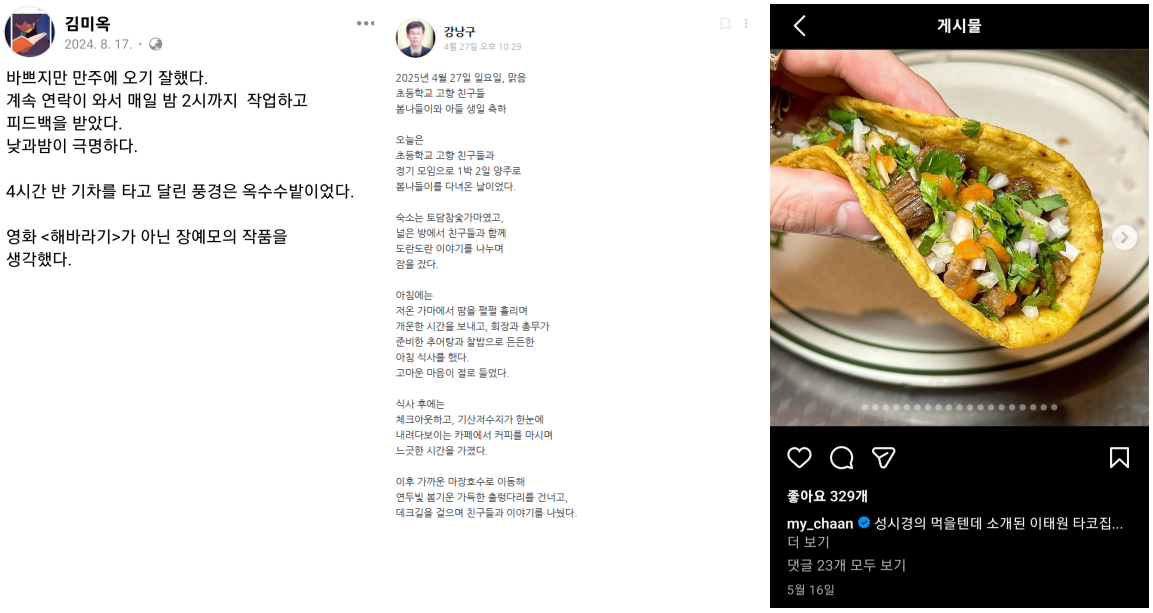

Figure 4: Typical posts showing content differences. Facebook and KakaoStory (left, center) feature text-heavy content, while Instagram (right) emphasizes visual content.

While speculative, the estimates in Panel B of Table 7 suggest a potential divergence between privately and socially optimal levels of privacy protection. A multi-product firm such as Meta (Facebook) may have incentives to provide less privacy protection than would be socially optimal for two reasons. First, it does not internalize negative spillovers imposed on competing firms. Second, even modest positive spillovers to its other service, Instagram, may partially offset the firm's losses from privacy-related events. Although the data do not allow for direct investigation of these mechanisms, the relationship among multi-product firms, market structure, and incentives for privacy protection remains an important area for future research.

5.6 Robustness Checks

To further validate findings that rely on the privacy concerns measure, I conduct a series of robustness checks. First, I use an alternative version of the measure of privacy concerns constructed from more explicitly privacy-related questions. Second, I exclude panelists under age 10. Third, I focus on dependent variables that are less likely to be affected by privacy concerns and run placebo regressions. The results are presented in Appendix A.

Alternative Measure of Privacy Concerns Among the eight privacy-related questions mentioned in Section 2.2, questions (1) and (6) more directly measure privacy concerns.[13] I construct an alternative measure of privacy concerns using only these two questions.

Using this alternative measure, I replicate Tables 3, 4, 5, and 6. For comparison, I present results from this alternative specification alongside the original estimates. The overall results remain qualitatively similar. Appendix A presents the results.

Dropping Panelists with Age Under 10 The main dataset includes panelists under age 10, who may not have an accurate understanding of privacy. As another robustness check, I replicate Tables 3, 4, 5, and 6 after excluding these observations. The results remain qualitatively similar, as panelists under age 10 account for only about 1.5% of observations. The result tables are presented in Appendix A.

Placebo Tests To address concerns about spurious correlation, I use two dependent variables: a newspaper subscription dummy and the frequency of reading posts on internet communities. Both dependent variables are less likely to affect or be affected by privacy concerns, in contrast to previous dependent variables such as posting frequency on internet communities.

Regression results using the same fixed-effects panel regression and difference-in-differences–style analysis are presented in Table A4. The results show that the estimates are not statistically significant across all specifications.

6 Discussion and Concluding Remarks

6.1 Potential Welfare Impacts of Data Protection Laws

To summarize, the empirical analyses show that heightened privacy concerns have significant negative associations with a wide range of digital activities: engagement in internet communities, online content contribution (e.g., answering questions on Quora), email usage, cloud service adoption, mobile app spending, and smartphone replacement.

As mentioned in 1.1, an implication of these empirical findings is the substantial social costs of privacy concerns and the potential benefits of effective data protection laws, such as the GDPR and CCPA. Key values of the internet—community building and knowledge sharing—benefit billions of users worldwide and hinge on active user engagement. For example, Reddit, with more than a billion monthly active users as of 2023, enables users to find travel destinations, enjoy curated content, and share investment opinions. Quora boasts over 400 million monthly visitors with content spanning 300,000 topics across 24 languages and helps its users access knowledge, from startup experiences to legal misconceptions. User engagement is the fundamental driver of these platforms' success and their social benefits.

[13] These questions are: "I worry that someone I don't know might obtain my personal information from my online activity" and "In general, I worry about my privacy when I use the internet."

Given this sizable social value, if data protection laws successfully mitigate privacy concerns and encourage online engagement, the welfare benefits could be immense. However, evidence on whether laws like GDPR actually reduce privacy concerns remains sparse. Johnson (2024, sec. 4.1) summarizes GDPR's impact on consumers and reports that existing research finds no clear improvement in perceived control over personal data (Presthus and Sørum, 2021). Therefore, making personal data protection laws more effective in improving consumer perceptions requires further investigation.

6.2 Limitations

These results should be interpreted with caution and not as clean causal estimates. Although the analysis addresses prominent confounding factors through individual fixed effects with lagged measures of privacy concerns, and by exploiting the plausibly exogenous event of the Facebook–Cambridge Analytica scandal disclosure, important limitations remain. Fixed-effects regressions alone do not guarantee a causal interpretation. Moreover, this analysis should be understood as a difference-in-differences–style approach that exploits differential exposure to an unexpected event, rather than a canonical difference-in-differences design. Because only a single pre-scandal period is available, formal tests of the parallel trends assumption are difficult to conduct. Identifying more convincing sources of exogenous variation in privacy concerns is therefore an important but challenging direction for future research.

6.3 Conclusion

This paper uses national-level panel survey data to study the relationship between privacy concerns and digital engagement and consumption. I take two approaches: fixed-effects panel regressions using directly measured privacy concerns, and a difference-in-differences–style analysis that exploits differential exposure to the Facebook–Cambridge Analytica scandal.

Both higher levels of self-reported privacy concerns and greater exposure to the Facebook–Cambridge Analytica scandal are associated with reductions in digital activities across multiple dimensions, including engagement in internet communities (e.g., Reddit), online content contribution (e.g., Wikipedia), and digital consumption, such as paid app purchases and smartphone replacement. Moreover, scandal exposure is associated with reduced SNS usage more broadly, affecting not only Facebook but also competing platforms, while Instagram may have benefited from substitution effects.

Taken together, these findings suggest that privacy-related factors are closely linked to digital engagement and consumption decisions. While the empirical strategies do not deliver clean causal estimates, the results point to the potential for welfare gains from personal data protection policies that mitigate privacy-related risks and support sustained internet user engagement.

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar (2022) "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 14 (4), 218–256.

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein (2015) "Privacy and human behavior in the age of information," *Science*, 347 (6221), 509–514.

Acquisti, Alessandro, Leslie K John, and George Loewenstein (2013) "What is privacy worth?" *The Journal of Legal Studies*, 42 (2), 249–274.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016) "The economics of privacy," *Journal of Economic Literature*, 54 (2), 442–92.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz (2023) "The effect of privacy regulation on the data industry: empirical evidence from GDPR," *The RAND Journal of Economics*, 54 (4), 695–730.

Athey, Susan, Christian Catalini, and Catherine Tucker (2017) "The digital privacy paradox: Small money, small costs, small talk," *NBER Working Paper No. 23488*.

Barnes, Susan B (2006) "A privacy paradox: Social networking in the United States," *First Monday*.

Beresford, Alastair R, Dorothea Kübler, and Sören Preibusch (2012) "Unwillingness to pay for privacy: A field experiment," *Economics letters*, 117 (1), 25–27.

Cecere, Grazia, Vincent Lefrere, and Fabrice Le Guel (2022) "Third parties in the app market and economics of privacy," *Economics Bulletin*, 42 (2), 1040–1049.

Chen, Long, Yadong Huang, Shumiao Ouyang, and Wei Xiong (2021) "The Data Privacy Paradox and Digital Demand," Working Paper 28854, National Bureau of Economic Research, Cambridge, MA, 10.3386/w28854.

Goldberg, Samuel G, Garrett A Johnson, and Scott K Shriver (2024) "Regulating privacy online: An economic evaluation of the GDPR," *American Economic Journal: Economic Policy*, 16 (1), 325–358.

Goldfarb, Avi and Catherine Tucker (2012) "Shifts in privacy concerns," *American Economic Review P&P*, 102 (3), 349–353.

Ichihashi, Shota (2023) "Dynamic Privacy Choices," *American Economic Journal: Microeconomics*, 15 (2), 1–40.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman (2021) "The short-run effects of the General Data Protection Regulation on Technology Venture Investment," *Marketing Science*, 40 (4), 661–684.

Johnson, Garrett A. (2024) "Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond," in Acquisti, Alessandro, Curtis Taylor, and Liad Wagman eds. *The Economics of Privacy*, 97–126: University of Chicago Press.

Johnson, Garrett A, Scott K Shriver, and Samuel G Goldberg (2023) "Privacy and market concentration: intended and unintended consequences of the GDPR," *Management Science*, 69 (10), 5695–5721.

Kesler, Reinhold (2023) "The Impact of Apple's App Tracking Transparency on App Monetization," Working Paper 4090786, SSRN, 10.2139/ssrn.4090786, Available at SSRN.

Kummer, Michael and Patrick Schulte (2019) "When private information settles the bill: Money and privacy in Google's market for smartphone applications," *Management Science*, 65 (8), 3470–3494.

Lee, Stephanie (2018) "Quantifying the Consumer Surplus from Smartphones," *SSRN working paper No. 3270047*.

Lee, Yi-Shan and Roberto A Weber (2025) "Revealed privacy preferences: Are privacy choices rational?" *Management Science*, 71 (3), 2657–2677.

Liao, Guocheng, Yu Su, Juba Ziani, Adam Wierman, and Jianwei Huang (2024) "The Privacy Paradox and Optimal Bias–Variance Trade-offs in Data Acquisition," *Mathematics of Operations Research*, 49 (4), 2749–2767.

Lin, Tesary (2022) "Valuing intrinsic and instrumental preferences for privacy," *Marketing Science*, 41 (4), 663–681.

Miklós-Thal, Jeanine, Avi Goldfarb, Avery Haviv, and Catherine Tucker (2024) "Frontiers: digital hermits," *Marketing Science*, 43 (4), 697–708.

Park, Jin-Hwan (2023) "Estimating the Size of the Domestic Secondhand Smartphone Market and Its Implications," *KISDI Perspectives*, 2023 December No. 1, 1–9, https://www.kisdi.re.kr/report/view.do?key=m2102058837181&masterId=4334696&arrMasterId=4334696&artId=1168676, Published December 6, 2023 (in Korean).

Pew Research (2019) "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information,"Technical report, Pew Research Center, https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/11/Pew-Research-Center_PI_2019.11.15_Privacy_FINAL.pdf, Accessed June 3, 2025.

——— (2022) "Social Media Seen as Mostly Good for Democracy Across Many Nations, But U.S. is a Major Outlier,"Technical report, Pew Research Center, https://www.pewresearch.org/global/2022/12/06/social-media-seen-as-mostly-good-for-democracy-across-many-nations-but-u-s-is-a-major-outlier/, Accessed June 3, 2025.

——— (2023) "Majority of Americans say TikTok is a threat to national security," July, https://www.pewresearch.org/short-reads/2023/07/10/majority-of-americans-say-tiktok-is-a-threat-to-national-security/, Survey conducted May 15-21, 2023 among 5,101 U.S. adults.

Presthus, Wanda and Hanne Sørum (2021) "A three-year study of the GDPR and the consumer," in *Proceedings at 14th IADIS International Conference Information Systems*, 3–5.

Prince, Jeffrey T and Scott Wallsten (2022) "How much is privacy worth around the world and across platforms?" *Journal of Economics & Management Strategy*, 31 (4), 841–861.

Prince, Jeffrey and Scott Wallsten (2025) "Do People Around the World Care Where Their Data Are Stored?" *Information Economics and Policy*, 101132.

Roberts, Richard, Julio Poveda, Raley Roberts, and Dave Levin (2023) "Blue Is the New Black (Market): Privacy Leaks and Re-Victimization from Police-Auctioned Cellphones," in *2023 IEEE Symposium on Security and Privacy (SP)*, 3332–3336, 10.1109/SP46215.2023.10179348.

Stutzman, Frederic D, Ralph Gross, and Alessandro Acquisti (2013) "Silent listeners: The evolution of privacy and disclosure on Facebook," *Journal of privacy and confidentiality*, 4 (2), 2.

Tsai, Janice Y, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti (2011) "The effect of online privacy information on purchasing behavior: An experimental study," *Information systems research*, 22 (2), 254–268.

Tucker, Catherine E (2014) "Social networks, personalized advertising, and privacy controls," *Journal of marketing research*, 51 (5), 546–562.

A Additional Tables and Figures

A.1 Alternative Measure of Privacy Concerns and Excluding Responses from Children

In this subsection, I show that the main results remain almost unchanged when I use an alternative measure of privacy concerns based on questions (1) and (6) and when I drop observations with ages under 10.

For this purpose, I replicate Tables 3, 4, 5, and 6 for each robustness check. Tables A1, A2, and A3 display the results. For comparison, I also repeat the baseline estimates that appeared in the main text.

Table A1: Robustness Check Results for Internet Community Engagements

	Internet Community (IC) dummy and activities			
	(1)	(2)	(3)	(4)
Dep. Var.	$\mathbb{1}\{\text{UsesIC}\}$	Posting	Sharing (a post)	Commenting
(Baseline)				
$\beta^{PrivConc}$	-0.007^{**}	-0.059^{**}	-0.082^{***}	-0.062
	(0.004)	(0.027)	(0.030)	(0.041)
(Alternative Privacy Concerns Measure)				
$\beta^{PrivConc}$	-0.010^{***}	-0.049^{*}	-0.099^{***}	-0.078^{**}
	(0.003)	(0.025)	(0.028)	(0.037)
(Excluding Child Panelists)				
$\beta^{PrivConc}$	-0.008^{**}	-0.059^{**}	-0.083^{***}	-0.063
	(0.004)	(0.027)	(0.031)	(0.041)

Notes: The unit of activities is times per month. Three sets of results show (1) baseline results from the main text, (2) results from an alternative measure of privacy concerns using more clearly privacy-related questions 1 and 6, and (3) results when dropping child panelists (under 10 years old). Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A2: Robustness Check Results for Online Content Contribution

	Online Content Contribution			
	(1)	(2)	(3)	(4)
Dep. Var.	Question posting	Answering	Voting	Recommendation
(Baseline)				
$\beta^{PrivConc}$	-0.069^{***}	-0.036^{*}	-0.046^{**}	-0.005
	(0.021)	(0.021)	(0.023)	(0.024)
(Alternative Privacy Concerns Measure)				
$\beta^{PrivConc}$	-0.075^{***}	-0.042^{*}	-0.036	0.009
	(0.022)	(0.023)	(0.024)	(0.025)
(Excluding Child Panelists)				
$\beta^{PrivConc}$	-0.071^{***}	-0.036^{*}	-0.047^{**}	-0.005
	(0.021)	(0.021)	(0.024)	(0.024)

Notes: The unit of activities is times per month. Three sets of results show (1) baseline results from the main text, (2) results from an alternative measure of privacy concerns, and (3) results when dropping child panelists (under 10 years old). Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A3: Robustness Check Results for Digital Service Usage and Consumption

	Digital Service Usage		Mobile App Purchases			Smartphone
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	$\mathbb{1}\{\text{Email}\}$	$\mathbb{1}\{\text{CloudSvc}\}$	$\mathbb{1}\{\text{Purchase}\}$	# of apps	Spending	Replacement time
(Baseline)						
$\beta^{PrivConc}$	-0.009^{***}	-0.006^{*}	-0.006^{***}	-0.021^{***}	-0.348^{***}	0.473^{**}
	(0.004)	(0.004)	(0.002)	(0.008)	(0.125)	(0.210)
(Alternative Privacy Concerns Measure)						
$\beta^{PrivConc}$	-0.009^{***}	-0.003	-0.004^{*}	-0.013^{**}	-0.343^{***}	0.450^{**}
	(0.003)	(0.003)	(0.002)	(0.007)	(0.115)	(0.194)
(Excluding Child Panelists)						
$\beta^{PrivConc}$	-0.010^{***}	-0.007^{*}	-0.006^{***}	-0.022^{***}	-0.357^{***}	0.469^{**}
	(0.003)	(0.004)	(0.002)	(0.008)	(0.126)	(0.210)

Notes: # of apps and spending represent, respectively, the number of paid apps purchased and the amount spent on app purchases (unit: 1,000 KRW $\approx$ 1 USD) in the last 12 months. Replacement time refers to the expected remaining time for replacing the current smartphone (unit: months). Three sets of results show (1) baseline results from the main text, (2) results from an alternative measure of privacy concerns, and (3) results when dropping child panelists (under 10 years old). Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

A.2 Placebo Tests

For placebo test purposes, I run the same regressions but using dependent variables that are less likely to be affected by privacy concerns: frequency of reading posts in internet communities and newspaper subscription dummy. Table A4 shows that these variables are not significantly correlated with privacy concerns, as expected.

Table A4: Placebo tests using dependent variables less likely to be affected by privacy concerns.

	(1)	(2)
Dep. Var.	$\mathbb{1}\{$Newspaper Subscription$\}$	Reading Posts
Panel A: Fixed-Effects Specification		
$\beta^{PrivConc}$	-0.005	0.038
	(0.004)	(0.079)
Observations	35,987	27,416
R^2	0.586	0.609
Panel B: Difference-in-Differences–Style Specification		
FB user × Post-2018	-0.002	-0.055
	(0.016)	(0.202)
Observations	36,604	31,033
R^2	0.524	0.537
Indiv FE	✓	✓
Year FE	✓	✓

Notes: Dependent variables are a dummy for newspaper subscription and the number of times per month a panelist reads posts in internet communities such as Reddit. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

A.3 Dropped Panelists

Figure A1 shows the age profiles of dropped panelists who responded that they do not engage in online activities, likely reflecting age-related patterns, and are thus not relevant for this privacy concerns study.

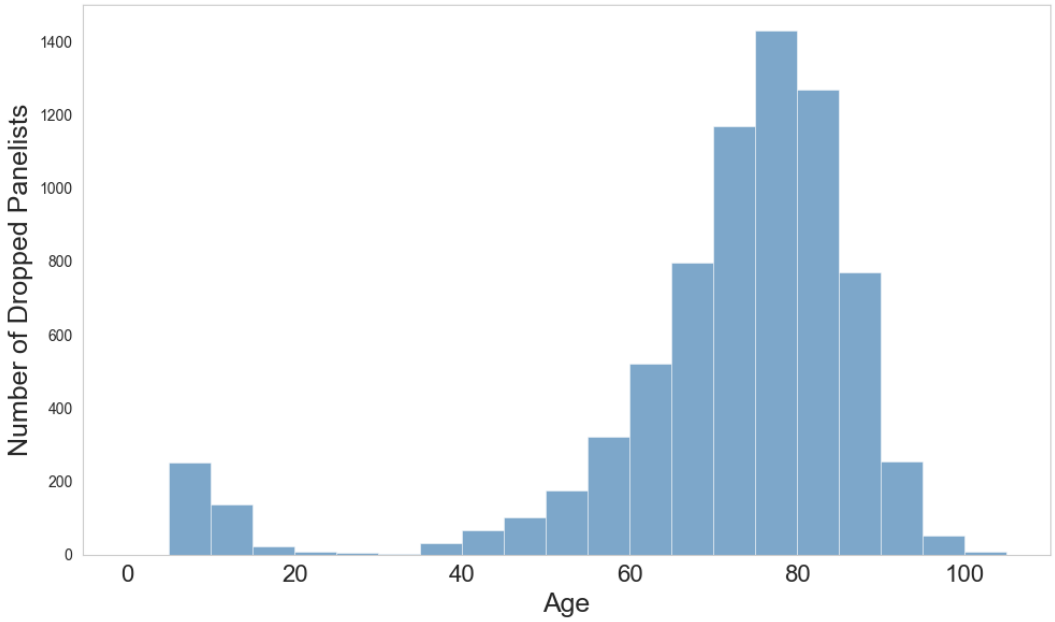

Figure A1: Age distribution of dropped panelists

Notes: This figure shows the age distribution of panelists dropped from the analysis for selecting "I do not engage in online activities" to privacy-related questions. For dropped samples, $N^{\text{panelists}} = 3,217$ and $N^{\text{obs}} = 7,392$. For comparison, the main dataset contains $N^{\text{panelists}} = 12,763$ and $N^{\text{obs}} = 52,720$. The mean age is 70.39 and 44.74, respectively.

B Variable Coding Details

The KMPS survey includes questions that ask respondents about their frequency of engagement in various activities. For each activity, respondents select from options such as "rarely," "1-3 times per week," and "almost everyday." These categorical responses are converted to numeric values (times per month) for regression analysis. Table A5 presents the mapping between response categories and their corresponding numeric values.

Table A5: Mapping between categorical responses and the corresponding numeric values.

Categorical Response	**Numeric value (times per month)**
rarely	0
once every three months	1/3
1-3 times per month	2
1-3 times per week	8
4-6 times per week	20
almost every day	30