Tech's Touch on Mental Health:
Revolutionizing Mental Health Care with Telemedicine and Online Empowerment

BY

YI YANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY

ROBINSON COLLEGE OF BUSINESS

2025

ACCEPTANCE

This dissertation was prepared under the direction of the *Yi Yang* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Dr. Arun Rai, Dr. Aaron Baird, Dr. Yu-kai Lin, Dr. Elena Karahanna

Tech’s Touch on Mental Health:

Revolutionizing Mental Health Care with Telemedicine and Online Empowerment

BY

YI YANG

15TH JULY 2025

Committee Chair: *Dr. Arun Rai and Dr. Aaron Baird*

Major Academic Unit: *Center for Digital Innovation*

ABSTRACT

This dissertation explores how digital technologies help mental health outcomes at both community and individual levels, offering a multi-level perspective on digital mental health support. The first essay investigates the impact of hospital telehealth adoption on county-level mental health outcomes across the United States. Guided by social capital theory and employing causal forest modeling on a panel dataset of over 2,800 counties, the study reveals that telehealth's effectiveness varies significantly by community context. Counties with limited broadband access but high economic or cultural capital see the greatest improvements, highlighting the need for complementary community resources to realize the full benefits of digital infrastructure. The second essay examines how anonymity affects user engagement in online mental health discussions. Leveraging a natural policy change on Quora and grounded in signaling theory, it demonstrates that anonymity leads to substantial increases in user participation, content specificity, and emotional expression—particularly among users with fewer disclosed credentials—underscoring the role of platform design in empowering marginalized voices. Together, these essays demonstrate how digital technologies can improve mental health outcome and foster inclusive mental health discourse.

ACKNOWLEDGEMENTS

Completing this dissertation has been one of the most challenging and transformative experiences of my life. I am deeply grateful to those who have supported, guided, and encouraged me throughout this journey.

First and foremost, I would like to express my heartfelt gratitude to my advisor, Dr. Arun Rai, for your unwavering support, thoughtful guidance, and invaluable mentorship. Your belief in my work and consistent encouragement played a central role in shaping both my research and my academic identity. I could not have asked for a better advisor. I am also thankful to my dissertation committee: Dr. Aaron Baird (co-chair), Dr. Yu-Kai Lin, and Dr. Elena Karahana for your insightful feedback, encouragement, and commitment to my development as a scholar. Your advice challenged me to think more deeply and critically, for which I am sincerely appreciative. To my cohort mates, mentors, and friends in the Center for Digital Innovation at Georgia State University, thank you for the intellectual camaraderie, collaboration, and kindness. You made this experience not only productive but also meaningful and memorable. To my family, especially my mom Weiran Liu and dad Tianliang Yang, thank you for your unconditional love and support. Your faith in me provided strength during the moments when I doubted myself.

My PhD journey began at the onset of the COVID-19 pandemic, a time of global uncertainty that unexpectedly helped me focus on healthcare and inspired my interest in exploring digital interventions in mental health. It concludes during a period marked by political turmoil and social division, where academic institutions face unexpected funding cuts and increasing uncertainty. Navigating this journey through such extraordinary times has taught me resilience, adaptability, and the importance of staying grounded in the pursuit of meaningful work. This dissertation is not only a product of research and writing, but also a testament to perseverance, growth, and the generosity of those who stood by me. I am sincerely thankful.

CONTENTS

Chapter 1: Introduction

Mental health represents one of the most pressing public health challenges of our time, with traditional care systems struggling to meet growing demand. The COVID-19 pandemic has intensified these challenges, contributing to unprecedented increases in anxiety, depression, and other mental health conditions while simultaneously highlighting persistent barriers to accessing traditional mental health services. As digital technologies rapidly transform healthcare delivery, understanding their mental health implications has become critically important for both policymakers and health care providers.

In this dissertation, we investigate how digital platforms affect mental health outcomes through two complementary lenses: community-level impacts of formal telehealth adoption and individual-level effects of anonymity in online mental health communities. We use these essays to examine the multilevel ecosystem of digital mental health support. While formal telehealth services represent institutionalized care delivery that extends professional treatment beyond traditional settings, online mental health communities offer peer-driven support networks that provide accessible, immediate assistance. Understanding both formal and informal digital mental health channels is crucial for developing comprehensive mental health policies and interventions in the digital age.

1.1 Theoretical and Methodological Contributions

Despite growing optimism about the potential of digital technologies, prior research has often assumed that access to infrastructure or platform availability alone is sufficient to improve mental health outcomes (Barnett et al., 2021; Torous et al., 2020). However, this assumption overlooks critical nuances such as community capacity, social inequality, and user-level barriers like stigma and identity signaling. Existing studies tend to focus on either system-level interventions or individual behavior, but few investigate how digital tools function within a multilevel mental health ecosystem. This dissertation addresses this gap by examining both institutionalized care through telehealth and informal support systems in online mental health communities. By doing so, it challenges prevailing assumptions of technological determinism and demonstrates that context—both structural and psychological—plays a pivotal role in shaping mental health outcomes. The dissertation makes several key contributions to the digital health literature. Methodologically, it applies causal forest approach to mental health research and leverages natural experiments in online platform settings to address endogeneity concerns. Theoretically, it extends social capital theory to digital health contexts and provides new empirical insights into signaling effects in mental health settings. Collectively, these contributions offer a more nuanced understanding of how digital technologies can support mental health at community and individual level.

1.2 Essay 1: Telehealth as Digital Capital in Community Mental Health

In the first essay, we investigate how hospital adoption of telehealth services affects county-level mental health outcomes. Telehealth emerges as a form of digital capital that communities can harness to democratize access to mental health care, offering virtual platforms for interaction with health care providers while eliminating geographical barriers and reducing stigma.

Drawing from social capital theory, we argue that telehealth's effectiveness depends on the interplay of various forms of capital—digital, social, economic, and cultural—within communities. Using a novel panel dataset covering over 2800 U.S. counties and applying causal forest modeling, we reveal substantial variation in telehealth's impact across different community contexts.

The findings demonstrate that counties with low broadband access but high economic or cultural capital experience the most significant improvements in mental health outcomes, including reductions in mentally unhealthy days and increases in service utilization. Conversely, counties with high broadband access but low social capital show minimal effects. These results suggest that technological infrastructure without supportive community resources may be insufficient for effective telehealth implementation, challenging assumptions that digital infrastructure alone determines success.

1.3 Essay 2: Identity Cues and Anonymity in Online Mental Health Discussion

In the second essay, we explore how transitioning from identity disclosure to anonymity affects user engagement in online mental health discussions. Leveraging a natural experiment on Quora where the platform rescinded its real-name policy, we investigate how anonymity affects both quantity and quality of user-generated mental health content.

Grounding the study in signaling theory and social stigma frameworks, we hypothesize that anonymity reduces participation barriers and increases authentic self-disclosure, especially for stigmatized mental health topics. Using a weekly panel dataset of over 2,700 user-week observations and employing LLM-based text analysis, we employ difference-in-differences methodology to identify causal effects.

The findings demonstrate that anonymity substantially increases user engagement, with response rates increasing by approximately 23%, content specificity improving by 18%, and emotional expression (anger emotion) intensifying by 31%. Importantly, users with fewer disclosed credentials exhibit the greatest behavioral shifts under anonymity, suggesting that anonymity disproportionately empowers vulnerable users who may face greater stigma in identified settings.

1.4 Synthesis and Implications

Together, these studies underscore technology's dual facets as both a health care delivery tool and a social medium. The first essay illuminates how telehealth can transcend traditional barriers to care while underscoring the need for nuanced understanding of community dynamics to maximize benefits. The second essay demonstrates how platform design features can influence the quality of support and engagement in mental health discussions.

This dissertation highlights technology's integral role in reshaping mental health care, demonstrating how digital solutions can augment traditional health care models and foster more inclusive, supportive environments for discussing sensitive health topics. The synergistic relationship between telehealth and online mental health forums creates a comprehensive ecosystem that addresses diverse needs in the mental health space.

The findings have direct implications for health care policymakers and platform designers, as summarized in Table 1. Essay 1 informs evidence-based telehealth expansion strategies that consider community context rather than pursuing uniform technological solutions. Essay 2 provides guidance for online platform policies that can enhance mental health support while protecting user privacy and promoting authentic discourse.

As the world continues to grapple with mental health complexities, these insights provide a road map for leveraging technology to enhance care and community well-being, contributing significantly to our understanding of how digital innovations can address pressing challenges in community mental health.

Following the introduction, Chapter 2 presents the first essay, which examines the impact of telehealth adoption on community mental health outcomes. Chapter 3 contains the second essay, focusing on the effects of anonymity in online mental health communities. Chapter 4 concludes the dissertation, summarizing key findings and implications.

Table 1

Summary of the Essays

Essay	Theoretical Lens	Research Question	Data Sources	Methodology	Findings
1	Social Capital Theory	How does county-level telehealth adoption affect community mental health outcomes, and how do contextual factors moderate these effects?	SafeGraph location data, American Community Survey, American Hospital Association Annual Survey	Quasi natural experiment + Causal Forests Approach	Telehealth adoption has context-dependent effects. Greatest benefits occur in areas with low broadband access but high economic or cultural capital; under-resourced counties benefit in suicide reduction.
2	Signaling Theory	How does a shift from real name to anonymous posting affect user engagement, content specificity, and emotional expression in online mental health discussions?	Mental health Discussion Data on Quora	Quasi natural experiment + LLM-based text analysis	Anonymous posting increases answer rates, depth, and emotional expressiveness, especially among users lacking formal credentials or social authority.

Chapter 2: Uncovering Average and Heterogeneous Treatment Effects of Digital and Social Capital on Community Mental Health Outcomes

2.1 Introduction

Mental health issues are becoming more prevalent in society. A recent survey indicated 19.86% of adults have experienced a mental illness, equivalent to nearly 50 million Americans.[1] Yet, 94% of individuals with mental health issues did not receive any treatment (Reinert et al., 2021). Further, vulnerable populations are the most affected, which results in significant mental health disparities (Diaz et al., 2021). Overall, unwell mental health status has adverse impacts on people's personal lives by potentially resulting in social isolation, friction with family and friends, and adverse economic impact and occupational stress due to loss of wages (Pirkkalainen et al., 2019). As such, the magnitude and prevalence of mental health problems are not only an individual problem but also affect communities (Reinert et al., 2021). This distinction is important and is a key motivation for focusing on communities in this study.

While individual mental health refers to the psychological well-being of a single person, community mental health refers to the overall mental well-being of a defined group of people or population. Individual mental health involves the person's ability to manage stress, be productive (Bassett & Lloyd, 2001), maintain relationships (Pilgrim et al., 2009), and make meaningful contributions (Bond & Bunce, 2003). A variety of factors influences individual mental health, including genetics (Cloninger, 1999), personal experiences (Girdhar et al., 2020; Zarse et al., 2019), physical health (Ohrnberger et al., 2017; Prince et al., 2007), socioeconomic status (Compton & Shim, 2015; Yu & Williams, 1999), and environmental factors (Cianconi et al., 2020). However, community mental health well-being is not merely an aggregation of individual well-being. It

[1] https://mhanational.org/issues/2022/mental-health-america-adult-data

requires looking at the prevalence of mental health conditions in a community, alongside understanding the access to mental health services, social and cultural attitudes toward mental health, and communal support systems. Despite its importance, research on community-level mental health outcomes remains limited. Most existing studies prioritize individual-level mental health predictors and outcomes, often overlooking the collective dimensions of mental health such as social capital, structural inequities, or systemic access barriers. This narrow focus constrains our ability to develop policies and interventions that address mental health as a shared, societal issue. As mental health crises increasingly manifest across populations, it becomes crucial to expand research efforts to capture the broader community context in which communities experience, support, or neglect mental health.

Telehealth, which enables health care providers to provide virtual services for patients, is a potential solution to broaden access and mitigate disparities in access to care, specifically in the realm of mental health care. A their systematic review, Bashshur et al. (2016) highlighted the positive impact of telehealth interventions in increasing access to mental health services for underserved populations. Additionally, Hubley et al. (2016) demonstrated how telehealth programs can effectively address mental health disparities in rural and remote areas, where traditional in-person services may be limited. Furthermore, Hilty et al. (2013) found that telepsychiatry services have been successful in reducing barriers to mental health care and improving patient outcomes.

However, most of the research so far has focused on either the determinants of telehealth adoption at the hospital level (Adler-Milstein et al., 2014) or the impacts of telehealth on individuals' physical health (Chellaiyan et al., 2019) and mental health (Hubley et al., 2016). We have learned a great deal about the adoption and effective use of telehealth by hospitals and impacts of telehealth on individuals (Adler-Milstein et al., 2014; Angrist & Pischke, 2008; Baird et al., 2022). However, few researchers have investigated the impacts of the telehealth adoption on community health, including mental health. In addition, it is crucial to understand how telehealth as a potential solution for mitigating mental health problems in communities intertwines with socioeconomic determinants that can influence the effectiveness of telehealth as a solution. Such an understanding is essential to develop context-tailored solutions based on needs and pragmatics for individual communities and avoid the risk of imposing one solution without regard to key contextual considerations.

From a theoretical perspective, we draw on Bourdieu's (1977) theory of forms of capital of a collective to understand how telehealth adoption by hospitals in a community and other forms of capital in the community affect the mental health of citizens in the community. Bourdieu's theory explains why understanding the welfare of communities requires going beyond economic capital and requires consideration of other forms of capital, particularly social, cultural, and economic capital. Adding to these forms of capital, the IS literature, including that on health IT (HIT), has revealed the crucial role of digital capital and further differentiated between two types of digital capital: (a) application-specific digital resources (i.e., telehealth to virtualize health care consultation)(Ayabakan et al., 2023; Overby, 2008; Overby et al., 2010), and (b) infrastructure digital resources (i.e., internet connectivity and internet speed which influence the propensity of people to go online)(Agarwal et al., 2009). Accordingly, we conceptualize telehealth as a form of digital capital that virtualizes the health care process and redefines how individuals in a community can connect with care providers. However, we expect the benefits of telehealth as a form of capital for mental health to mental health of communities to intertwine with other forms of capital specifically economic, cultural, social, and digital infrastructure. Therefore, we ask the following questions:

1. What is the impact of telehealth adoption by hospitals on community mental health outcomes?
2. Under which combinations of other forms of capital (economic, cultural, social, and digital infrastructure) does telehealth adoption improve the community mental health outcomes?

To identify the effect of telehealth adoptions by hospitals in a community on community-level mental health outcomes, we consider a county to represent a community. We exploit the phased rollout of telehealth services by the hospitals in the United States using a difference-in-differences (DID) approach. This provides several benefits, primarily the creation of a natural control group, that is, those counties that have not yet adopted telehealth services. We then examine the heterogeneous association between telehealth adoptions along with the other forms of capital and mental health outcomes using a novel ML-based causal inference approach (causal forest). Rather than limit exploration of predefined interactions, the approach enables us to evaluate the combinations of the different forms of capital.

Our empirical analyses indicate that the adoption of telehealth services is associated with a significant increase in community mental health outcomes and reinforce empirical support to prior research indicating that telehealth influences mental well-being (Schriger et al., 2022). However, results from our causal forest analysis reveal substantial heterogeneity in its effectiveness, underscoring that the impact of telehealth adoption is highly context dependent. Specifically, communities with limited broadband infrastructure but strong economic or cultural capital tend to benefit the most, even in the absence of robust mental health providers. These nuanced patterns are consistent across multiple mental health outcome indicators. For example, the effect on in-person mental health visits varies by digital infrastructure and demographic composition, while suicide rates decline most in under-resourced areas where telehealth fills critical care gaps (Baird et al., 2022). Overall, our results challenge the assumption that expanding telehealth alone leads to equally improved outcomes among communities. Instead, they highlight that telehealth's effectiveness depends on the interplay of digital, social, economic, and cultural capital. Importantly, while telehealth adoption is associated with increase in mental health outcome, these benefits are highly heterogeneous across communities.

We contribute to a more comprehensive understanding of how telehealth technology impacts community-level mental health outcomes. Our work advances the theoretical understanding of information systems (IS) in health by examining community mental health outcomes through two key contributions. First, we explain how telehealth for mental health care services delivered by hospitals can benefit the mental well-being of entire communities, thereby extending beyond the micro-level individual impacts that have been previously examined (Adler-Milstein et al., 2014; Baird et al., 2022; Polinski et al., 2016). Second, we reveal heterogeneous treatment effects by demonstrating how different forms of capital (social, economic, and cultural capital) interact with telehealth adoption to influence its effectiveness. From a practical standpoint, our findings illuminate the specific conditions under which telehealth adoption proves most effective for community mental health, providing valuable guidance for future adoption strategies and policy decisions related to telehealth implementation.

2.2 Theory and Hypotheses

In what follows, we provide an overview of two streams of literature that inform our research questions. First, we conceptualize digital capital by examining the epidemiological effects of digital technologies on mental health outcomes. Second, we review mental health research through a social capital perspective, focusing on how economic, social, and cultural capital intertwine with digital technologies in mental healthcare delivery.

2.2.1 Telehealth as Digital Capital to Virtualize Mental Health Care Processes

We have positioned digital capital as a novel form of capital for mental health, different from the other forms of capital that defined. Specifically, we differentiate telehealth (i.e., application-specific digital capital) and digital infrastructure (i.e., digital penetration rate and internet speed as two components of digital capital). While telehealth adoption in a community is about virtualizing the care process for people to access mental health service in a confidential, convenient manner, the latter is about the broader capacity of the community to access digital services such as telehealth efficiently and effectively.

In extant research, researchers have mainly examined the potential of digital technologies (e.g., telehealth, self-help apps, online communities) to improve physical health outcomes, but there has been limited research on mental health well-being. Moreover, findings are mixed with respect to the plausible impacts of digital technologies on mental health. They found social features on mobile apps influence peers and found positive effects on mental health. Barney et al. (2011) reported that online depression forums cater to diverse informational needs and aid in alleviating depressive symptoms. Conversely, indicated a correlation between high broadband internet subscription rates and increased suicide rates. Social media has also been linked to negative mental health outcomes (Braghieri et al., 2022; Steers et al., 2014).

To be more specific, researchers have investigated whether telehealth adoption has had negative or positive consequences for domains such as drug abuse and mental health (Pinxten & Lievens, 2014). In addition, the COVID-19 pandemic has further accelerated the adoption of telehealth for health because it offers a safe and convenient alternative to in-person visits. Torous et al. (2020) highlighted the rapid implementation of tele-psychiatry during the pandemic, enabling continuity of care for individuals with mental health conditions. The study underscored the importance of maintaining and expanding telehealth adoption beyond the pandemic to ensure long-term access to mental health care. Moreover, in a systematic review, it is suggested that telepsychology interventions during COVID-19 have been effective in reducing symptoms of specific mental diseases such as anxiety, depression, and PTSD (Barnett et al., 2021). These findings demonstrate telehealth's potential to transform mental health service delivery during crises and in routine care. Research on telehealth's effectiveness for specific mental health conditions suggests broader applications for community mental health promotion. However, this evidence also reveals the need for IS scholars to develop more nuanced theoretical frameworks that account for the complex interactions between technology adoption and mental health outcomes at the community level.

Despite growing interest in digital technologies in healthcare, limited research has examined their impact on community-level mental health outcomes. In particular, there is a lack of understanding regarding how two distinct forms of digital capital—telehealth services and digital infrastructure—jointly influence community mental health and well-being. This gap is critical, as community mental health represents a higher level of analysis with far-reaching implications. It shapes not only individual well-being but also the broader social and economic health of the community.

Addressing mental health concerns at the community level is essential to ensure that geographically proximate populations have equitable access to mental health resources and support systems. We argue that telehealth, as a form of digital capital, holds significant promise in enhancing community mental health. Prior research has shown that telehealth can reduce barriers to care (Haleem et al., 2021), expand access to underserved populations (Silberholz et al., 2017), and deliver effective and convenient mental health services (Hilty et al., 2013). By integrating telehealth platforms into local care ecosystems, communities can make mental health services more accessible, timely, and responsive to resident needs—ultimately leading to improved psychological well-being at the population level.

Accordingly, we hypothesize the following:

Hypothesis 1: Telehealth adoption by a community's (county's) hospitals will improve a community's mental health outcomes.

2.2.2 Other Forms of Capital for Mental Health

Bourdieu's (1977, 2011b) concept of capital is particularly relevant to this study as it provides a robust theoretical lens to examine the complex interplay between various forms of capital and telehealth in shaping community mental health outcomes. This framework facilitates the generation of novel insights by highlighting that the effectiveness of telehealth interventions is not uniform, but rather contingent upon the availability and interaction of other forms of capital within a community.

As a foundational step, we conducted a comprehensive review of the mental health literature to identify key factors that influence mental health outcomes. Rather than focusing on isolated or idiosyncratic variables, this review emphasized systemic and structural determinants (Table 2). We then mapped these factors onto Bourdieu's typology of capital, demonstrating that his theory provides a coherent and integrative structure for organizing the diverse determinants of mental health. This mapping enables a more nuanced understanding of how social, economic, cultural, and digital forms of capital collectively shape community mental health and well-being.

Table 2

Review of Studies on Mental Health from Forms of Capital Lens

Authors (Year)	Forms of Capital Lens	Proxy	Health	Effect	Findings
Umberson & Karas Montez. (2010)	Social	Social connectedness	Mental Health	Positive	Social connectedness is beneficial for people's health and mental health well-being
Kawachi & Berkman. (2001)	Social	Social support networks	Depressive Symptoms and Anxiety	Positive	Social support networks play a positive role for people's psychological distress, including depressive symptoms and anxiety
Holt-Lunstad et al. (2015)	Social	Social isolation	Anxiety, Depression, and Mortality	Negative	Social isolation is contributing to the worsened mental health outcomes among retired people
Alegria et al. (2013)	Economic	Financial barriers	PTSD	Negative	Financial barriers were associated with decreased PTSD health service utilization
Kim &Von Dem (2015)	Economic	Job insecurity and unemployment	Mental Health	Negative	Job insecurity and unemployment increase the risk of developing mental health problems
Kämpfen et al. (2012)	Economic	Economic prosperity and supportive social safety nets	Mental Health	Negative	Economic prosperity and supportive social safety nets promote mental health well-being
Byrow et al. (2022)	Cultural	Language barriers	Depression	Negative	Language barriers prevent ethnic minority, immigrants, and refugees from accessing mental health services
Ganesan et al. (2021)	Cultural	Social Stigma	Suicide	Negative	People are not willing to seek mental health services due to social stigma
Kramer et al. (2002)	Cultural	Values	Distress	Negative	Asian patients are likely to express psychological distress as physical complaints

However, as illustrated in Table 2, while prior mental health research has acknowledged the importance of digital capital—both application-specific and infrastructural—as well as other forms of capital, these factors have largely been examined in isolation. This fragmented approach overlooks a critical insight from Bourdieu's theorization: forms of capital are interrelated and mutually reinforcing rather than independent. Although this integrative perspective has received limited attention in the mental health literature, a few studies in the IS domain have begun to explore such interactions. For example, Karahanna et al. (2019) demonstrated how social, economic, and cultural capital jointly contribute to hospital competitive advantage. Similarly, Hsieh et al. (2011) investigated digital inequality through the interplay of social, economic, and cultural capital. These examples underscore the value of examining how different forms of capital coalesce to influence outcomes.

Building on this perspective, our study highlights the importance of understanding how interactions among multiple forms of capital in shaping the influence of telehealth adoption on community mental health outcomes —a theme we explore in greater detail in subsequent sections.

Social Capital for Mental Health

Social capital, conceptualized as a collective feature of society rather than an individual asset, comprises formal and informal networks that serve as a public good (Bourdieu, 2011b). These networks generate valuable intangible resources—such as trust, civic engagement, and shared norms—that can be harnessed to promote community well-being (Henrich et al., 2001). Strong community social capital has been linked to greater social cohesion and equity (Hero, 2007). A growing body of research indicates that social capital influences both the risk of developing mental illnesses and the outcomes for those affected (Pinxten & Lievens, 2014). For example, Umberson and Montez (2010) emphasize the critical role of social relationships in shaping mental and physical health, underscoring the need for policies that enhance social connectedness. Similarly, Kawachi and Berkman (2001) explored the relationship between social ties and mental health while Holt-Lunstad et al. (2015) find that social isolation significantly increases mortality risk, illustrating the broader consequences of weakened social networks. Beyond social cohesion among peers and friends, Ehsan et al. (2019) shows that professional connections such as the coverage of professional health provider and the relationship between health provider and patients also affect health outcomes.

It is also important not to adopt a narrow definition of social capital limited only to family, friends, and peers. In the case of professional services, there can be greater connectedness of providers with consumers in a community if the density of providers in the community is higher. For example, a higher density of teachers in a community can lead to greater connections between teachers and students and consequently promote the flow of instrumental advice and support related to education and learning (Felten & Lambert, 2020). In our case, a higher density of both primary care and mental health providers is likely to lead to greater connections between mental health professionals and patients and thereby promote the flow of instrumental advice and support related to mental well-being.

Economic Capital for Mental Health

Community economic capital represents the material resources immediately and directly convertible into money (Bourdieu, 2011a). This form of capital plays a crucial role in shaping various aspects of life and can significantly influence overall well-being (Bourdieu, 1977). An increasing number of studies show that economic capital plays a crucial role in shaping mental health outcomes for individuals and communities. Lund et al. (2011) highlighted the bidirectional relationship between poverty and mental disorders, emphasizing the need to address poverty as a determinant of mental health. Knapp et al. (2011) underscored the economic burden of mental health disorders, estimating that the global costs associated with mental illness will surpass $6 trillion by 2030. The researchers emphasized the direct costs of health care services as well as indirect costs such as productivity loss and reduced quality of life. Moreover, economic disparities impact mental health because individuals from lower-income backgrounds often face barriers in accessing mental health care due to financial constraints and limited insurance coverage. Alegría et al. (2013) found that financial barriers were associated with decreased mental health service utilization, highlighting the importance of addressing economic factors to ensure equitable access to care. Individuals facing economic hardships often experience heightened stress, anxiety, and depression due to the constant struggle to meet basic needs and maintain a decent standard of living (Dohrenwend et al., 1992). Limited access to quality health care, including mental health services, can further compound the challenges that economically disadvantaged individuals face (Torous et al., 2020). Moreover, job insecurity, unemployment, and precarious work conditions can contribute to feelings of low self-esteem, hopelessness, and a sense of being trapped in a cycle of economic disadvantage, thereby increasing the risk of developing mental health problems (Marazziti et al., 2021). Conversely, economic prosperity, equitable distribution of resources, and supportive social safety nets can promote mental health well-being and contribute to the overall resilience of individuals and communities (Kämpfen et al., 2020). Experiences associated with economic disadvantage, including unemployment, low income, poverty, debt and poor housing, are consistently associated with poorer mental health (Allen et al., 2014; Marazziti et al., 2021). Lorant et al. (2003) investigated the association between socioeconomic status (SES) and depression, finding that lower SES is consistently linked to a higher risk of depression.

Cultural Capital for Mental Health

Community cultural capital includes features that bond a community together (Jeannotte, 2003), such as language and events that provide the means to establish traditions and expectations (Bourdieu, 2018). Cultural assets play a significant role in shaping the collective behaviors of community members (Bourdieu, 1977). In existing literature, researchers have highlighted the importance of cultural factors in understanding mental health, emphasizing how cultural beliefs and practices shape the expression and interpretation of symptoms among refugees and immigrants (Kirmayer et al., 2011). Cultural variations in understanding mental illness can impact diagnosis, treatment preferences, and stigma associated with seeking help (Griner & Smith, 2006). Moreover, language barriers can create challenges in accessing mental health care because individuals with limited English proficiency may face difficulties in effectively communicating their concerns (Sentell et al., 2007). Language barriers are associated with lower utilization of mental health services among ethnic minority populations, immigrants, and refugees (Byrow et al., 2020). These findings underscore the need for culturally sensitive and linguistically appropriate mental health services to ensure accessibility and quality of care for diverse populations.

Cultural norms and linguistic factors also influence the delivery and effectiveness of mental health interventions. Griner and Smith (2006) demonstrated that providing culturally adapted therapy materials and bilingual therapists improved treatment satisfaction and engagement among Hispanic clients. These findings emphasize the need for culturally informed approaches in mental health interventions to address effectively the diverse needs of individuals from different cultural and linguistic backgrounds.

Intertwining of Digital Capital with Other Forms of Capital

Understanding how different forms of capital interact is crucial for comprehending their collective impact on a community's mental health well-being, particularly for developing interventions and policies that promote community mental health. A better understanding of the heterogeneous effects of telehealth on mental health will likely help clinicians and policymakers to identify subareas where they prioritize telehealth for mental health service.

While existing literature suggests that telehealth adoption holds promise for expanding access to mental health care, its impact on community-level mental health outcomes is unlikely to be uniform across contexts. Communities differ significantly in their social, economic, infrastructural, and cultural characteristics, which shape both the availability of health care services and the ability of residents to engage with them. These contextual factors are likely to moderate the relationship between telehealth adoption and mental health outcomes individually and collectively. For example, variations in broadband infrastructure, provider density, economic capital, and cultural resources may interact in complex ways to either amplify or constrain the potential benefits of telehealth.

In this sense, we expect the causal impact of telehealth adoption to be conditional rather than universal. Rather than assume a homogeneous treatment effect, we posit that heterogeneous treatment effects are likely—where the size and even direction of the effect depend on the underlying community context. To investigate this empirically, we apply a causal forest approach (Athey & Imbens, 2017; Athey & Wager, 2019), which is well suited to uncovering complex, nonlinear interactions and conditional effects without imposing restrictive parametric assumptions. This methodological choice allows us to explore the interdependent nature of moderating variables while maintaining flexibility in capturing nuanced effect heterogeneity. Therefore, we posit the following:

Hypothesis 2: Combinations of other forms of capital (economic, social, cultural, and digital infrastructure) will moderate the effect of community (county) hospital adoption on community mental health.

2.3 Data and Measurement

2.3.1 Data Sources

We compiled a longitudinal county-level dataset from 2017 to 2021, using the American Hospital Association (AHA), Area Health Resources Files (AHRF), American Community Survey, and County Health Ranking (CHR) data from the University of Wisconsin Population Health Institute. We obtained outcome data for average mentally unhealthy days per U.S. County from CHR, which the University of Wisconsin Population Health Institute developed and extracted from other secondary sources such as from the Centers for Disease Control (CDC). We acquired covariates used for coarsened exact matching (CEM) from the AHA and AHRF datasets. We selected controls and covariates following a literature review. After removing counties with missing data (6%), our resulting dataset included variables for 2,794 counties.

Dependent Variables: Community Mental Health Outcomes

We conceptualize community mental health as a multidimensional construct with three key components. Mental Health Clinic Visits measure community engagement with formal care infrastructure, capturing behavioral indicators of treatment access and utilization. Poor Mental Health Days quantify subjective psychological distress by measuring the percentage of days in the past month that individuals report their mental health as “not good.” Suicide Rate represents the most severe outcome, indicating complete system failure when both individual coping mechanisms and community support prove insufficient. This framework captures the full spectrum of community mental health—from proactive care seeking to subjective distress to tragic crisis outcomes. Below, I we examine each component accordingly.

Poor Mental Health Days. This variable was originally from the nation's premier system of health-related telephone surveys, the Behavioral Risk Factor Surveillance System (BRFSS). With technical and methodological assistance from CDC, state health departments use in-house interviewers or contract with telephone call centers or universities to administer the BRFSS surveys continuously through the year. The states use a standardized core questionnaire, optional modules, and state-added questions.[2] The variable "mentally unhealthy days" is from the following survey question: "Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" The aggregate county-level average number of mentally unhealthy days reported in the past 30 days (age-adjusted) have been compiled in the annual CHR data from the University of Wisconsin Population Health Institute, which is a general measure of health-related quality of life in a population. Self-reported health status is a widely used measure of health-related quality of life (Nyman et al., 2007). In addition to measuring how long people live, it is essential to include measures that consider how well people live. Further, reports of days of unwell mental health are a reliable estimate of recent health.[3]

Suicide Rates. Suicide represents the most severe and terrible result of poor mental health, which frequently happens when people feel psychological discomfort, hopelessness, or a lack of support overtakes them (Kyung et al., 2017). It is not only a personal problem; rather, it reflects

[2] https://www.cdc.gov/brfss/about/brfss_faq.htm
[3] https://www.countyhealthrankings.org/explore-health-rankings/county-health-rankings-model/health-outcomes/quality-of-life/poor-mental-health-days?year=2022

larger institutional and community-level shortcomings in recognizing, assisting, and treating those who are at risk. High suicide rates are frequently a sign of poor community mental health outcomes (Barnett et al., 2021), indicating underlying issues such as widespread psychological distress, social disconnection, economic hardship, and inadequate access to mental health care. They reflect not only individual struggles but also broader societal conditions that may contribute to feelings of hopelessness and despair. As such, suicide rates can serve as a critical barometer for assessing the effectiveness of mental health systems and community support structures. This data is originally from the CDC National Vital Statistics System Mortality Dataset, and we extracted county-level suicide numbers and created the ratio of suicide to the total population of the county.

In-Person Mental Health Service Visitation. It measures the number of SafeGraph-based mental health service visits per capital for a given year (the ratio of SafeGraph-based visits to the total population of the county). We obtained visitation data from the SafeGraph point of interest database.[4] SafeGraph is a company that tracks more than 45 million mobile device access from its household census block groups (CBG) to about seven million points of interest (POIs) in the United States and Canada. SafeGraph visitation data has been used broadly in public health and business research (Y. Wang et al., 2022; H. Zhou et al., 2022). The company collects this data by monitoring mobile devices that have opted in through an array of navigation and mobile devices that have opted in through an array of navigation and mobile application, where users agree to share their geographic location. The methodology for determining the CBG for each device involves analyzing the most frequent nighttime location over 6 weeks. This process ensures that the data accurately reflect the typical residential location of the device's user. SafeGraph utilizes the GPS coordinates for a device and compares them with a POI's building footprint to verify

[4] SafeGraph: https://docs.safegraph.com/docs/monthly-patterns

access to a particular location. To ensure the accuracy of the collected data, a device must remain at the POI for a minimum of 4 minutes, eliminating any brief or accidental visit. The company also considers additional relevant location information and ML algorithms to identify the visitation at a specific location. SafeGraph Technical Guidance has further details on how SafeGraph identifies POI visitation.[5]

In this study, we used the SafeGraph Patterns dataset, recording the number of visits and median dwell time of POIs. We further aggregated CBG-level visitation to the county level and computed the yearly visit as in-person mental health service visits for the POIs that are categorized as office of mental health practitioner (except physicians) based on the North American Industry Classification System (NAICS)[6] with NAICS code 621330.

Independent Variables: Telehealth Adoption in the Community (County).

We used the responses from the AHA Annual Survey: "Does your hospital provide telehealth office visit service?" or "Does your hospital and their affiliates provide tele psychiatry service?" to determine whether a hospital (and affiliates such as clinics) offer telehealth. We selected to use both questions because psychiatrists, other specialists, and primary care providers can provide mental health services directly. Thus, we aggregated responses for these two questions per hospital and generated a single dichotomous variable for each hospital. We aggregated the hospital-level data into county level, which tells us how many hospitals per county provide telehealth-related services per year. We assigned 1 to those counties with hospitals providing telehealth services, otherwise 0.

[5] SafeGraph Technical Guidance: https://www.safegraph.com/guides/places-data-technical-guide
[6] North American Industry Classification System: https://www.census.gov/naics

Covariates

We included demographic characteristics, SES measures, health care provider conditions, and behavioral factors at the county level as covariates (Table 3). We identified these factors because they were likely to operate as confounders (i.e., predictors of mental health outcomes distributed differently among the levels of telehealth) or moderators (i.e., associations between telehealth and mental health outcomes differ by the levels of these factors). We chose those covariates not via a data-driven approach but rather based on the subject-matter knowledge and the literature on mental health (Byrow et al., 2020; Griner & Smith, 2006). Guided by the literature, we mapped them to the other forms of capital and with demographic factors. Then, we used the same set of variables in our causal forest analyses because they can be important determinants of heterogeneous treatment effects, based on existing literature (Allen et al., 2014; Griner & Smith, 2006; Lorant et al., 2003).

2.3.2 Empirical Framework

Our empirical strategy consists of three components designed to capture both the causal effects and mechanisms through which telehealth adoption affects community mental health outcomes. First, we employed a DID model specification combined with matching-based sample construction to estimate the average treatment effects of telehealth adoption. This approach leverages temporal and cross-sectional variation to isolate the impact of adoption on key outcomes such as poor mental health days, suicide rates, and mental health clinic visits. Second, we used causal ML, specifically causal forests (Athey et al., 2019; Athey & Imbens, 2017), to uncover heterogeneous treatment effects. This allowed us to identify which types of counties—based on demographic, economic, or health system characteristics—experience the greatest improvements in mental health outcomes from telehealth adoption. These first two methods align with recent developments in marketing and accounting research that emphasize causal inference and predictive heterogeneity (Guo et al., 2021; Iyengar et al., 2022).

Table 3

Definitions of Other Forms of Capital and Demographics

Forms of capital	Definition	Proxy	Reference	Definition of Proxy
Cultural capital (CC)	The embodied competencies for adopting Tele-health Services	County-level Not Proficient in English Rate	Bourdieu, 1991, Pinxten & Lievens, 2014	Percentage of population not proficient in English
Social capital (SC)	The resources from social networks for adopting tele-health services	County-level Primary Care Physicians Rate	Mohnen et al., 2012	Primary Care Providers per 100,000 population
		County-level Mental Health Provider Rate	Mohnen et al., 2012	Mental Health Providers per 100,000 population
Economic capital (EC)	The monetary means to acquire and access Telehealth	County-level Unemployment Rates	Alessie et al., 1997	Percentage of population without employment
		County-level Uninsured Rates	Alessie et al., 1997	Percentage of population under 65 without insurance
Digital Capital (Infrastructure)	Access rates and speed of the Internet	County-level Broadband Rates	Ragnedda, 2018	Percentage of households with broadband internet connection
		County-level average Broadband Speed	Ragnedda, 2018	Average internet download speed
Demographic Factors (used as control variables)		County-level Hispanic Population Rate	Paccoud et al., 2020	Percentage of population who are Hispanic
		County-level Black Population Rate	Paccoud et al., 2020	Percentage of population who are Black or African American
		County-level Rural Population Rate	Paccoud et al., 2020	Percentage of population who located in rural areas

Third, we conducted a causal mediation analysis to explore why telehealth adoption affects community mental health outcomes. Specifically, we tested whether and to what extent changes in in-person mental health clinic visits mediate the relationship between telehealth adoption and key mental health indicators, such as poor mental health days or suicide rates. This approach enabled us to decompose the total treatment effect into direct and indirect pathways, offering deeper insight into the mechanisms of impact. By integrating these three methods—DID for identification, causal forests for heterogeneity, and mediation analysis for mechanism—we provide a comprehensive, multilayered assessment of how telehealth adoption shape community-level mental health outcomes.

Identification Strategy

To test whether the adoption of telehealth can impact community mental health well-being, we conducted a standard DID regression with multiple groups and time periods following previous literature (Angrist & Pischke, 2008). To predict the counterfactual, we needed a natural way to identify counties that share similar characteristics but did not receive the telehealth adoption treatment during the same time. We defined our treatment groups as those counties that did not have telehealth in 2017 but subsequently adopted them in 2018, 2019, or 2020, while our control groups are those counties that did not adopt telehealth from 2017 to 2021. We performed estimation using ordinary least squares. The unit of analysis is the community's (county's) mental health well-being, measured by mentally unhealthy days. The DID equation representing our model is as follows:

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 p_{it} + \beta_3 (T_{it} \times p_{it}) + \mu_i + \vartheta_t + \varepsilon_{it}$$

β_0 is the constant, β_1 is the effect from the treatment group, β_2 represents posttreatment periods, and β_3 is the treatment effect (i.e., the DID effect), which is the expected value difference in the time trend and the difference between treatment and control groups after treatment. We included county fixed effects (μ_i) to address any time-invariant county heterogeneity and time fixed effects to address time trends (ϑ_t).

In the DID analysis, we accounted for time-invariant region-level heterogeneity using location fixed effects (i.e., county fixed effects). These fixed effects capture time invariant region-level factors (e.g., public health service conditions, mobile internet development) that may correlate with the adoption of telehealth. For example, demographic (e.g., age or gender) or socioeconomic conditions of a region are relatively stable across years and are thus captured by the region fixed effect. Second, we included time fixed effects at the year level to control for seasonal or temporal changes in mentally unhealthy days. Table 3 above provides details regarding the data sources and measures.

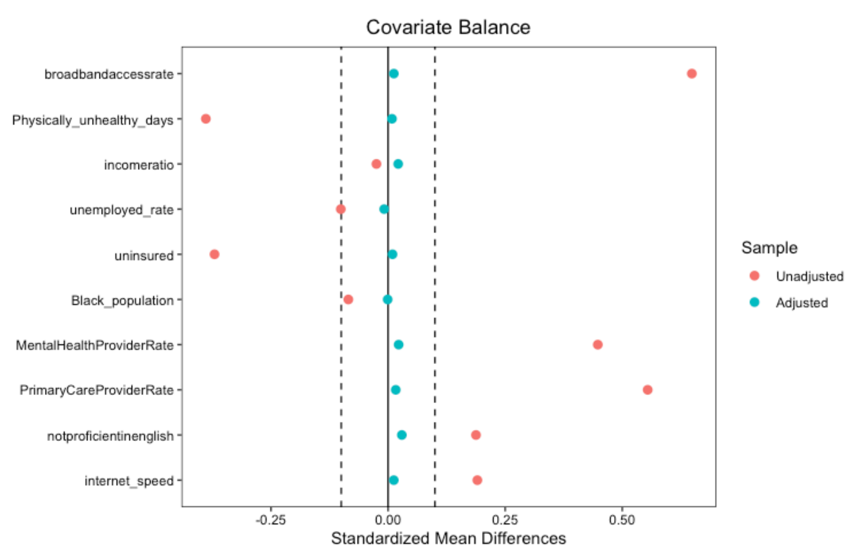
Matching

A naive DID approach will lead to biased estimates if the adoption of telehealth is not made randomly. To correct for endogenous selection of treatment groups and to balance the covariates between the treatment and control groups in our analysis, we used a one-to-many CEM procedure using covariates to mitigate self-selection bias (Iacus et al., 2011). Our covariates consisted of county-level variables, derived from related literature (Byrow et al., 2020; Griner & Smith, 2006). The matching algorithm produces a cell for each combination of attributes used in the matching. We retained cells with at least one treated and one control county (some retained cells have more). This is required to estimate the treatment effect and is standard for matching estimators (Abadie et al., 2004). The remaining observations comprise the matched sample, which consists of 3,251 matched counties in 788 cells. Matched counties have similar demographic characteristics and socioeconomic levels.

We also assessed whether the distribution of the covariates is similar across the treated and control groups after matching. Following Austin (2009) and Imbens and Rubin (2015), we examined the standardized differences in covariate means between the treatment and control groups. Figure 1 presents the absolute value of the standardized differences for each covariate. Matching leads to a substantial improvement in balance. After matching, all the standardized differences are below 0.1, a degree of balance comparable to what one might expect in a completely randomized experiment (Imbens & Rubin, 2015; Stuart, 2010). We verified the validity of our CEM by checking whether there are systematic differences in covariates of interest across the two groups. There are no significant covariate differences between the treatment and control groups.

Figure 1

Covariate Balance After CEM



Method for Estimating Heterogeneous Treatment Effects: Causal Forest

Knowing that a treatment works on average is often not sufficient. Understanding which counties benefit the most from the telehealth is essential to the evaluation of existing adoption practices and the design and targeting of new telehealth adoption initiatives. Thus, we need to estimate heterogeneous treatment effects. Heterogeneity refers to the existence of differences in the effect of treatment or intervention across different subgroups. This implies that the treatment's effectiveness can fluctuate based on the county's characteristics. Acknowledging heterogeneity is critical because it shapes intervention design and evaluation. Uniform interventions are suitable if effects are consistent across subgroups, but varying effects require tailored approaches for maximum efficacy. Neglecting this diversity can yield biased treatment effect estimates. Thus, employing methods that accommodate heterogeneity is paramount when gauging a treatment's causal effect.

The causal forest method (Athey & Imbens, 2017; Athey & Wager, 2019; Luo et al., 2019) adapts the random forest technique from ML and applies it within a causal inference framework. Unlike traditional regression models, where incorporating interactions relies on assuming linear or nonlinear relationships, this offers a nonparametric, computationally efficient way to identify differences in outcomes between treatment and control units across multidimensional covariate space. With this splitting process, we aim to find optimal splits that maximize the difference in treatment effects while maintaining accurate effect estimation. Consequently, the model provides conditional average treatment effects (CATE) for specific subgroups, enabling predictions of individual effects for future observations with similar characteristics (Athey & Wager, 2019). We find it appropriate to explore the heterogeneous treatments of telehealth adoption on community (county)'s mental health well-being because we look beyond the two-way interaction of telehealth adoption with one county characteristic, and we explore how additional characteristics combined play a role in the impact of telehealth on community's mental health outcome.

2.4 Results

In this section, we execute each of the empirical strategies and report the corresponding results.

2.4.1 Average Treatment Effects

Table 4 reports the CEM DID regression result. We observed that telehealth adoption has a significant and meaningful impact on mental health outcomes at the county level. Specifically, telehealth is associated with a modest but statistically significant reduction in self-reported mentally unhealthy days, indicating potential improvements in general psychological well-being. More notably, it is linked to a substantial decrease in suicide rates, suggesting that expanded access to remote mental health services may play a critical role in preventing severe mental health crises. Additionally, telehealth adoption corresponds with a significant decline in in-person mental health visits, implying a substitution effect where remote care may replace face-to-face services. These findings underscore the potential of telehealth to enhance community mental health outcomes.

Table 4

CEM DID Result for Telehealth Adoption on Community Mental Health Outcomes

Variables	**Mental Unhealthy Days**	**Suicide Rate**	**In-person Mental Health Visit**
Post* Treated(β_3)	-0.0014 (0.005) ***	-0.156 (0.052) ***	-0.173 (0.058) ***
Controls	√	√	√
County Fixed Effects	√	√	√
Time Fixed Effects	√	√	√
R^2	0.06	0.2	0.07

Note: ***p <0.01, Robust standard errors reported in parentheses

While the results of the baseline DID estimations are compelling, it is important to note they are subject to several assumptions. The parallel trend assumption is the most critical to ensure internal validity of DID models. Violation of parallel trend assumption will lead to biased estimation of the causal effect. To test parallel trends, we visualized the average community mental health well-being for each group at a point in time (years) relative to when telehealth was adopted. To further test the parallel trends assumption statistically, we interacted pretreatment values with corresponding time dummies within the DID model. None of the coefficients was significant, which suggests that the trends are sufficiently common between the control and treatment groups.

2.4.2 Heterogeneous Treatment Effects

We employed the causal random forest approach with honest tree algorithms (Athey & Imbens, 2016; Athey & Wager, 2019) to estimate the heterogeneous treatment effects (HTEs) in our study. By leveraging the flexibility of causal forest modeling, we aimed to identify well-matched subgroups and understand the variations in treatment effects. In implementing this method, we first estimated the contribution of county features toward the heterogeneity in the treatment effect of telehealth adoption on community (county)'s mental health well-being. The variable importance score from this analysis denotes the percentage of times a variable is used in the tree splitting decision. Table 5 reports the results of the variable importance analysis, which broadband access rate determines as 12.6% of tree partition followed by 11.17% by mental health professional rate, and unemployment rate by 8.4%. The high variable importance score may be due to the splits at the bottom of the tree, which are less critical than the splits at the top of the tree. We examined this fact and found that the high importance score variables in Table 5 determined the top slits in our causal trees. We report the variable importance score of the top 10 county features in Table 5. These results show that county features under different categories, such as digital infrastructure (e.g., broadband access rate), social capital (e.g., health professional rate), and economic capital (e.g., unemployment rate) significantly contribute to elucidating the heterogeneity in telehealth effect on mental health.

Table 5

Variable Importance Score of Top 10 Features

Ranking	Feature	Score
1	Broadband access rate	0.126
2	Mental health professional rate	0.117
3	Unemployment rate	0.084
4	Primary care professional rate	0.075
5	Hispanic persons rate	0.061
6	Black persons rate	0.053
7	Not insured rate	0.051
8	Not proficient in English rate	0.042
9	People living in rural area rate	0.035
10	Internet speed	0.035

Note: The importance of the score denotes how many times a feature was split at each depth in the causal forest.

Second, we conditioned our model to identify heterogeneous effects across the top 10 variables shown in Table 5 to ensure the coverage of important features (Athey & Imbens, 2016). We optimized and tuned our causal forest model based on the recommendations from Athey and Wager (2019).

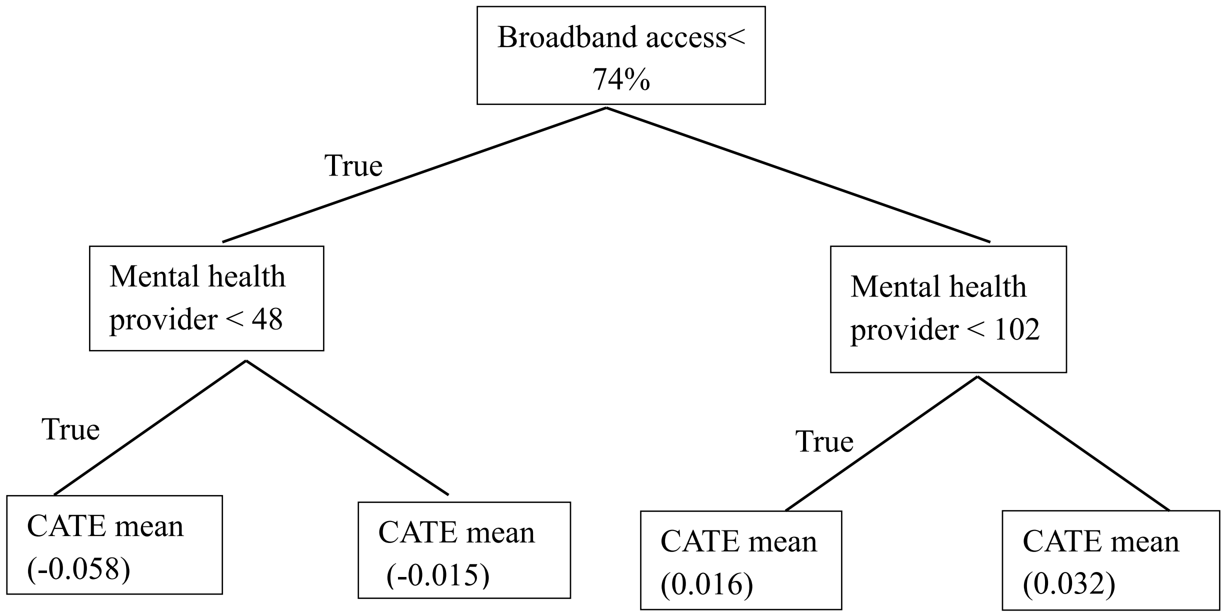
Overall, our causal forest indicates that adoption of telehealth has an average treatment effect of -.013 on community (county)'s mental health well-beings (measured by county's mentally unhealthy days).

To further investigate how community characteristics influence the effectiveness of telehealth adoption, we constructed a depth-2 policy tree on the Conditional Average Treatment Effect (CATE) for average mentally unhealthy days. The tree begins by splitting counties based on broadband access, using a threshold of 74%. Among counties with less than 74% broadband coverage, the next split considers the availability of mental health providers, with a threshold of 48. Counties falling below both thresholds—those with limited broadband and fewer than 48 mental health providers—show the strongest reduction in mentally unhealthy days, with a CATE of −0.058. This suggests that telehealth offers the greatest benefit in the most resource-constrained environments. Counties with limited broadband but a higher number of providers see a smaller but still negative effect (CATE = -0.015). On the other side of the tree, for counties with broadband access ⩾ 74%, the model splits again at 102 providers. These relatively well-resourced counties see positive CATE values—indicating a slight increase in mentally unhealthy days (CATE = 0.016 and 0.032), implying that telehealth may be less effective or even counterproductive in some higher-capacity environments. As figure 2 indicated, this policy tree highlights the importance of local digital and healthcare capacity in shaping telehealth's impact and underscores the need for context-aware implementation strategies.

Similarly, we conducted causal forest on other components of community mental health outcome including in-person mental health visit and suicide rates. Figure 3 and Figure 4 illustrates how the effectiveness of telehealth varies across counties with high broadband access, based on contextual factors such as provider availability, insurance coverage, racial composition, and internet speed. Overall, the CATE is modestly negative (-0.023), but significant heterogeneity emerges across subgroups. In counties with limited mental health provider availability and high uninsured rates, telehealth is associated with strongly negative effects (CATE = -0.132), suggesting that virtual care may be insufficient when baseline access to care is already severely constrained. In contrast, in counties with relatively low Hispanic population rates and high internet speed, telehealth shows a positive impact (CATE = 0.061), indicating that infrastructure and possibly trust or usability factors amplify its benefits.

Figure 2

Depth 2 Policy Tree on CATE on Mental Unhealthy Days

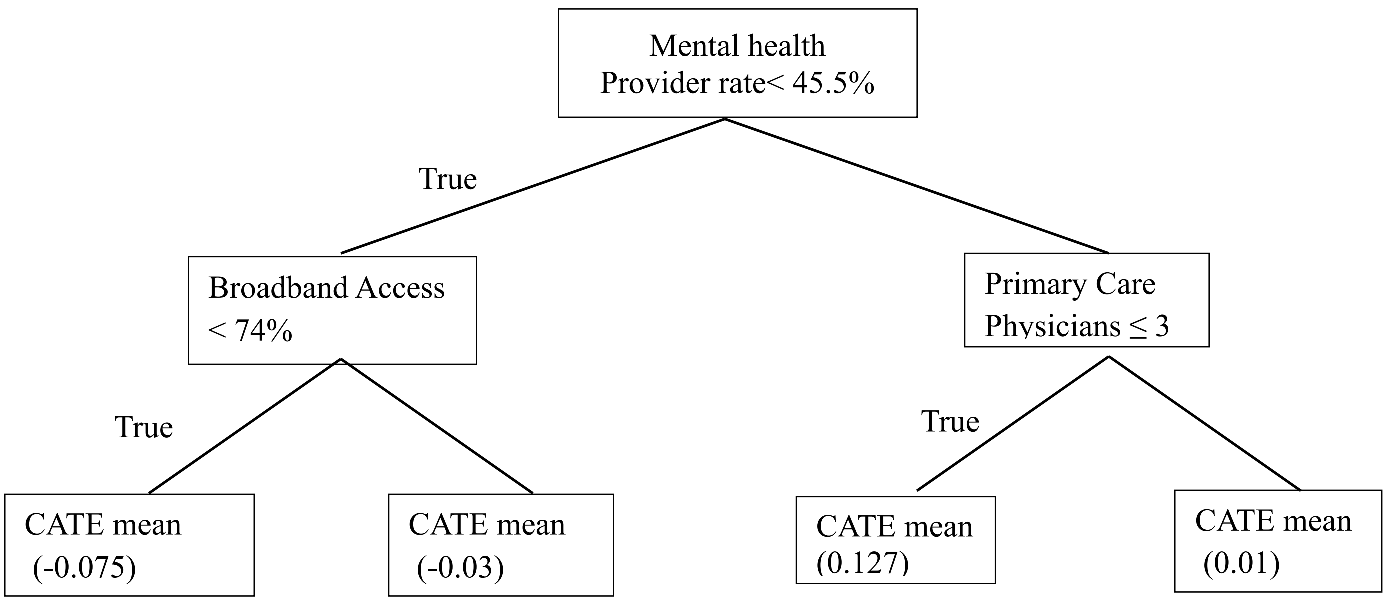


Note: Communities with less social resources, proxied by mental health provider rate, have the best treatment effect of telehealth adoption when the broadband access in the community is relatively low. (The left branch displays counties where a condition is smaller than the value, and the right branch shows counties where the condition is bigger than the value. The "< >=" value serves as the continuous cutoff for splitting, and the boxed value represents the conditional average treatment effect.)

The causal forest analysis reveals meaningful heterogeneity in the impact of telehealth adoption on in-person. Among counties with a low density of mental health providers (below 45.5%), further stratification by broadband access indicates that the treatment effect is more negative in areas with broadband access below 74% (CATE =-0.075), suggesting that poor internet infrastructure may limit the effectiveness of telehealth in preventing suicides. Even in low-provider regions with slightly better broadband (≥ 74%), the effect remains negative but is attenuated (CATE = -0.03), indicating marginal improvement. In contrast, counties with more mental health providers but a limited number of primary care physicians (≤ 3) experience a strong positive effect of telehealth on reducing suicide rates (CATE = 0.127). This subgroup likely benefits from the supplementary role of telehealth where traditional primary care is limited. Finally, in areas with both sufficient provider capacity and primary care availability, the effect is positive but modest (CATE = 0.01). Overall, these findings emphasize that telehealth's effectiveness in lowering suicide rates is contingent on both internet infrastructure and provider availability, with the greatest gains observed in under-resourced areas where telehealth fills a primary care gap.

Figure 3

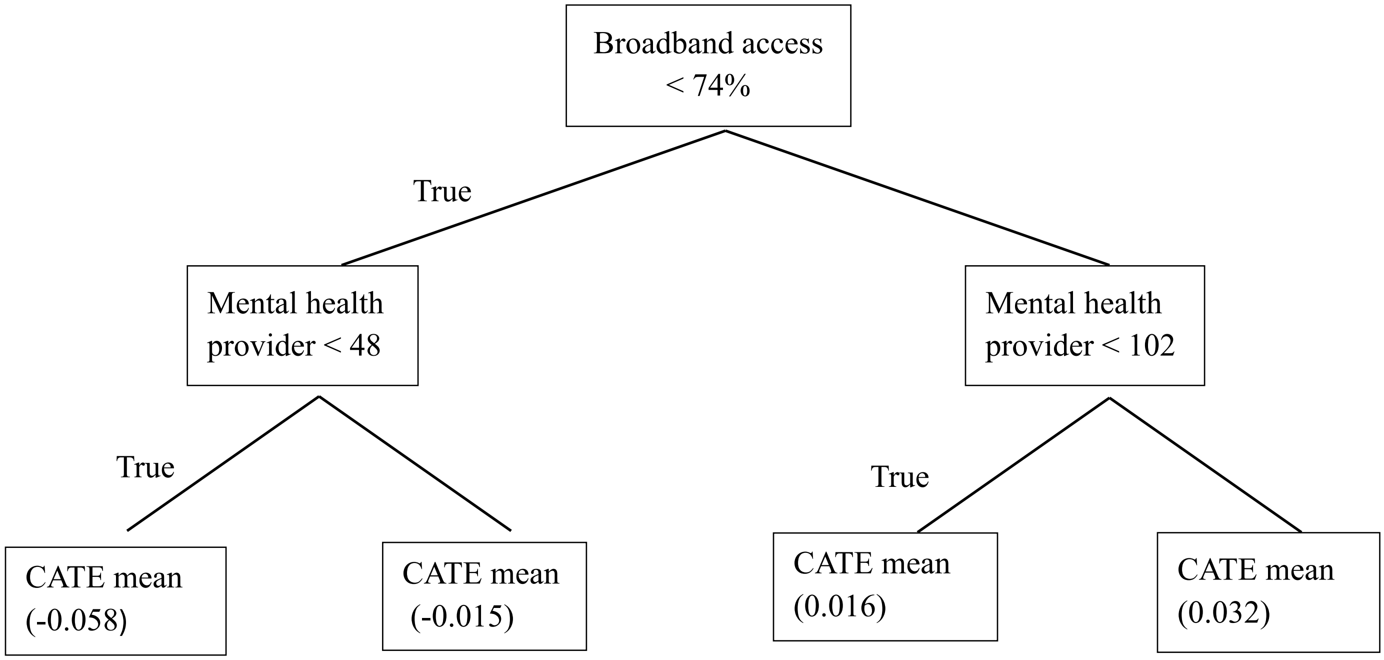
Depth 2 Policy Tree on CATE on In-Person Mental Health Visit



Note: Communities with lower mental health provider rates (less than 45.5%) show heterogeneous treatment effects of telehealth adoption depending on their broadband infrastructure. Among these less socially resourced areas, counties with limited broadband access (< 74%) experience the most favorable treatment effect, as indicated by a CATE of -0.075. This suggests that telehealth is especially effective in digitally and clinically underserved environments. Conversely, counties with better broadband access in the same low-provider group show a smaller improvement (CATE = -0.03). For counties with higher provider availability (⩾ 45.5%), the presence of primary care physicians further differentiates the effect: those with ⩽ 3 physicians show an unexpectedly positive CATE (0.127), while those with more resources see a modest effect (CATE = 0.01). As with all splits, the left branch represents counties where a condition is **less than** the threshold, and the right branch represents counties where the condition is **greater than or equal to** it. The cutoff values are selected based on model-driven partitioning, and the boxed values represent the conditional average treatment effects **(CATEs) for** each subgroup.

Figure 4

Depth 2 Policy Tree on CATE on Suicide Rates



Note: Communities with lower levels of broadband access (< 74%) derive the most substantial benefits from telehealth adoption when they also have limited mental health provider availability. Specifically, counties with both low broadband access and fewer than 48 mental health providers experience the largest treatment effect (CATE = −0.058), indicating meaningful improvements in mental health outcomes. Within the same broadband-constrained group, counties with relatively more providers (≥ 48) show a smaller benefit (CATE = −0.015). On the other side of the split, where broadband access is higher (≥ 74%), treatment effects become positive: counties with fewer than 102 mental health providers exhibit a modest positive CATE (0.016), while those with more providers show an even greater increase (CATE = 0.032). As before, the left branches reflect counties where the condition is less than the threshold, and right branches reflect counties where the condition is greater than or equal to it. The boxed values represent the conditional average treatment effect (CATE) for each subgroup identified through model-based partitioning.

The tree first splits counties based on broadband access, identifying 74% as a critical threshold. This split reflects the foundational role of digital infrastructure in enabling virtual care. For counties with broadband access below 74%, the second-level split is determined by the availability of mental health providers, using a cutoff of 48 providers. In this segment, counties with both limited broadband and fewer than 48 mental health providers experience the largest reduction in suicide-related visits, with a CATE of -0.058: indicating telehealth had a strong protective effect. In contrast, counties with similarly low broadband but more mental health providers still benefit (CATE = -0.015), though to a lesser extent. On the opposite side of the tree, counties with broadband access ≥ 74% are further split at 102 providers. Interestingly, these relatively well-resourced counties show positive CATE values (0.016 and 0.032), implying either neutral or slightly adverse effects of telehealth adoption on suicide-related visit rates. This result suggests that in some high-capacity areas, telehealth may be a supplement rather than a critical enabler, and its deployment may not reduce suicide-related crises without targeted intervention strategies. Overall, this policy tree offers actionable insight: telehealth has the greatest impact in digitally and medically underserved communities, and policies should prioritize these areas to achieve the strongest mental health outcomes. This finding challenges a one-size-fits-all approach and emphasizes the importance of equity-focused telehealth expansion strategies.

2.4.3 Causal Mediation Mechanism

We conducted additional analyses to gain deeper insight into the mechanism by which telehealth adoption influences *community mental health outcomes*. Drawing from prior literature, we theorized that one key channel through which telehealth adoption affects community mental health is by altering in-person mental health clinic visitation. Specifically, we expect that as telehealth becomes more accessible, it may substitute for in-person care, thereby reducing clinic visits, which in turn influences mental health outcomes such as poor mental health days and suicide rates**.** To investigate this mechanism, we employed a causal mediation analysis (Imai et al., 2010) a method increasingly used in empirical social science and health research to decompose total treatment effects into direct and indirect pathways. In our context, the total effect of *telehealth adoption* on a given mental health outcome is partitioned into (a) a direct effect, capturing the influence of telehealth independent of in-person care; and (b) an indirect effect**,** which operates through changes in in-person mental health clinic visits. We applied this framework using the outcome of poor mental health days as our dependent variable, with in-person clinic visits as the mediator.

The results, reported in Table 6, indicate that a statistically significant portion of the total effect of telehealth adoption on poor mental health days is mediated through reduced in-person clinic visitation. This suggests that telehealth's substitution effect plays a key role in shaping community-level mental health outcomes. Specifically, we found that telehealth adoption significantly decreases in-person visits, and that lower in-person care is associated with fewer reported poor mental health days, supporting our theorized indirect pathway. The average causal mediation effect is both statistically significant and substantively meaningful, accounting for a notable share of the total treatment effect.

Our causal mediation analysis supports the view that telehealth adoption improves mental health outcomes not only through direct channels (e.g., increased mental health days and decreased suicide rates) but also indirectly by reshaping how communities engage with existing care systems. These findings highlight both the mechanism of telehealth's effectiveness and the importance of considering substitution effects when evaluating digital health interventions.

Table 6

Mediating Role of In-Person Mental Health Visits in the Telehealth Effect on Mental Health Outcomes

Treated Units	All Counties
Mediation effect (Telehealth → In-person visits → Poor Mental Health Days)	−0.0021 [-0.0034, 0.0011] ***
Direct effect (Telehealth → Poor Mental Health Days)	−0.0046 [−0.0068, -0.0027] ***
Total effect (Telehealth on Poor Mental Health Days)	−0.0067 [−0.0093, 0.0042] ***
Proportion mediated	0.313 [0.193, 0.447] ***

Note: Confidence intervals appear in brackets and are based on 1,000 bootstrap simulations.

***p < 0.01. All models include controls, county fixed effects, and time fixed effects.

2.5 Robustness Checks

While the results hold across several specifications and consistently support the notion that the adoption of telehealth services is associated with an increase in community mental health well-being, alternative explanations of this relationship exist. We thus extended our empirical analyses to ensure the robustness of the observed effect. First, we evaluated the sensitivity of our DiD results with respect to the matching method. Second, we analyzed whether modifying the treated group changes our findings. Specifically, we redefined our treatment groups as those counties or their neighboring counties that did not have telehealth services in 2017 but subsequently adopted them in 2018, 2019, or 2020. Finally, we reiterate that our results are similarly valid under nonparametric estimation.

2.5.1 Sample Selection Bias

Community's hospitals make telehealth adoption decisions themselves. Therefore, our sample has potential self-selection bias. To address this statistically, we used a Heckman model (Heckman, 1979), which corrects this selection bias in the first stage through the selection equation together with the conditional expectation of the dependent variable through outcome equation. The selection equation helps to understand the factors that determine whether an individual county belongs to the selected treated group or not. It models the probability of selection as a function of various exogenous variables. For the selection model to control for endogeneity effectively, the exclusion restriction variables must be important determinants of the dependent variable in the selection model (i.e., isadopt) but unrelated to the dependent variable in the treatment effect model (i.e., mentally unhealthy days, suicide rate)(Lennox et al., 2012). We introduced an exclusion restriction variable county_size $_{(i,t)}$, in the first stage. We note that county_size $_{(i,t)}$affects county's adoption of telehealth in the selection equation but does not directly affect the county's overall mentally unhealthy days in the outcome equation once the telehealth adoption has been established. The following are the reasons for this. On the one hand, the geographic availability of health resources determines hospitals' decisions on whether to adopt telehealth (Cummings et al., 2017); thus, county_size $_{(i,t)}$ can directly affect the dependent variable of the selection equation on whether to adopt telehealth. On the other hand, mentally unhealthy days vary significantly across counties; therefore, the county size is not likely to affect the county's mentally unhealthy days. Consequently, county_size $_{(i,t)}$ will not directly affect the dependent variable of the outcome equation on the mentally unhealthy days. Thus, it can serve as an exclusion restriction variable in the first stage. We can now write down the first stage selection equation as the following probit model:

$$\begin{aligned} \text{isadopt}_{i,t} &= \begin{cases} 1 & \text{if isadopt} \\ 0 & \text{otherwise} \end{cases} \\ \text{isadopt}^{*}_{i,t} &= \alpha_{0,i} + \alpha_1 w_{i,t} + \boldsymbol{\alpha} \boldsymbol{X}_{i,t} + u_{i,t} \end{aligned}$$

where $w_{i,t}$ is the additional exogenous variable county_size $_{i,t}$,$i$sadopt$^{*}_{i,t}$ is a latent variable, and $u_{i,t}$ is the idiosyncratic error term. Vector $\boldsymbol{X}_{i,t}$ contains the covariates for the selection equation. In the second stage, we estimated the outcome equation on the community mentally unhealthy days by adding the inverse Mills ratio (IMR) to control for the selection bias, which is computed from the first-stage regression. The IMR is significant in the second stage, thereby indicating that there is indeed a sample selection bias in the yearly telehealth adoption decision (i.e., whether to adopt or not), and this bias is corrected by the first stage selection equation in the Heckman selection model. The results remain consistent with our primary results.

2.5.2 Expanded Treatment Groups

For counties that do not provide telehealth services, there is an alternative way for people living in those counties to get access to telehealth: receiving telehealth services provided by providers in neighboring counties. To include those areas, we defined our treatment groups as those counties or their neighboring counties that did not have telehealth services in 2017 but subsequently adopted them in 2018, 2019, or 2020, while our control groups are those counties or neighboring counties that did not adopt telehealth services from 2017 to 2021. We have 1,750 counties in treated groups and 1,089 counties in control groups. The result is consistent with our primary results.

2.5.3 Alternate Staggered DID Estimation

Because telehealth adoption has occurred in many different counties during different years. We adopted lead and lag estimator by Angrist and Pischke (2008) and Autor (2003) to demonstrate the robustness of our results. The estimator defines group-time average treatment effects, ATE (g, t), to be the average treatment effect for a particular group at a particular time. The result is consistent with our main findings.

2.5.4 DID With Look-Ahead Propensity Score Matching (LA-PSM)

Following the methodology of Bapna et al. (2018), we utilized look-ahead propensity score matching (LA-PSM) as a supplementary matching technique to our primary propensity score matching (PSM) method. LA-PSM offers an advantage over traditional PSM by mitigating biases arising from omitted unobserved variables. It does so by including control units subjected to the treatment at a later stage. In our study, this means we first identified non-banned users comparable to banned users using PSM and then refined this to include only those banned subsequently.

2.5.5 Alternate Functional Form

Another robustness test involves evaluating the linear and additive structure of our main estimation. The estimation via causal forest allows us to relax this assumption because of its nonparametric structure, and as we discussed in section on HTEs estimated via causal forests, this assumption does not influence our findings.

2.6 Implications

2.6.1 Key Findings

Drawing from the theoretical tenets of forms of social capital (Calderon Gomez, 2021; Gilbert et al., 2022; Safiollah et al., 2009), we conceptualize and empirically test the relationship between county-level telehealth adoption and county-level mental health well-being. We establish the ATE as well as the HTEs. Building on burgeoning work in information systems and public health (Baird et al., 2022; Barney et al., 2011; Byrow et al., 2020; Greenwood & Wattal, 2017; T. Guo et al., 2021), we exploit the phased rollout of telehealth adoption for approximately 2,800 counties in the United States using consolidated archival data from 2017 to 2021. Results indicate a significant decline in community mentally unhealthy days, following the adoption of telehealth services. This findings extend our understanding on the positive impact of telehealth services on individual level mental health from specific disease such as depression (Barney et al., 2011) and PTSD (Morres et al., 2019) to overall community mental health well-being. Overall, the finding suggests that adoption in a community of telehealth, as a novel form of digital capital that virtualizes mental health care services, can improve the community's mental well-being.

Additionally, empirical evidence from the causal forests also suggests that the impact of telehealth services is heterogeneous across counties with different forms of capital. The main findings from the causal forest analysis are as follows.

First, our finding that broadband access itself does not lead to mental health outcome improvement is consistent with existing evidence (Barney et al., 2011; Henrich et al., 2001). In addition, we found that for counties with high digital access rates, the role of telemedicine plays a smaller role in improving people's mental health outcomes, compared to those counties with relatively low digital access rates. This is likely because digital access itself is a double-edged sword, which on the one hand provides access to health care resources while on the other hand causes unnecessary comparison, peer pressures, and over-consumption, worsening mental health outcomes. Therefore, having a high digital access rate alone does not amplify the positive impact of telehealth on mental health.

Second, we found when digital capital is intertwined with other forms of social capital, there are synergistic effects for beneficial mental health outcomes, echoing some existing studies (Allen et al., 2014; Pinxten & Lievens, 2014). Specifically, digital infrastructure measured by broadband access rate is a key effect modifier. Counterintuitively, our study reveals community's (county's) with low broadband access rates benefit more from telehealth adoptions than their counterparts with high broadband access rates do. This is likely because only basic digital infrastructure is necessary to unleash the full potential of telehealth so that the influence of strong digital infrastructure (measured by high broadband access rates) on telehealth's impact is constrained. We reinforced our findings by observing that, within counties with high bandwidth, those with slower internet speed derived greater benefits compared to counterparts with faster internet. The availability of high-speed internet in communities allows access and utilization of high-bandwidth digital resources such as social media, potentially exerting a negative impact on mental health (Braghieri et al., 2022). For counties with limited digital infrastructure, telehealth can be advantageous in communities with low social capital but high cultural capital (measured by

English-speaking rate) or economic capital (measured by unemployment rate), which are consistent with the existing research (Marazziti et al., 2021; Sentell et al., 2007). These findings extend our understanding on how different forms of capital underscore the importance of assessing heterogeneous effect of telehealth adoption on community mental health well-being because examining only ATEs may mask important heterogeneity and result in the misleading conclusion that the effect is distributed evenly.

2.6.2 Implications for Theory and Research

This paper contributes to both the literature and practice on HIT. By bridging the literature on telehealth and the literature on mental health, our study offers solid evidence for the promise of telehealth adoption on community (county) mental health well-being, which is not readily available from the existing literature. Our findings also open a new channel on how telehealth adoption contributes to mental health well-being. While previous studies suggest that telehealth could improve individuals' mental well-being (Connolly et al., 2020), our paper shows that telehealth achieves greater effect when combined with other forms of capital. We believe that such findings are critical for community mental health due to the intertwined socioeconomic factors leading to rising challenges of mental well-being.

Specifically, in the context of mental health and telehealth adoptions, the contribution to theory highlights the importance of considering multiple dimensions of capital. It emphasizes that the benefits of telehealth in improving community's mental well-being are enhanced when combined with digital infrastructure and economic, cultural, and social capital. This perspective challenges a narrow focus on digital technology (Torous et al., 2020) alone and recognizes the broader social and economic factors that shape the effectiveness and reach of telehealth interventions.

The contribution to theory also highlights the importance of addressing disparities and promoting equity (Gilbert et al., 2022). Whereas prior work provides evidence of how telehealth under a defined condition (e.g., during the COVID-19 pandemic) can reduce disparities in mental health outcomes (Qian et al., 2021), we explore the impact of telehealth adoption intertwining different forms of capital to enhance community mental health conditions. Further, our research provides valuable insights and has important implications for narrowing mental health disparities within and between counties.

Methodologically, ours is, to the best of our knowledge, one of the first articles to address this problem where we use both a DID approach to identify ATEs and causal ML to explore HTEs. We obtained new insights that a purely deductive approach—a commonly used method examining a single source of heterogeneity at a time—could have missed. The counties that benefit the most from telehealth adoptions tended to report lower unemployment rates, lower broadband access at baseline, and more mental health providers. In prior work examining heterogeneity deductively, higher SES such as higher internet access alone is typically linked to resilience to people's health in general. Our ML-based inductive approach for effect heterogeneity allowed interactions between multiple characteristics and revealed that higher internet access rate could backfire and make people unhealthier when coupled with other adverse backgrounds such as economic hardship and education problems such as not speaking English well.

2.6.3 Implications for Practice and Public Policy

The contributions and implications for practice and public policy stem from recognizing the significance of digital capital, economic capital, and other forms of social capital in shaping outcomes and informing strategies for improving mental health and health care access. Here are some key contributions and implications.

Understanding the interplay between different forms of capital can inform the design and implementation of telehealth strategies. Practitioners can leverage digital resources while considering economic and social factors to ensure the effectiveness and reach of telehealth interventions. This may involve tailoring services to specific cultural contexts, providing financial support for individuals with limited economic resources, and fostering community engagement to promote telehealth adoption.

Recognizing the impact of different forms of capital helps identify and address health disparities. Public policies can focus on reducing barriers to access for underserved populations, particularly those experiencing poverty. This may involve initiatives such as subsidized internet connectivity, increased funding for telehealth infrastructure in disadvantaged areas, and targeted outreach programs to promote awareness and utilization of telehealth services among marginalized communities.

This study provides evidence on the HTEs of telehealth services that can assist policy makers and health care stakeholders in scoring and targeting customers and in allocating resources across individual counties when designing policy or activity associated with telehealth on mental health.

2.6.4 Implications for Future Work

It is important to note that this research is subject to some limitations. First, we are unable to observe telehealth adoption at the more granular level such as at the zip code, street, or individual level. Our work fills the gap of addressing the effect of telehealth adoptions on community-level mental health status. However, additional work with more granular level data is necessary if we want to understand heterogeneity better at lower levels of analysis.

Second, the exclusion of telehealth services offered by private practices in our study stems from data limitations. Conventional wisdom suggests that patients from vulnerable populations tend to opt for public hospitals rather than private practices. Therefore, it is essential to underscore that our research findings on uncovering HTE of telehealth on mental health to address mental health disparities likely remain unaffected. Nevertheless, we recommend that future researchers consider incorporating telehealth services provided by private practices to facilitate comparative analysis and enrich the scope of our findings.

Third, while our study has demonstrated the significant impact of telehealth adoption on improving mental health in counties, and we have identified heterogeneity in the effects across different counties, we have not considered the efficiency or clinical quality of such services. Future researchers should explore strategies to optimize the efficiency and quality of telehealth interventions and improve their impact on mental health in diverse county settings.

2.7 Conclusion

This study contributes to the theoretical understanding and empirical investigation of the intricate relationship between digital capital—telehealth and associated digital infrastructure—and overall community mental health status. Mental health is a significant concern for public health and society, attracting growing attention in medical and population health research. Nonetheless, these important matters have received limited attention within the IS literature. In previous studies, researchers have mainly discussed the social and health effects of digital technologies in the context of the spread of public health problems (Bourdieu, 2011a; Lorant et al., 2003) and the contagious effect of lifestyle decisions (Alegría et al., 2013). We hope that our study ushers in an important area of IS research: the potential of telehealth to improve community mental health. Our study highlights the positive impact technology can have on community mental health. We specifically examine the role of telehealth adoptions as a form of technology. Furthermore, we highlight how contextual combinations of different forms of capital could dynamically influence the benefits of telehealth adoptions on community mental health outcomes. We hope this research sparks a new intellectual discussion around technology and mental health in the IS field and hopefully in other literature. In this way, we hope to gain a deeper understanding of whether, how, and why the use of digital technologies can gain positive effects on community mental health.

References

Abadie, A., Drukker, D., Herr, J.L., & Imbens, G.W. (2004). Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal*, *4*(3), 290–311.

Adler-Milstein, J., Kvedar, J., & Bates, D.W. (2014). Telehealth among US Hospitals: Several Factors, Including State Reimbursement and Licensure Policies, Influence Adoption. *Health Affairs*, *33*(2), 207–215.

Agarwal, R., Animesh, A., & Prasad, K. (2009). Research note—Social Interactions and the "Digital Divide": Explaining Variations in Internet Use. *Information Systems Research*, *20*(2), 277–294.

Alegría, M., Fortuna, L. R., Lin, J. Y., Norris, L. F., Gao, S., Takeuchi, D. T., Jackson, J. S., Shrout, P. E., & Valentine, A. (2013). Prevalence, Risk, and Correlates of Posttraumatic Stress Disorder across Ethnic and Racial Minority Groups in the US. *Medical Care*, *51*(12), 1114.

Alessie, R., Lusardi, A., & Aldershof, T. (1997) Income and Wealth over the Life Cycle Evidence from Panel Data. *Review of Income and Wealth*, *43*(1), 1–32.

Allen, J., Balfour, R., Bell, R., & Marmot, M. (2014) Social Determinants of Mental Health. *International Review of Psychiatry, 26*(4), 392–407.

Angrist, J.D., & Pischke, J.S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (pp. 27-112). Princeton University Press.

Aral, S. & Nicolaides, C. (2017). Exercise Contagion in a Global Social Network. *Nature Communications, 8*(1), 14753.

Ashwood, J.S., Mehrotra, A., Cowling, D., & Uscher-Pines, L. (2017). Direct-to-Consumer Telehealth May Increase Access to Care but Does Not Decrease Spending. *Health Affairs, 36*(3), 485–491.

Athey, S., & Imbens, G. (2016). Recursive Partitioning for Heterogeneous Causal Effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., & Imbens, G.W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized Random Forests. *The Annals of Statistics*, *47*(2), 1148 – 1178.

Athey, S., & Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies*, *5*(2) 37-51.

Austin, P.C. (2009). Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity‐Score Matched Samples. *Statistics in Medicine*, *28*(25), 3083–3107.

Autor, D.H. (2003). Outsourcing at will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, *21*(1), 1–42.

Ayabakan, S., Bardhan, I.R., & Zheng, Z. (2023). Impact of Telehealth and Process Virtualization on Healthcare Utilization. *Information Systems Research*, *35*(1), 45-65.

Baird, A., Cheng, Y., & Xia, Y. (2022). Telehealth Adoption and Discontinuation by US Hospitals: Results From 2 Quasi-Natural Experiments. *JMIR Formative Research*, *6*(2), e28979.

Barney, L.J., Griffiths, K.M., & Banfield, (2011). M.A. Explicit and Implicit Information Needs of People with Depression: A Qualitative Investigation of Problems Reported on An Online Depression Support Forum. *BMC psychiatry*, *11*(1), 88.

Bashshur, R.L., Shannon, G.W., Bashshur, N., & Yellowlees, P.M. (2016). The Empirical Evidence for Telemedicine Interventions in Mental Disorders. *Telemedicine and E-Health*, 22(2), 87–113.

Bassett, H., & Lloyd, C. (2001). Occupational Therapy in Mental Health: Managing Stress and Burnout. *British Journal of Occupational Therapy*, *64*(8), 406–411.

Bond, F.W., & Bunce, D. (2003). The Role of Acceptance and Job Control in Mental Health, Job Satisfaction, and Work Performance. *Journal of Applied Psychology*, *88*(6), 1057.

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press.

Bourdieu, P. (1991). *Language and symbolic power*. Harvard university press.

Bourdieu, P. (1986). The Forms of Capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.

Bourdieu, P. (2018). The Forms of Capital. In M. Granovetter (Ed.), *The Sociology of Economic Life* (pp. 78-92). Routledge.

Bourdieu, P. (2018). Cultural Reproduction and Social Reproduction. In R. Brown (Ed.), *Knowledge, Education, and Cultural Change* (pp. 71–112). Routledge.

Braghieri, L., Levy, R., & Makarin, A. (2022). Social Media and Mental Health. *American Economic Review*, *112*(11), 3660–3693.

Breck, A., & Wakar, B. (2021). Methods, Challenges, and Best Practices for Conducting Subgroup Analysis (OPRE Report No. 2021-17). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.gov/sites/default/files/documents/opre/methods-challenges-best-practices-jan-2021.pdf

Byrow, Y., Pajak, R., Specker, P., & Nickerson, A. (2020). Perceptions of Mental Health and Perceived Barriers to Mental Health Help-Seeking amongst Refugees: A Systematic Review. *Clinical Psychology Review*, *75*, 101812.

Calderon Gomez, D. (2021). The Third Digital Divide and Bourdieu: Bidirectional Conversion of Economic, Cultural, and Social Capital to (and from) Digital Capital among Young People in Madrid. *New Media & Society*, *23*(9), 2534–2553.

Chellaiyan, V.G., Nirupama, A.Y., & Taneja, N. (2019). Telemedicine in India: Where Do We Stand? *Journal of Family Medicine and Primary Care*, *8*(6), 1872.

Cianconi, P., Betrò, S., & Janiri, L. (2020). The Impact of Climate Change on Mental Health: A Systematic Descriptive Review. *Frontiers in Psychiatry*, *11*, 490206.

Cloninger, C.R. (1999). A New Conceptual Paradigm from Genetics and Psychobiology for the Science of Mental Health. *Australian & New Zealand Journal of Psychiatry*, *33*(2), 174–186.

Compton, M.T., & Shim, R.S. (2015). The Social Determinants of Mental Health. *Focus*, *13*(4), 419–425.

Connolly, S.L., Miller, C.J., Lindsay, J.A., & Bauer, M.S. (2020), A Systematic Review of Providers' Attitudes toward Telemental Health via Videoconferencing. *Clinical Psychology: Science and Practice*, *27*(2), e12311.

Cummings, J.R., Allen, L., Clennon, J., Ji, X., & Druss, B.G. (2017) Geographic Access to Specialty Mental Health Care across High-and Low-Income US Communities. *JAMA Psychiatry*, *74*(5), 476–484.

Diaz, A., Baweja, R., Bonatakis, J.K., & Baweja, R. (2021). Global Health Disparities in Vulnerable Populations of Psychiatric Patients during the COVID-19 Pandemic. *World Journal of Psychiatry*,*11*(4), 94.

Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B. G., Skodol, A. E., & Stueve, A. (1992). Socioeconomic Status and Psychiatric Disorders: The Causation-Selection Issue. *Science*, *255*(5047), 946–952.

Ehsan, A., Klaas, H.S., Bastianen, A., & Spini, D. (2019). Social Capital and Health: A Systematic Review of Systematic Reviews. *SSM-Population Health*, *8*, 100425.

Faccio, M. & McConnell, J.J. (2020). Death by Pokémon GO: The Economic and Human Cost of Using Apps while Driving. *Journal of Risk and Insurance*, *87*(3), 815–849.

Felten, P., & Lambert, L. M. (2020). *Relationship-Rich Education: How Human Connections Drive Success in College*. JHU Press.

Gilbert, K.L., Ransome, Y., Dean, L.T., DeCaille, J., & Kawachi, I. (2022). Social Capital, Black Social Mobility, and Health Disparities. *Annual Review of Public Health*, *43*(1), 173–191.

Girdhar, R., Srivastava, V., & Sethi, S. (2020). Managing Mental Health Issues among Elderly during COVID-19 Pandemic. *Journal of Geriatric Care and Research*, *7*(1), 32–35.

Goh, J.M., Gao, G., & Agarwal, R. (2016) The Creation of Social Value. *MIS Quarterly*, *40* (1), 247–264.

Greenwood, B.N., & Wattal, S. (2017). Show Me the Way to Go Home. *MIS quarterly*, *41*(1), 163–188.

Griner, D., & Smith, T.B. (2006). Culturally Adapted Mental Health Intervention: A Meta-Analytic Review. *Psychotherapy: Theory, Research, Practice, Training*, *43*(4), 531.

Guo, T., Sriram, S., & Manchanda, P. (2021). The Effect of Information Disclosure on Industry Payments to Physicians. *Journal of Marketing Research*, *58*(1), 115–140.

Haleem, A., Javaid, M., Singh, R.P., & Suman, R. (2021). Telemedicine for Healthcare: Capabilities, Features, Barriers, and Applications. *Sensors International*, *2*, 100117.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica: Journal of the Econometric Society*,153–161.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review*, *91*(2), 73–78.

Hero, R.E. (2007). *Racial Diversity and Social Capital: Equality and Community in America*. Cambridge University Press.

Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., & Yellowlees, P. M. (2013). The Effectiveness of Telemental Health: A 2013 Review. *Telemedicine and E-Health*, *19*(6), 444–454.

Holt-Lunstad, J., Smith, T.B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review. *Perspectives on Psychological Science*, 10(2), 227–237.

Hsieh, J.P.-A., Rai, A., & Keil, M. (2011). Addressing Digital Inequality for the Socioeconomically Disadvantaged through Government Initiatives: Forms of Capital that Affect ICT Utilization. *Information Systems Research*, *22*(2), 233–253.

Hubley, S., Lynch, S.B., Schneck, C., Thomas, M., & Shore, J. (2016). Review of Key Telepsychiatry Outcomes. *World Journal of Psychiatry*, *6*(2) 269.

Iacus, S.M., King, G., &Porro, G. (2011). Multivariate Matching Methods that are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, *106*(493), 345–361.

Imbens, G.W. &Rubin, D.B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Iyengar, R., Park, Y.-H., &Yu, Q. (2022). The Impact of Subscription Programs on Customer Purchases. *Journal of Marketing Research*, *59*(6), 1101–1119.

Jeannotte, M.S. (2003). Singing Alone? The Contribution of Cultural Capital to Social Cohesion and Sustainable Communities. *The International Journal of Cultural Policy*, *9*(1), 35–49.

Kämpfen, F., Kohler, I.V., Ciancio, A., Bruine de Bruin, W., Maurer, J., & Kohler, H.-P. (2020). Predictors of Mental Health during the Covid-19 Pandemic in the US: Role of Economic Concerns, Health Worries and Social Distancing. *PloS one*, *15*(11), e0241895.

Karahanna, E., Chen, A., Liu, Q.B., & Serrano, C. (2019). Capitalizing on Health Information Technology to Enable Advantage in U.S. Hospitals. *MIS Quarterly*, *43*(1), 113–140.

Kawachi, I., & Berkman, L.F. (2001). Social Ties and Mental Health. *Journal of Urban Health*,*78*(3), 458–467.

Kirmayer, L. J., Narasiah, L., Munoz, M., Rashid, M., Ryder, A. G., Guzder, J., Hassan, G., Rousseau, C., & Pottie, K. (2011). Common Mental Health Problems in Immigrants and Refugees: General Approach in Primary Care. *Cmaj*, *183*(12), E959–E967.

Knapp, M., McDaid, D., & Parsonage, M. (2011). *Mental Health Promotion and Mental Illness Prevention: The Economic Case*. Department of Health, London, United Kingdom. https://www.gov.uk/government/publications/mental-health-promotion-and-mental-illness-prevention-the-economic-case

Kyung, N., Lim, S., & Lee, B. (2017). A Depressing Internet Tale: Empirical Analysis of the Internet's Impact on Suicide. *In Proceedings of the 38th International Conference on Information Systems (ICIS)*. Association for Information Systems.

Lennox, C.S., Francis, J.R., &Wang, Z. (2012). Selection Models in Accounting Research. *The Accounting Review*, *87*(2), 589–616.

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic Inequalities in Depression: A Meta-Analysis. *American Journal of Epidemiology*, *157*(2), 98–112.

Lund, C., De Silva, M., Plagerson, S., Cooper, S., Chisholm, D., Das, J., Knapp, M., & Patel, V. (2011). Poverty and Mental disorders: Breaking the Cycle in Low-Income and Middle-Income Countries. *The Lancet*, *378*(9801), 1502–1514.

Marazziti, D., Avella, M. T., Mucci, N., Della Vecchia, A., Ivaldi, T., Palermo, S., & Mucci, F. (2021). Impact of economic crisis on mental health: A 10-year challenge. *CNS Spectrums*, *26*(1), 7–13.

Mohnen, S. M., Völker, B., Flap, H., & Groenewegen, P. P. (2012). Health-related behavior as a mechanism behind the relationship between neighborhood social capital and individual health-a multilevel analysis. *BMC Public Health*, *12*(1), 1–12.

Morres, I. D., Hatzigeorgiadis, A., Stathi, A., Comoutos, N., Arpin-Cribbie, C., Krommidas, C., & Theodorakis, Y. (2019). Aerobic Exercise for Adult Patients with Major Depressive Disorder in Mental Health Services: A Systematic Review and Meta-Analysis. *Depression and Anxiety*, *36*(1), 39–53.

Nyman, J.A., Barleen, N.A., Dowd, B.E., Russell, D.W., Coons, S.J., & Sullivan, P.W. (2007). Quality-of-Life Weights for the US Population: Self-Reported Health Status and Priority Health Conditions, by Demographic Characteristics. *Medical Care*, *45*(7), 618–628.

Ohrnberger, J., Fichera, E., & Sutton, M. (2017). The Relationship between Physical and Mental Health: A Mediation Analysis. *Social Science & Medicine*,*195*,42–49.

Overby, E. (2008). Process Virtualization Theory and the Impact of Information Technology. *Organization Science*, *19*(2), 277–291.

Overby, E., Slaughter, S.A., & Konsynski, B. (2010). Research Commentary—the Design, Use, and Consequences of Virtual Processes. *Information Systems Research*, 21(4), 700–710.

Paccoud, I., Nazroo, J., & Leist, A. (2020). A Bourdieusian Approach to Class-related Inequalities: the Role of Capitals and Capital Structure in the Utilisation of Healthcare Services in Later Life. *Sociology of Health & Illness*, *42*(3), 510–525.

Pilgrim, D., Rogers, A., & Bentall, R. (2009). The Centrality of Personal Relationships in the Creation and Amelioration of Mental Health Problems: The Current interdisciplinary Case. *Health*, *13*(2), 235–254.

Pinxten, W., & Lievens, J. (2014). The Importance of Economic, Social and Cultural Capital in Understanding Health Inequalities: using a Bourdieu-based Approach in Research on Physical and Mental Health Perceptions. *Sociology of Health & Illness*, *36*(7), 1095–1110.

Pirkkalainen, H., Salo, M., Tarafdar, M., & Makkonen, M. (2019). Deliberate or Instinctive? Proactive and Reactive Coping for Technostress. *Journal of Management Information Systems*, *36*(4), 1179–1212.

Polinski, J.M., Barker, T., Gagliano, N., Sussman, A., Brennan, T.A., & Shrank, W.H. (2016). Patients' Satisfaction with and Preference for Telehealth Visits. *Journal of General Internal Medicine*, *31*(3), 269–275.

Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No Health without Mental Health. *The Lancet*, *370*(9590), 859-877.

Qian, F., Hastings, J.F., & Ahmed, R. Overcoming Telemental Health Disparities during the COVID-19 Pandemic. (2021). *International Journal for Quality in Health Care*, *33*(4), mzab127.

Ragnedda, M. (2018). Conceptualizing Digital Capital. *Telematics and Informatics*, *35*(8), 2366–2375.

Reinert, M., Fritze, D., & Nguyen, T. (2024). *The State of Mental Health in America 2024*. Mental Health America. https://mhanational.org/research/state-of-mental-health-in-america-2024/

Safiollah, S., Alimohammad, J., & Maryam, J. (2009). Effect of the Internet on Social Capital. *International Journal of Humanities and Social Sciences*, *3*(12), 2133–2140.

Schriger, S. H., Klein, M. R., Last, B. S., Fernandez-Marcote, S., Dallard, N., Jones, B., & Beidas, R. S. (2022). Community Mental Health Clinicians' Perspectives on Telehealth during the COVID-19 Pandemic: Mixed Methods Study. *JMIR Pediatrics and Parenting*, *5*(1), e29250.

Sentell, T., Shumway, M., & Snowden, L. (2007). Access to Mental Health Treatment by English Language Proficiency and Race/Ethnicity. *Journal of General Internal Medicine*, *22*(Suppl2), 289–293.

Silberholz, E.A., Brodie, N., Spector, N.D., & Pattishall, A.E. (2017). Disparities in Access to Care in Marginalized Populations. *Current Opinion in Pediatrics*, *29*(6), 718–727.

Snoswell, C.L., Taylor, M.L., Comans, T.A., Smith, A.C., Gray, L.C., & Caffery, L.J. (2020). Determining if Telehealth can Reduce Health System Costs: Scoping Review. *Journal of Medical Internet Research*, *22*(10), e17298.

Steers, L.-L.N., Robert E., W., & Acitelli, L.K. (2014). Seeing Everyone Else's Highlight Reels: How Facebook Usage is linked to Depressive Symptoms. *Journal of Social and Clinical Psychology*, *33*(8), 701–731.

Stewart, C., Konstantellou, A., Kassamali, F., McLaughlin, N., Cutinha, D., Bryant-Waugh, R., Simic, M., Eisler, I., & Baudinet, J. (2021). Is This the 'New Normal'? A Mixed Method Investigation of Young Person, Parent and Clinician Experience of Online Eating Disorder Treatment during the COVID-19 Pandemic. *Journal of Eating Disorders*, *9*(1), 1–11.

Stuart, E.A. Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25, 1 (2010), 1. Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. *JMIR Mental Health*, *7*(3), e18848.

Umberson, D., & Karas Montez, J. (2010). Social Relationships and Health: A Flashpoint for Health Policy. *Journal of Health and Social Behavior*, *51*(1_suppl), S54-S66.

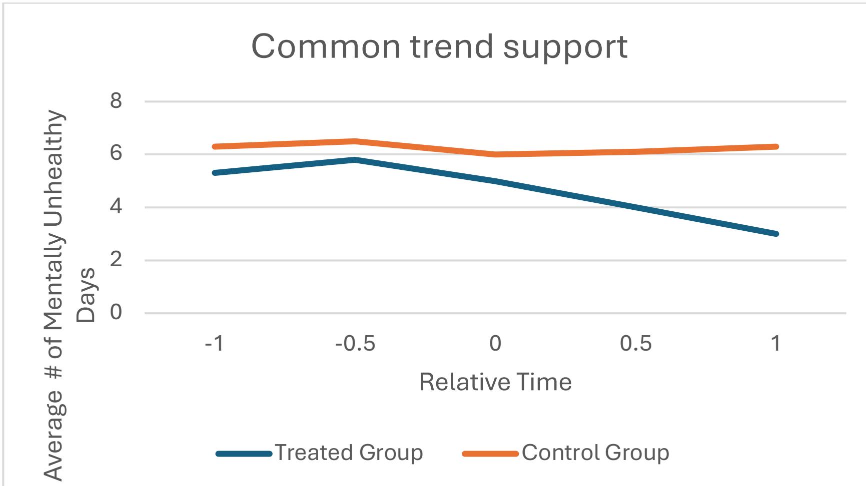
Yu, Y., & Williams, D. R. (1999). Socioeconomic Status and Mental Health. In C. S. Aneshensel & J. C. Phelan (Eds.), *Handbook of the Sociology of Mental Health* (pp. 151–166). Springer.

Zarse, E. M., Neff, M. R., Yoder, R., Hulvershorn, L., Chambers, J. E., & Chambers, R. A. (2019). The Adverse Childhood Experiences Questionnaire: Two Decades of Research on Childhood Rrauma as a Primary Cause of Adult Mental Illness, Addiction, and Medical Diseases. *Cogent Medicine*, *6*(1), 1581447.

Appendix

Figure 1

Common Trend Support for Average Number of Mentally Unhealthy Days



Chapter 3: Give a Man a Mask to Tell You the Truth: Anonymity and User Engagement: Sensitive Information Exchange in Mental Health Context

3.1 Introduction

Mental health is a critical component of overall health and well-being. According to the World Health Organization (WHO), one in eight people worldwide lives with a mental illness.[7] Mental health conditions affect an estimated one in seven children and adolescents aged 10 to 19 years.[8] In the United States, an estimated more than one in five adults live with a mental illness (59.3 million in 2022; 23.1% of the U.S. adult population).[9]

Community-based platforms such as social networking services (e.g., Facebook, Instagram) and knowledge exchange websites (e.g., Quora, Stack Overflow) serve as vital venues for individuals seeking support, information, and social connection. This is especially true for those dealing with sensitive personal issues, such as mental health challenges, where traditional resources such as therapy may be inaccessible or stigmatized. Recent studies show that 84% of internet users have looked online for health information, and 41% have read or watched someone else's health experience on social media platforms (Chen & Wang, 2021; Farsi, 2021; Pew Research Center, 2024). The COVID-19 pandemic accelerated this trend, with health-related social media engagement increasing by 70% between 2019 and 2022 (Xin & Lim, 2024) Social media have emerged as particularly critical platforms for health information interaction, serving over million users globally who turn to them for support with sensitive health conditions (Wang &

---

[7] Global Mental Illness Statistics: https://huntingtonpsych.com/the-latest-mental-health-statistics-what-the-numbers-say-about-the-state-of-our-minds-in-2024.

[8] https://www.who.int/news/item/09-10-2024-who-and-unicef-launch-guidance-to-improve-access-to-mental-health-care-for-children-and-young-people

[9] US Mental Illness Statistics 2025, National Institute of Health: https://www.nimh.nih.gov/health/statistics/mental-illness.

Cohen, 2023). Recent data indicates that people in the United States have shared health-related information online, mental health content representing the fastest-growing category of health discussions on social platforms (Miller, 2025).

A key determinant of the success and sustainability of these platforms is user participation. Prior research has highlighted several platform features that influence participation, including user ranking systems (Goes et al., 2016; Shen et al., 2015), symbolic incentives (Gallus, 2017), interactivity (Jiang et al., 2010), and social network integration (Huang et al., 2017). Among these, anonymity has emerged as a particularly powerful yet contentious feature. On the one hand, anonymity can reduce barriers to participation by alleviating fears of judgment or social stigma, making it especially valuable in health-related and mental health contexts (Bapna et al., 2016; Cho & Kwon, 2015). On the other hand, it may also pose challenges related to accountability and content quality. This sets up a critical tension between identity and anonymity —while identity cues can enhance trust and engagement, anonymity can offer psychological safety and openness, especially in communities where vulnerability and stigma are prevalent. This tension rises as online health discussions evolve into mental health-specific spaces, where the stakes of identity disclosure are significantly higher.

Major digital platforms have adopted diverse identity policies reflecting different philosophical approaches to balancing user safety, content quality, and engagement. These policies exist along a spectrum: LinkedIn and Facebook require real names and credential verification, assuming accountability enhances trust and content quality; platforms such as Medium and Twitch allow pseudonyms as middle-ground solutions; and Reddit embrace complete anonymity, arguing that identity masking enables more diverse expression (Haimson & Hoffmann, 2016).

Research has consistently demonstrated that anonymity facilitates self-disclosure across both offline and online contexts. The classic "stranger on a train" phenomenon (Rubin, 1975) revealed that individuals often share intimate details with anonymous strangers, challenging assumptions about the necessity of established relationships for meaningful disclosure. Digital environments amplify this effect through what Suler (2004) termed the "online disinhibition effect," where computer-mediated communication systematically reduces psychological barriers to self-expression. For individuals with stigmatized conditions, these dynamics become particularly salient. Social identity theory (Tajfel et al., 1979) suggests that people manage multiple identities across different contexts, while social stigma theory (Goffman, 2014; Link & Phelan, 2001) explains how individuals develop protective strategies when managing potentially damaging information about themselves. Online anonymity represents one such strategy, allowing access to peer support while maintaining protective boundaries around primary social identities. In addition, in online communities such as Quora, users often have the option to share information about their credentials besides identity, such as educational background, professional affiliation, or other personal attributes—collectively referred to as disclosed credentials. These credentials serve as visible indicators of a user's authority, expertise, or social standing within the platform. For instance, a user with a profile indicating they are a licensed therapist or hold a PhD in psychology may be perceived as more credible in mental health discussions than an anonymous user without such disclosures. In the context of this study, we use disclosed credentials as a moderator to examine how the effects of anonymity on user engagement and emotional expression vary depending on a user's perceived status or authority within the platform. Specifically, we explore whether users with fewer disclosed credentials—who may feel more socially vulnerable or less authoritative—experience greater increases in participation and openness when anonymity is introduced, compared to users with more established or identifiable profiles.

In most existing studies, researchers have examined transitions from anonymous to identified participation, focusing on how increased accountability affects content quality and engagement (Chan et al., 2023; Guo & Caine, 2021; Guo et al., 2021; Huang et al., 2017; Pu et al., 2020; Pu et al., 2024). Research on Facebook's real-name policy implementation, Google+'s identity requirements, and LinkedIn's professional verification systems has similarly focused on how increased accountability mechanisms affect user behavior, content quality, and community dynamics (Cho et al., 2012; Lingel & Boyd, 2013). This emphasis reflects platforms' general trend toward greater identity disclosure but leaves unexplored the potentially beneficial effects of transitions toward anonymity, particularly in sensitive health contexts where identity disclosure creates participation barriers. Existing literature has extensively documented the effects of transitioning from anonymous to identified participation across various online contexts. Researchers examining platforms such as Wikipedia, Reddit, and various news comment sections have consistently found that requiring identity disclosure tends to reduce overall participation while potentially improving content civility and reducing harmful behaviors (Huang et al., 2017; Pu et al., 2020; Resnick et al., 2017). This research reflects broader platform trends toward identity disclosure, driven by concerns about misinformation, harassment, and the commercial value of identifiable user data. Platforms increasingly view anonymity as a problem to be solved rather than a feature to be preserved, leading to policy changes that require progressively more identity verification. Consequently, most of the empirical researchers have followed these platform decisions, investigating how communities adapt when anonymity is removed or reduced.

However, this extensive focus on anonymity-to-identity transitions has created a significant research gap: we know remarkably little about what happens when platforms move in the opposite direction. The reverse perspective—examining how user-generated content changes when communities transition from identity disclosure requirements toward greater anonymity—remains largely unexplored, despite its critical importance for understanding the full spectrum of identity policy effects. This gap is particularly problematic in health contexts, where the stakes of identity disclosure differ fundamentally from those in general social media or news commentary settings. While requiring real names in political discussions might improve civility, the same requirement in mental health communities could exclude individuals facing employment discrimination, family rejection, or social stigma. The absence of research on anonymity-enabling transitions means that platforms lack empirical guidance when considering policies that might improve access for vulnerable populations.

The reverse-perspective investigation matters for several critical reasons. First, it challenges the fundamental assumption that effects of identity policies are symmetrical. Research has shown that psychological processes often exhibit asymmetries—gaining something and losing the same thing can produce different effects due to loss aversion, framing effects, and different cognitive pathways (Kahneman & Tversky, 1984). In the context of online identity, removing anonymity may activate different social and psychological mechanisms than granting it, particularly when stigmatized identities are involved. Understanding these potential asymmetries is essential for developing nuanced theoretical models of online identity management. Second, the reverse perspective offers practical insights for platform governance in an evolving digital landscape. As concerns about online harassment, doxxing, and privacy violations intensify, many platforms are reconsidering blanket identity disclosure requirements, particularly for sensitive topics. However, these policy discussions currently rely on theoretical speculation rather than empirical evidence. By systematically examining how communities respond when anonymity protections are introduced rather than removed, researchers provide actionable guidance for platform designers seeking to create safer, more inclusive digital health spaces without sacrificing the quality and credibility that make these communities valuable.

We examine two specific research questions that directly address the core tension:

1. What are the effects of transitioning from identity disclosure to anonymity on the quantity and quality of user-generated content in online mental health communities?
2. How do user credentials moderate the impacts of transitioning from identity disclosure to anonymity on content quality in online mental health communities?

The first research question is to challenge the assumption that anonymity necessarily degrades content while potentially revealing how identity protection affects participation barriers. And the second question acknowledges the heterogeneous nature of community participants and examines whether anonymity's effects vary across different types of contributors.

We conduct this research through a quasi-experimental design on Quora, a question-and-answer platform with over 400 million monthly users that has undergone significant evolution in its identity policies.[10] Quora presents an ideal natural shock for several reasons: its explicit focus on knowledge sharing creates clear metrics for content quality assessment, its structured Q&A format facilitates systematic analysis of information-seeking and information-provision behaviors, and its policy evolution regarding anonymity provides natural experimental conditions. Importantly, Quora exemplifies the central tension this research addresses. The platform must

[10] Reach over 400 million monthly unique visitors on Quora: https://business.quora.com/resources/reach-over-400-million-monthly-unique-visitors-on-quora/.

balance its reputation for expert-driven, credible content with the need to provide accessible spaces for individuals seeking peer support around stigmatized conditions. Quora's initial emphasis on real-name policies and credential verification has gradually shifted toward allowing total anonymity, creating an opportunity to examine how these changes affect community dynamics and content characteristics.

This research contributes to our understanding of digital health community design by providing empirical evidence about the relationship between identity policies and community outcomes in sensitive health contexts. Rather than accepting assumed trade-offs between access and quality, we examine whether alternative approaches to identity management might resolve rather than merely navigate the central tension. The findings have implications extending beyond platform design to public health policy regarding digital health resources, health equity considerations in technology design, and the broader question of how digital technologies can be leveraged to reduce rather than reproduce existing health disparities. By focusing specifically on the transition toward anonymity in mental health contexts, we address a critical gap in our understanding of how to design inclusive digital health infrastructure that serves both individual and community needs.

3.2 Literature Review

Our research draws on two streams of literature: identity disclosure and online user behavior.

3.2.1 Identity Disclosure in Online Environments

Identity disclosure represents a fundamental aspect of human social interaction, encompassing the voluntary revelation of personal information, thoughts, feelings, and experiences to others (Pu et al., 2020). Researchers have defined identity disclosure as "what individuals verbally reveal about themselves to others (including thoughts, feelings, and experiences)" (Kim & Dindia, 2011). Building on foundational theories such as Goffman's (1959) dramaturgical perspective on self-presentation and Suler's (2004) online disinhibition effect, recent research has expanded our understanding of identity disclosure in online settings. Specifically, Balani and De Choudhury (2015) focused on explaining the decision to reveal a secret or divulge that one has an illness, such as HIV. Individuals who perceive their illness to be stigmatized may find it challenging to engage in self-disclosure. Zhou et al. (2025) showed that identity-related roles in self-governance (e.g., community arbitrators) create new layers of disclosure. While not requiring real-name identification, participation in moderation confers status and visibility, altering behavioral incentives. Similarly, Wang et al. (2021) explored how identity signals influence peer feedback dynamics, finding that identifiability enhances accountability and social learning but may also produce selective disclosure. These insights align with broader studies in mass communication and information systems, such as those by Rains (2007, 2014), who showed that identity cues and source credibility significantly shape perceptions of online health information. In academic Q&A forums, identifiability can simultaneously facilitate trust and inhibit controversial contributions, depending on topic sensitivity (Lee et al., 2014; Wang et al., 2013). Across these works, a consensus emerges: identity disclosure is neither inherently beneficial nor detrimental. Its effects depend on the platform architecture, user role, and topic domain.

Researchers have long been interested in social technologies for online anonymity. People frequently use anonymity to mitigate social risks and manage complicated social boundaries. The concept of anonymity online is best understood as a spectrum involving varying degrees of identifiability (Andalibi et al., 2016; Van der Nagel & Frith, 2015). A growing body of research highlights the double-sided effects of anonymity—defined as the absence or masking of identifiable user information—on online communication. Prior studies have emphasized the advantages of anonymity: it promotes more candid and open self-expression and makes self-disclosure and social support easier (Deng et al., 2021; Guo & Caine, 2021), particularly when discussing delicate and taboo subjects such as socioeconomic difficulties associated with poverty (Rho et al., 2017) as well as revelations of sexual assault disclosure (Andalibi et al., 2016) and pregnancy loss (Andalibi, 2019). This safe space contributes to a sense of psychological relief and social connection, even in the absence of physical proximity. Similarly, in political and civic forums, anonymous participation can reduce the influence of social hierarchies and promote more equitable dialogue (Zhou et al., 2025). Existing research also shows that, under certain conditions, anonymity can heighten group identity salience and enhance norm-conforming behaviors (Deng et al., 2021). Anonymity is an essential factor in designing social technologies that facilitate support seeking and provision in socially stigmatized contexts (Andalibi, 2019; Andalibi et al., 2016). However, anonymity can reduce accountability, lower perceived source credibility, and increase the likelihood of incivility and antisocial behavior. Wang (2020) found that anonymous commenters on news websites were perceived as less credible and their messages less persuasive, particularly when incivility was present. These findings align with earlier work by Rains (2007), which showed that a lack of identity cues diminishes perceptions of trustworthiness and social presence. Deindividuation theory (Reicher et al., 1995; Suler, 2004) further suggests that reduced identifiability can lead to disinhibited behavior, such as flaming, trolling (Chan et al., 2023; Guo & Caine, 2021), cyberbullying (Chan et al., 2023; Law et al., 2012) and other forms of norm violation. Additionally, researchers suggested that anonymity can make it difficult for people to establish trust or get credit for one’s contributions in groups and thus hinder online community building (Gupta et al., 2020).

Overall, internet users are supportive of anonymity: 59% of internet users say that they should be able to use the internet anonymously (Ireland, 2024). However, many online platforms are implementing increasingly strict name verification policies. These policies often require users to provide real names and integrate their accounts with established social networks such as Facebook or Google, effectively reducing anonymity. In addition, prior researchers have heavily focused on demonstrating the inhibition effect of real identity disclosure in digital platforms (Huang et al., 2017; Pu et al., 2020). We found no systematic study of the inhibition effect of real identity masking: how the transition from real name to anonymity impacts user behavior, participation levels, and content sharing patterns. Given the contrasting theoretical possibilities of anonymity—that it may either encourage authentic expression and participation by reducing social inhibition or conversely lead to decreased engagement and lower-quality content due to reduced accountability—it remains empirically unclear how much users value their online identity masking in mental health context. In addition, our emphasis on user behaviors when identity was masked from real name extends the focus beyond the predominantly identity disclosure behaviors examined in extant studies (see Table 1). This is meaningful because it helps the literature cover a fuller spectrum of user participation dynamics, addressing the understudied area of how identity concealment—as opposed to identity revelation—affects engagement in online community-based platforms

Despite extensive research on online behavior, no researcher has systematically examined how transitioning from real names to anonymity affects user participation, content sharing, and engagement patterns. Given competing theoretical predictions that anonymity either enhances authentic participation through reduced inhibition or diminishes engagement through reduced accountability, the actual behavioral effects of identity masking remain unclear, especially in sensitive contexts such as mental health where anonymity may be particularly important.

Our focus on behaviors under anonymity extends beyond existing studies in which researchers primarily examined voluntary identity disclosure (see Table 1). We address a significant literature gap by investigating how identity concealment—rather than revelation—affects user engagement, help the literature cover a fuller spectrum of user participation dynamics, and address the understudied area of how identity concealment—as opposed to identity revelation—affects engagement in online community-based platforms.

Table 1

Example Studies on Identity and User Behaviors

Study	**Research Context**	**Identity Strategy**	**Posting vs "Liking"**
Fredheim et al. 2015	Huffington Post	Anonymity to identity disclosure	Posting
Paskuda and Lewkowicz 2017	Quora	Anonymity to identity disclosure	Posting
Huang et al. 2017	Yelp and TripAdvisor	Anonymity to identity disclosure	Posting
Brown et al., 2018	Reddit	Anonymity to identity disclosure	Posting
Pu et al., 2020	Corporate Online Community	Anonymity to identity disclosure	Posting
Guo and Caine, 2021	Yahoo Answers	Anonymity to identity disclosure	Posting
Chan et al., 2023	Facebook	Anonymity to identity disclosure	Posting
Qian et al., 2024	Douban	Platform’s anonymous policy	Liking
This study	Quora	Platform’s anonymous policy	Posting

3.2.2 User Engagement in Online Settings

User engagement in Q&A sites is commonly measured through interactivity, including core activities such as asking, answering, following, commenting, upvoting, and downvoting (Li et al., 2012; Qian et al., 2024). Differences in engagement may also stem from identity disclosure policies. For instance, two news sites saw fewer comments after disabling anonymity and requiring Facebook login (Fredheim et al., 2015; Guo & Caine, 2021; Huang et al., 2017; Pu et al., 2020). These studies compared engagement when the platform policies transited from anonymity to identity disclosure.

Existing literature evaluates and predicts information quality on Q&A sites (Li et al., 2015; Ponzanelli et al., 2014). The most relevant predictors of quality include the price required for an expert to post an answer (Song et al., 2024), length of answer (Lou et al., 2025), language characteristics (Huang et al., 2017), upvotes (Graham & Rodriguez, 2021;Wang et al., 2013), downvotes (Ponzanelli et al., 2014), and factual information (Li et al., 2015). Information quality may also be associated with different types of identity. For example, after sites removed the option to engage anonymously (i.e., the identity model changed from allowing anonymity to disallowing anonymity), comment quality improved on online news sites (Fredheim et al., 2015) and online corporate communities (Pu et al., 2020).

Identity, topic sensitivity, and user engagement are related in online communities such as Q&A sites. When people use their real names to pose questions on social networks to their friends, family, and colleagues, they tend to provide answers to less sensitive questions (Hart et al., 2021; Moriña, 2021). In contrast, when people ask and answer questions to strangers via anonymous platforms such as Facebook confession boards, the topics tend toward more sensitive topics (e.g., health, illegal substances, sex) (Birnholtz et al., 2015).

There is no universally adopted definition of what constitutes sensitive content. Each online service provider defines sensitive data independently, describing it in the service's privacy policy or terms of service. For example, according to Google's privacy policy, sensitive personal data includes private medical details, racial or ethnic background, political or religious affiliations, and sexual orientation (Duffourc & Gerke, 2024), and Google does not link cookies or anonymous identifiers to sensitive ad categories, such as those related to race, religion, sexual orientation, or health (Chandrasekharan et al., 2018). Facebook's advertising policies restrict advertisers from targeting individuals based on sensitive personal attributes. These include characteristics such as race or ethnicity, religious or philosophical beliefs, age, sexual orientation or activity, gender identity, disabilities or health conditions (both physical and mental), financial status, trade union membership, and criminal history (Bond, 2021; Broniatowski et al., 2022). Quora considers adult content to be sensitive on questions and answers related to that type of content (Guo & Caine, 2021). Legal and data protection authorities have proposed their own definitions of sensitive content, which parallel but do not perfectly align with those that online service providers use.

User perceptions of what constitutes sensitive content often diverge significantly from the categories that both online services and policymakers establish. Privacy experts call for definitions that would allow the concept of sensitive data to "evolve over time and depend on the context, technologies, the use made of the data and even the individuals themselves" (CNIL Annual Report 2024). Despite this recognized need for contextual and user-centric approaches to sensitivity, no researchers have empirically examined how users behave when engaging with sensitive topics. This research gap is particularly significant given the increasing emphasis on user agencies in data protection frameworks and the growing recognition that one-size-fits-all content policies may be inadequate. While substantial research exists on general user behavior and content moderation policies, the specific question of how users behave in sensitive content contexts—such as sharing behavior, commenting patterns, or emotion expression — remains largely unexplored. Understanding these behavioral dynamics is crucial for developing more effective and user-centered approaches to content governance, particularly as platforms struggle to balance free expression with user protection in an era of increasingly diverse and global user bases.

3.3 Theoretic Framework and Hypotheses Development

We draw on Goffman's (1963) theory of stigma management and Suler's (2004) online disinhibition effect (ODE) to explain self-disclosure in anonymous online mental health communities. While Goffman emphasized the strategic concealment of stigmatized attributes in everyday interactions to preserve social identity, Suler offered a psychological account of how online affordances reduce behavioral inhibition. In the digital contexts, users can selectively manage their identities by leveraging dissociative anonymity, invisibility, and asynchronicity — affordances that mitigate the risks of social exposure (Suler, 2004). Suler defined ODE as the tendency for individuals to exhibit reduced behavioral inhibitions and express themselves more openly or intensely in online interactions compared to similar face-to-face situations. He identified several interrelated psychological factors that contribute to this phenomenon. Dissociative anonymity allows users to feel detached from their real-world identity, while invisibility removes the physical presence of others, reducing social cues and judgment. Asynchronicity gives individuals time to reflect and carefully prepare responses, unlike real-time conversations. Through solipsistic introjection, users may unconsciously assign imagined traits or voices to their online counterparts, merging their own internal dialogue with the other person. Dissociative imagination leads users to view online interactions as part of a make-believe world, disconnected from real-life consequences. Finally, the minimization of status and authority reduces visible hierarchies, fostering a sense of equality that can further embolden self-expression. Together, these factors help explain why people may disclose more personal information, show heightened emotion, or act more freely in digital spaces.

Suler's ODE framework is particularly relevant to the present study examining self-disclosure within anonymous online health communities. Specifically, the dissociative anonymity factor directly relates to the core feature of the community we are studying, where users can post without revealing their real names or identities. This lack of identifiability likely reduces the perceived social risks associated with sharing sensitive health information. Furthermore, the invisibility inherent in this text-based forum environment removes nonverbal cues (e.g., blushing, looking away) that might inhibit disclosure of potentially embarrassing health details in face-to-face settings. Similarly, asynchronicity allows users time to prepare potentially difficult or emotional disclosures about their health conditions, possibly leading to more deliberate and detailed sharing than might occur spontaneously.

Based on how these specific factors of the ODE are situated within the context of anonymous online health communities, we theorize the following effects regarding user self-disclosure.

Drawing on the principle of dissociative anonymity reducing perceived risk, we predict that posts made under conditions of higher anonymity (e.g., fully anonymous vs. pseudonymous, if applicable, or compared to hypothetical identifiable posts) will exhibit greater depth of self-disclosure (e.g., sharing more intimate details, vulnerabilities, or stigmatized information). In online Q&A platforms, anonymity removes the fear of being judged, labeled, or having personal disclosures tied back to one's real-world identity. This perceived safety net makes users more comfortable discussing vulnerable aspects of their health—particularly mental health issues, which are often associated with stigma and social sensitivity. The relationship between anonymity and self-disclosure is especially relevant in the context of mental health, where users may hesitate to share their experiences under conditions of identifiability. Anonymity allows users to speak more freely about emotional struggles, psychological symptoms, trauma, or diagnoses without fear of reputational harm. That is, users may be more likely to respond to mental health-related questions when doing so anonymously feels safer and less personally exposing. Therefore, we posit the following:

H1a: Transitioning to anonymity will significantly increase the number of answers provided by users in online mental health discussions.

Stemming from the principle of invisibility—which removes nonverbal feedback and social presence—and asynchronicity, which allows users to prepare responses more deliberately, anonymous online environments often lower the emotional and social barriers to disclosure. In particular, when users interact without revealing their real names or identities, they may feel less constrained by social norms, fear of judgment, or reputational concerns. This reduced inhibition can lead to greater openness, especially around distressing or stigmatized topics such as mental health. Therefore, we hypothesize the following:

H1b: Transitioning to anonymity will significantly increase the specific information related to mental health conditions and symptoms—as measured by a specificity score.

Dissociative anonymity can reduce the perceived connection between one's online actions and offline identity, making users feel psychologically distanced from the consequences of their disclosures. This detachment often lowers self-censorship, particularly when discussing topics that carry social risk or stigma. Health issues such as mental illness, sexually transmitted infections (STIs), or substance use are often subject to judgment or shame in public or semipublic spaces. An anonymous setting, by contrast, allows users to externalize these experiences without fear of recognition or backlash. The result is a digital environment that supports more candid, unfiltered conversations about stigmatizing topics. Considering that dissociative anonymity might make actions feel less tied to the "real self," we predict that users will be more likely to disclose socially sensitive or potentially stigmatizing health information (e.g., related to mental health, STIs, substance use) in anonymous posts compared to less anonymous conditions. Thus, we posit the following**:**

H1c: Transitioning to anonymity will result in significantly increased emotional expression, particularly the expression of negative or stigmatized emotions (e.g., anxiety, anger).

Although people generally believed that platform name policies influence user-generated content, this impact is not uniform across all users. It is important to recognize that specific credentials of content creators may moderate the effect of identity disclosure or anonymity. For instance, geographic location can shape both cultural attitudes toward privacy and the perceived risks of disclosure. As Forman et al. (2008) and Crossler and Belanger (2019) noted, users from regions with stronger privacy norms or higher stigma around mental health may respond more positively to anonymity, contributing more openly when their identity is masked. Similarly, professional background may moderate the effect. Peng (2021) found that individuals in prestigious or high-responsibility professions may be more cautious about disclosing sensitive information due to reputational concerns, making them more likely to reduce participation under real-name policies. Last, education level can influence how users perceive and navigate identity policies. Chi et al. (2022) suggested that individuals with higher education levels might be more aware of privacy risks or be strategic in managing their digital identities, which may either amplify or dampen the effects of anonymity depending on context.

Taken together, these factors suggest that the relationship between identity policies and online engagement is not one-size-fits-all—rather, who the user is, where they are from, and what they bring to the platform shapes the relationship. In our research we seek to uncover these nuanced dynamics by examining how user credentials interact with shifts from identity disclosure to anonymity in the mental health context.

Therefore, we posit the following hypotheses:

H2a: Users with high credential disclosure will experience a smaller increase in content quality under anonymity compared to users with low credential disclosure.

H2b: Users with high credential disclosure will experience a smaller increase in content specificity under anonymity compared to users with low credential disclosure.

H2c: Users with high credential disclosure will experience a smaller increase in content discrete emotion under anonymity compared to users with low credential disclosure.

3.4 Research Design and Methodology

The nature of online communities is that discussion of socially undesirable topics can be stigmatizing. Online Q&A communities offer a novel context in which to explore the implications of stigma and anonymity for ongoing self-disclosure about mental health. It became a popular arena for support and sharing of experience, knowledge, and advice among users. Users who help each other cope with different mental health conditions and learn together about mental health-related issues and behaviors use these platforms daily. Online Q&A communities are also noteworthy because they allow individuals to feel anonymous when self-disclosing. Users may strategically choose to conceal some or all information about their identity by using a pseudonym or nickname in place of their legal name, providing no identifying information, or using other strategies (Qian & Scott, 2007).

The preceding discussion of self-disclosure, stigma, and anonymity suggests several hypotheses. First, concerns about social disapproval associated with health-related stigma may serve to stifle self-disclosure. Anonymity may reduce concerns about social sanctions and enable individuals to relieve distress associated with discrimination and fear through self-disclosure. Anonymity could serve as a form of covering (Goffman, 2014) that allows users to share their health experiences without the stigma associated with their condition becoming obtrusive. Users may communicate their health experiences without the stigma connected with their condition becoming overt if they remain anonymous (Goffman, 2014). Existing literature shows users would be motivated to be anonymous to cancel their identity when they want to share greater levels of embarrassment about their illness (Warner et al., 2011). However, more research is necessary on the impact of identity intervention on promoting users' self-disclosure. Therefore, we propose that anonymity moderates the relationship between perceived health-related stigma and self-disclosure, which shifts the nature of contents (from more factual to more personal).

Second, although it is generally believed that platform name policies influence user-generated content, this impact is not uniform across all users. It is important to recognize that the effect of identity disclosure or anonymity may be moderated by the specific credentials of content creators. For instance, geographic location can shape both cultural attitudes toward privacy and the perceived risks of disclosure. As Forman et al. (2008) and Crossler & Belanger (2019) note, users from regions with stronger privacy norms or higher stigma around mental health may respond more positively to anonymity, contributing more openly when their identity is masked. Similarly, professional background may moderate the effect. Peng (2021) finds that individuals in prestigious or high-responsibility professions may be more cautious about disclosing sensitive information due to reputational concerns, making them more likely to reduce participation under real-name policies. Lastly, education level can influence how users perceive and navigate identity policies. Chi et al. (2022) suggest that individuals with higher education levels may be more aware of privacy risks or more strategic in managing their digital identities, which may either amplify or dampen the effects of anonymity depending on context.

Taken together, these factors suggest that the relationship between identity policies and online engagement is not one-size-fits-all—it is shaped by who the user is, where they are from, and what they bring to the platform. My research seeks to uncover these nuanced dynamics by examining how user credentials interact with shifts from identity disclosure to anonymity in the mental health context.

3.4.1 Research Context: Quora

We tested our research hypotheses in the context of Quora.com, the largest social Q&A platform in the US. Established in 2009, Quora is the largest peer-to-peer network of knowledge exchange and question and answers in the United States. We chose to investigate Quora for the following reasons. Quora is one of the largest social media platforms, with over 400 million monthly active users in 2024. Unlike other major social media platforms such as Reddit and Twitter, Quora strictly applied the name policy in 2014 but rescinded it in April 2021. The policy significantly reduced users' visibility and offered an alternative to being completely anonymous. The username status change and corresponding credentials are publicly observable on Quora.

Therefore, we could leverage the available data to extract comprehensive insights from the platform. In this study, we leveraged this exogenous shock as a natural experiment to explore the influence of anonymity on the quantity and duration of users' content engagement across sensitive topics. Specifically, we wanted to set our context in more socially undesirable topics such as mental health. We leveraged the readily available data to extract comprehensive insights from the platform.

We used this exogenous shock as a natural experiment to explore the impact of name change on the quantity and nature of users' content engagement across sensitive topics. We specifically set our context on more socially undesirable topics such as mental health.

Figure 1

Screenshot of Profiles Pages with Different Identity Display

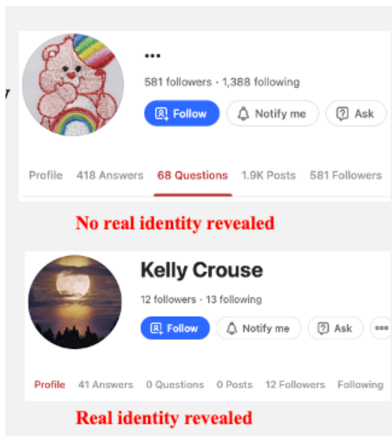
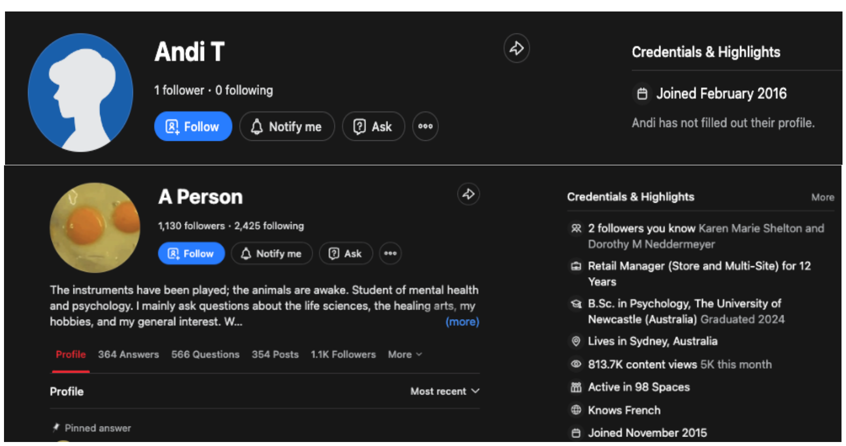


Figure 2

Screenshot of Profiles Pages with Different Types of Credentials



3.5 Sample and Variables

We based our analysis on a user-level weekly panel dataset drawn from a comprehensive online knowledge question and answer community. The final sample included 2,703 user-week observations from 760 matched users observed over approximately 26 weeks, spanning both the pre- and post-policy change periods.

3.5.1 Variable Definition

Table 2 succinctly describes the constructs of this study.

Table 2

Definitions of Constructs

Constructs	**Variables**	**Descriptions**
Perceived Identifiability	Anonymity (binary)	Binary construct on whether users' identity is revealed
User Engagement	Quantity	The number of answers per question
	Specificity	The score measures how detailed and precise each answer is in addressing mental health symptoms using LLMs
	Discrete Emotions	The score indicates how each answer reflected a specific emotional state, such as happiness, anger, fear, sadness, contempt, and surprise using LLMS
	User Credentials	The number of credentials users disclosed, such as profession, location, and registration age.

Dependent Variables

We used three outcome variables to capture both the quantity and quality of user-generated content within the mental health community. We measured each variable at the user-week level.

1. Quantity

We measured this variable as the number of answers per question; this variable captures the quantity of user contributions.

2. Specificity

This variable evaluates how detailed, precise, and symptom-relevant each answer is. Using LLMs, we scored answers based on the presence of symptom-specific information. The goal was to capture the informational quality of responses within a mental health support context. The rationale is that high-quality answers in clinical or diagnostic contexts should explicitly address the symptoms presented. LLMs, with their capacity to understand and extract nuanced information, can help quantify this alignment. By focusing on symptom-specific content, we rated responses that are more informative rather than generic or vague.

3. Discrete Emotions

This variable captures the emotional tone conveyed in each answer. Using LLMs, we scored responses based on the presence of specific, discrete emotions: happiness, anger, fear, contempt, surprise, and disgust, following the framework that Ekman et al. (1999) proposed. We chose these six emotions because of their universality—reliably recognized across cultures and expressed through consistent facial and verbal cues. By focusing on this set, we tapped into a well-established psychological model that helped us identify meaningful emotional signals. Measuring these emotions allowed us to assess whether responses exhibited empathy, distress, bias, or other affective cues that may shape user experience, trust, and perceived credibility. While other emotional taxonomies exist, Ekman's model provides a clear, manageable set of categories that are both theoretically grounded and computationally tractable.

LLMs are increasingly used to score text data in online communities due to their advanced natural language understanding, scalability, and adaptability (Tai et al., 2024). Unlike traditional keyword-based or lexicon-driven methods, LLMs can interpret the semantic and contextual nuances in user-generated content. This makes them particularly effective for analyzing complex text. First, an LLM provides consistent evaluations because it does not rely on other contemporaneous information (given an appropriate prompt) or personal opinions that could introduce biases, ensuring a reliable, consistent, and objective assessment of texts on online communities. Second, as an algorithm, LLM does not have the capacity constraints of humans and can process many texts in a short time frame.

In this research context, we used ChatGPT-4 to generate the specificity and discrete emotion score. When given a proper prompt, ChatGPT draws on this vast learned knowledge—acquired through a combination of supervised learning and reinforcement learning techniques—to interpret the input and generate the most contextually appropriate response based on statistical patterns in language. This allows it to notice intricate linguistic patterns and contextual correlations. In our study, we provided the following prompt for all the answers posts in the sample:

"Think of yourself as a social media guru with expertise on mental health. Based on the text, please answer the following questions:

1. How specifically does this answer post reveal the symptoms of the mental health related disease? Please give them a score.

2. What are the discrete emotions (happiness, sadness, fear, anger, disgust, surprise) this post reveals? Please give them a score based on different kinds of emotions"

Independent Variable

Our main independent variable is a dichotomous indicator: $\mathbf{Y}_{it}$, which equals 1 if anonymous users *i* generated answers in week *t*; otherwise, 0. We designed the independent variables to estimate the causal effect of the policy change from identity disclosure to anonymity, using a DID framework.

Moderators

To explore HTEs, we included a moderator that captures differences in users' identity disclosure behavior. Credential Disclosure reflects the degree to which users share personal credentials on the platform. We coded it as a binary variable, where *High Credential Disclosure* refers to users who disclose more than two types of credentials—such as education, profession, or location—while *Low Credential Disclosure* applies to users who disclose two or fewer types.

Control Variables

We included several time-varying control variables in our panel regression models to account for content visibility and platform-level engagement dynamics that may influence user behavior independently of the policy change. These controls included *Question Age*, which captures the time elapsed between the creation of a question and the implementation of the policy; *Number of Upvotes*, which reflects the positive reactions or thumbs-up an answer receives, indicating its perceived helpfulness or popularity among users; *Number of Shares*, representing how frequently an answer is shared, thereby capturing its reach within and beyond the platform; and *Number of Views*, which denotes how many times an answer is viewed each week, serving as a proxy for content exposure and visibility.

We included these controls to account for fluctuations in content popularity that could otherwise bias our outcome measures. By adjusting for these factors, we aimed to isolate the treatment effect from engagement driven by visibility or algorithmic amplification. Including these variables helped ensure that we are not mistakenly attributing the effects of high exposure or platform-driven traffic to policy intervention.

Additionally, our models incorporate user fixed effects, which control for unobserved time-invariant individual characteristics such as user baseline activity levels, writing ability, or previous platform engagement. We also included weekly fixed effects to account for broader platform-wide trends, such as seasonal fluctuations or overall growth in user activity. Together, these controls and fixed effects substantially strengthen the identification strategy by accounting for both observed and unobserved heterogeneity in content-level and temporal dynamics.

Table 4

Variables, Definition, and Descriptive Statistics

Variables	**Mean**	**Std dev.**	**Min.**	**Max.**
Specificity score	0.61	0.17	0.13	0.89
Anger	0.73	0.12	0.44	0.92
Joy	0.37	0.06	0.20	0.54
Sadness	0.31	0.05	0.31	0.67
Fear	0.24	0.12	0.14	0.56
Disgust	0.27	0.08	0.17	0.34
Suprise	0.25	0.04	0.12	0.56
#Credential	3	1.31	0	8
Question Age (year)	3.57	2.53	−1.92	10.33
# $Answers_{it}$	30.00	20.03	0.00	78.63
# $Voteups_{it}$	15.76	45.21	0.00	726.00
# $Shares_{it}$	12.37	10.26	0.00	68.00
# $Views_{it}$	135.25	230.36	0.00	63391

3.6 Empirical Approach

We applied the following two-step empirical approach to address all our hypotheses. First, we matched each answer without real usernames to an answer without real usernames using PSM. We conducted this step because a primary concern regarding endogeneity is the nonrandom selection of users to display users' names. One could argue that inherent differences may exist between our treatment and control groups. PSM ensures that answers by users with real usernames and answers by users without real usernames are as similar as possible with respect to an array of observable characteristics. Following PSM, we conducted DID analysis to evaluate pre- and post-policy changes in our variables of interest.

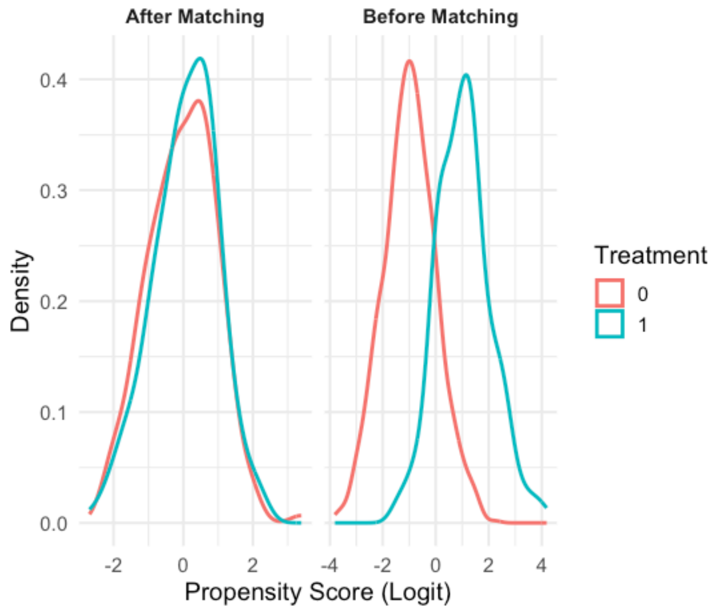
3.6.1 Step 1: PSM

We adopted the PSM method (Stuart et al., 2014) to mitigate potential selection bias concerns. Specifically, we included a series of variables, wherein time-varying covariates are measured using their pretreatment values from the period immediately before the policy introduction: the years since the creation of the question before the policy, the number of upvotes per answer, the number of shares per answer, the number of upvotes per answer, the number of views per answers, and whether the question was raised by a user awarded a Top Writer badge by Quora.

We employed the MatchIt package in R to perform an exact one-to-one matching. Notably, after matching, the differences for all covariates between the treatment and control groups become insignificant, suggesting that the treatment and control groups are appropriately matched, and that the covariate imbalance has been greatly reduced, as Figure 3 illustrates: the density distributions of propensity scores perfectly overlap, indicating a good match.

Figure 3

Density Distribution Plots of Logit Estimates for Propensity Scores Before and After Matching



3.6.2 Step 2: DID

In line with prior literature (Burtch et al., 2018; Shan & Qiu, 2023), we employed a two-way fixed-effects DID regression specification that allowed us to investigate our research questions:

$$Log(Y_{it}) = \beta\ (Answer_{it} \times PostAnonymity) + \gamma X_{it} + \lambda_i + \varepsilon_{it}\ (1)$$

There are three sets of dependent variables: the number of answers, the specificity score of the answers, and the discrete emotion score of the answers: Y_{it}, we considered two aspects of user's participation in user i in week t: the number of answers. The key coefficient $\beta\ (Answer_{it} \times PostAnonymity)$ indicates the causal effect of anonymity on the outcome. To improve the robustness and interpretability of our regression analysis, we normalized the scale of these variables using a logarithmic transformation because of their skewed distribution (J. Wang et al., 2022). The independent variable serves as a dummy variable indicating whether user *i* generated an answer during week t. The term X_{it} contains a series of control variables: the years since the creation of each question to week t, the number of upvotes per answer, the number of shares per answer, the number of comments per answer, and the number of views per answers per week. Finally, we included user-fixed effects λ_i to control potential variations associated with individual traits such as user tenure at the platform

3.7 Results and Interpretation

In this section we present the empirical results and interpret the estimated effects of the anonymity policy change on user engagement in terms of both quantity and quality. We first evaluate the parallel trends assumption to ascertain that the treatment and control groups exhibited similar outcome trajectories prior to the intervention, ensuring the validity of the DID design. Next, we employ the DID approach to establish causal estimates of the intervention of anonymity's impact on the outcome of interest (quantity, specificity, and discrete emotion). Subsequently, we explore potential moderators to assess whether the treatment effects vary across subgroups or contextual factors. Finally, we conduct a series of robustness checks—including alternative model specifications and matching methods—to confirm the stability and reliability of our findings.

3.7.1 Validating the Parallel Trends Assumption

According to Angrist and Pischke (2008), the parallel trends assumption, a fundamental tenet of the DID model, asserts that the outcome variable for both the treatment and control groups would have exhibited a comparable trend over time if the treatment—in this case, the switch from identity disclosure to anonymity—had not occurred. We applied the relative time model to alleviate the concern regarding the pretreatment trend and estimate the dynamics of the impact after the name policy change in Quora. Following prior literature (Burtch et al., 2018; J. Chan & Ghose, 2014; Greenwood & Agarwal, 2016; Greenwood & Wattal, 2017). Equation (2) shows our econometric specification:

$$Log(Y_{it}) = \sum_{-3}^{2} \beta\tau \, (Answer_{it} \times PostAnonymity) + \gamma X_{it} + \lambda_i + \varepsilon_{it} \text{ (2)}$$

Instead of using the name change dummy in the main analysis, we include a vector of relative time dummies T_τ ($\tau \epsilon [-3, 2]$), which indicates whether the current month is τ month before ($\tau < 0$) or after ($\tau > 0$) the name policy implemented. Additionally, we pool those time periods outside of the range $[-3, 2]$ into two relative dummies, that is, $T_{<-3}$ and $T_{>2}$. $RelativeTime$ (0), that is, the month when the name policy is implemented is omitted to serve as the reference. The remaining part remained the same: we included the same set of control variables, user, and month fixed. From Table 4 below, we observe that the parallel trend assumption is valid.

Table 4

Relative Time Model Results

	(1)	**(2)**	**(3)**
Variable	***New Answer***	***Specificity***	***Discrete Emotion (Anger)***
T_{-3} * $TreatGroup_i$	0.021	0.006	0.013
	(0.015)	(0.009)	(0.014)
T_{-2} * $TreatGroup_i$	−0.005	0.003	−0.004
T_{-1} * $TreatGroup_i$	(0.013)	(0.009)	(0.013)
	0.009	0.006	−0.007
T_0 * $TreatGroup$			
T_{+1} * $TreatGroup_i$	0.216***	0.026***	0.135***
	(0.017)	(0.010)	(0.026)
T_{+2} * $TreatGroup_i$	0.354***	0.143***	0.114***
	(0.016)	(0.010)	(0.018)
Constant	0.021	0.234***	0.237***
	(0.031)	(0.025)	(0.054)
Observations	2703	2703	2703
Within R-squared	0.04	0.006	0.009
Month FE	Yes	Yes	Yes
User FE	Yes	Yes	Yes

*p < 0.1; **p < 0.05; ***p < 0.01
Note. Robust standard errors clustered at the individual user level appear in parentheses.

3.7.2 DID Results for H1: Quantity, Specificity, and Emotions

As Table 5 shows, we found empirical support for all three components of H1. First, transition to anonymity increased (a) the quantity of users' answer, (b) the specificity score, and (c) anger score of users' answer. After users switched to anonymous profiles, they posted more answers per week on mental health questions than before (and more than matched users who did not switch). Specifically, users contributed 10% ($p < 0.01$), more answers per week following the adoption of anonymity, suggesting a significant increase in content quantity. Second, supporting H1b, the specificity of users' answers significantly increased by 30% ($p < 0.01$), indicating users with anonymous names provided more targeted and specific responses related to mental health symptoms. Third, in line with H1c, emotional expression also rose, with a notable increase (18%) in the anger emotion ($p < 0.05$).

Table 5

Difference-in-Differences Results

	Quantity	Specificity	Discrete Emotion (Anger)
Answer ×	0.102***	0.302***	0.176**
Post Anonymity	-0.016	-0.102	-0.072
User FE	Yes	Yes	Yes
Week FE	Yes	yes	Yes
# Observations	2307	2307	2307
Adjusted $R2$	0.02	0.016	0.015

*p < 0.1; **p < 0.05; ***p < 0.01

Note. Robust standard errors clustered at the individual user level appear in parentheses.

These findings collectively suggest that anonymity enhanced both the quantity and the quality—measured in terms of specificity and emotionality—of user-generated content. Specifically, the increase in anger may therefore signal a deeper and more emotional engagement with mental health topics, revealing issues that users might not express under their real names. This supports the idea that anonymity does not only increase expression generally—it also changes what kinds of emotions users feel safe to express These findings suggest that the shift to anonymous posting encouraged users to engage more actively and authentically in mental health discussions. We now turn to H4, in which we explore how users' credential disclosure moderates these changes in user behavior.

3.7.3 Moderation Analyses

In this section, we investigate our second research question: how do the user's credentials moderate the impacts of transitioning from identity disclosure to anonymity on the quality and quantity of user-generated content? To be more specific, we aim to understand how users' sharing of credentials moderates the real name policy's impact on quantity and quality. Table 6 presents the findings, on which we will elaborate in the subsequent paragraph.

$$Log(Y_{it}) = \beta_1\, Post_{it} + \beta_2\, Answer_{it} + \beta_3(Answer_{it} \times PostAnonymity) + \gamma X_{it} + \lambda_i + \varepsilon_{it} \quad (3)$$

This suggests that users with high levels of self-identification may have felt less incentivized or less compelled to engage even when their identity was no longer visible. In this context, identity visibility likely served as a signaling mechanism, motivating content creation among users seeking to build credibility or social capital.

In addition to this quantity effect, credential disclosure also significantly moderates emotional expression. Following the policy change, users with more disclosed credentials exhibited a greater decline (−45%) in anger-related word usage ($p < 0.01$). This pattern suggests that identity disclosure may have previously enabled high-credential users to express stronger emotional tones—perhaps to assert authority or convey frustration more confidently. Once the platform adopted anonymity, these users may have become cautious, neutral, or emotionally restrained, possibly due to reduced reputational accountability or shifting audience expectations.

Table 6

Moderation Effects Results

	Quantity	Specificity	Discrete Emotion (Anger)
Answer × Post Anonymity	0.132** (0.031)	0.302** (0.102)	0.124** (0.056)
Answer × Post Anonymity× Credential	−0.506*** (0.149)	−0.273 (0.174)	−0.445*** (0.137)
User FE	Yes	Yes	Yes
Week FE	Yes	yes	Yes
# Observations	2307	2307	2307
Adjusted *R*2	0.03	0.018	0.017

*p < 0.1; **p < 0.05; ***p < 0.01
Note. Robust standard errors clustered at the individual user level appear in parentheses.

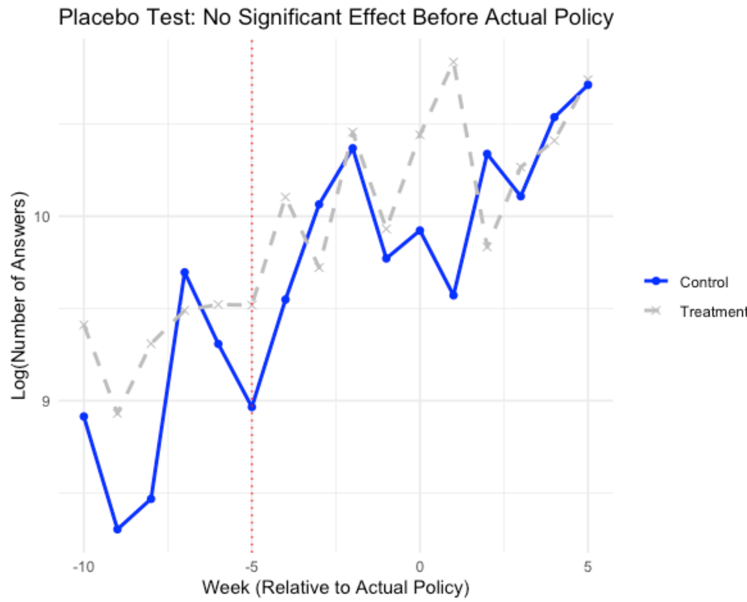
Interestingly, we find no significant moderating effect of credential disclosure on the specificity of answers (β = -0.273, n.s.). That is, regardless of how many credentials users disclosed, their level of detail or precision in addressing mental health symptoms remained statistically unchanged. This finding implies that while identity cues may influence user participation and emotional tone, they do not necessarily impact the informational depth of responses. One possible explanation is that topic seriousness and community norms drive specificity in mental health discussions more than self-presentation motives do. These results suggest that credibility signaling and content depth may operate independently in sensitive online discussions.

3.8 Robustness Tests

3.8.1 Placebo Test

We conducted a placebo test by assigning a fictitious policy change 5 weeks prior to the actual policy implementation. Specifically, we defined a placebo treatment indicator that switches to 1 at week −5, pretending the policy was enacted earlier. We re-estimated our DID model using this placebo treatment ($Log(Y_{it}) = \beta\ (PlaceboTreat_{it}) + \gamma X_{it} + \lambda_i + \varepsilon_{it}$). The results show that the placebo treatment effect is statistically insignificant (Figure 4), suggesting that there were no preexisting differential trends between the treatment and control groups prior to the actual policy change regarding the number of answers. We did the same test for the other two dependent variables—specific score and discrete emotional score—and obtained similar results. This finding strengthens the credibility of our main results by ruling out concerns related to spurious pre-trends.

Figure 4



3.8.2 Different Time Windows

We varied the pre- and post-policy observation window (2 months on each side instead of 3) and found consistent results, shown in Table 7.

Table 7

Difference-in-Differences Results

Signaling Theory	Quantity	Specificity	Discrete Emotion (Anger)
Answer × Post Anonymity	0.174***	0.214***	0.125***
	-0.016	-0.093	-0.064
User FE	Yes	Yes	Yes
Week FE	Yes	yes	Yes
# Observations	2307	2307	2307
Adjusted *R*2	0.1	0.023	0.017

*p < 0.1; **p < 0.05; ***p < 0.01
Note. Robust standard errors clustered at the individual user level appear in parentheses.

3.8.3 Different Matching Method and Specification Method

To ensure the robustness of our findings, we employed alternative matching methods and model specifications. Specifically, we utilized CEM, which improves covariate balance between treated and control groups by temporarily grouping variables into broader categories and then matching units exactly within those categories. This method reduces model dependence and potential bias in observational data analysis. In addition to matching approaches, we also adopted an interrupted time-series (ITS) strategy as an alternative specification. This method allowed us to evaluate the impact of an intervention by examining changes in outcome trends before and after a specific policy or treatment implementation. By incorporating these different analytical strategies, we aimed to test whether the estimated treatment effects remained consistent across methodological approaches, thereby increasing the credibility and generalizability of our results.

$$Y_t = \beta_0 + \beta_1 \cdot \text{Time}_t + \beta_2 \cdot \text{Intervention}_t + \beta_3 \cdot \text{TimeAfter}_t + \epsilon_t$$

In an ITS model, the outcome variable at each time point Y_t is modeled as a function of time and the intervention. The variable Time_t represents the underlying time trend across the entire study period, capturing any existing trajectory in the outcome before the intervention. The binary variable Intervention_t indicates whether the intervention has occurred, taking the value 0 before the intervention and 1 afterward; its coefficient estimates the immediate level change in the outcome following the intervention. To capture changes in the slope or trajectory after the intervention, the model includes TimeAfter_t, which counts the number of time units since the intervention and takes the value 0 prior to it. This term allows the model to estimate whether the rate of change in the outcome differs before and after the intervention. Together, these terms help isolate both the immediate impact and the long-term trend change attributable to the intervention, while ϵ_t represents the usual error term accounting for variation not explained by the model. Table

8 below demonstrated that the intervention had a positive and significant effect on all three outcomes (especially immediately after it was introduced).

Table 8

Interrupted Time Series Results

Variable	Quantity (Posts)	Specificity (Details)	Anger (Emotion)	Notes
Intercept (β_0)	15.20 (1.05) ***	3.10 (0.22) ***	1.25 (0.18) ***	Baseline level
Time (β_1)	0.12 (0.04) **	0.01 (0.01)	0.02 (0.01) *	Pre-intervention trend
Intervention (β_2)	2.35 (0.88) **	0.45 (0.18) **	0.31 (0.14) *	Immediate level change
Time After (β_3)	0.08 (0.03) **	0.02 (0.01) *	−0.01 (0.01)	Post-intervention trend change
R^2	0.21	0.18	0.15	Model fit

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

3.9 Implications

We have summarized our key findings in a structured table (Table 9) that outlines the main results, theoretical contributions, practical implications, and directions for future research. Drawing on this table, we dedicate each subsection to a major empirical result: (a) the impact of anonymity on engagement and emotional expression, (b) the moderating effect of credential disclosure on these changes, and (c) the lack of a moderation effect on content specificity. Each subsection addresses the implications of the findings for theory, practice, and future research.

3.9.1 Anonymity Increases Quantity of User Engagement

Our analysis revealed that the introduction of anonymity significantly increased user engagement within the online mental health community, as measured by the number of answers contributed per user per week. The uptick in participation suggests that the removal of identity disclosure requirements lowered barriers to entry and created a more comfortable environment for users to share and contribute (Chan et al., 2023; Huang et al., 2017; J. Pu et al., 2020). In particular, the increase was pronounced among users who had not previously disclosed credentials, indicating that anonymity may be especially empowering for users who are otherwise reluctant to participate due to identity-related concerns.

These findings strongly support the ODE (Suler, 2004), which posits that identity masking reduces social constraints, thereby encouraging freer participation in online environments. Our study adds nuance to this theory by showing that anonymity significantly increases the volume of user contributions in sensitive contexts such as mental health discussions. This suggests that identity disclosure can create self-presentation pressures that inhibit engagement—particularly in stigmatized domains—where individuals may fear judgment or reputational risk. Anonymity, in contrast, lowers psychological barriers (Fredheim et al., 2015; Qian et al., 2024) and encourages broader participation, especially from users who might otherwise remain silent.

Table 9

Summary of Main Findings and Implications

Key Findings	**Theoretical Implications**	**Practical Implications**	**Future Research Directions**
Anonymity increases quantity of answers (H1a)	Supports Suler's online disinhibition effect, specifically the reduction of self-presentation pressure and perceived risk, which encourages greater participation in sensitive discussions	Anonymity policies may help foster engagement in support communities by lowering barriers to participation	Explore differential effects across other sensitive domains (e.g., financial advice, trauma forums)
Anonymity improves answer specificity and emotional expression (e.g., more anger words) (H1b and H1c)	Reinforces Suler's notion of "toxic" and "benign" disinhibition: anonymity facilitates both emotionally intense and authentically detailed disclosures, challenging the idea that anonymity lowers quality	Platforms can consider anonymity as a design tool to encourage more expressive and symptom-specific communication	Investigate long-term effects of anonymity on community trust, and user activities

Credential disclosure reduces the positive effects of anonymity on both answer volume and emotional expression (H2a and H2c)	Demonstrates that identity signaling through credentials can reintroduce self-monitoring behaviors, tempering disinhibition; supports signaling theory and complicates Suler's framework by showing layered identity cues	Platforms should recognize that users with higher identity investment may disengage under anonymity and may need additional motivation	Study how identity cues interact with other platform features (e.g., profile photos)
No significant moderation effect of credential disclosure on content specificity (H2b, not supported)	Suggests that informational quality—defined as detail and precision—may be less affected by identity cues, challenging assumptions that expertise must be visible to enhance quality	Highlights that content depth can be preserved even in anonymous environments with credential disclosure	Examine how community norms shape informational quality across varying identity conditions

3.9.2 Anonymity Improves Quality of User Engagement (Specificity and Emotional Expression)

In addition to increasing the quantity of user contributions, our analysis reveals that anonymity improves the quality of user-generated content in two important dimensions: specificity and emotional expression. Specifically, users' answers became more detailed and symptom-focused following the adoption of anonymity, as assessed by an LLM. Furthermore, there was a notable increase in emotional expressiveness, especially in the use of anger-related words. These changes suggest that anonymity not only encourages users to participate more frequently but also enables them to communicate with greater emotional honesty, particularly in a negative one such as anger.

These findings extend the ODE by showing that anonymity can enhance not only the frequency but also the depth and intensity of user engagement. While ODE is often associated with reduced inhibition, our results demonstrate that this disinhibition extends to increased specificity in informational content, which prior literature has not emphasized. The findings also complicate concerns from earlier research that anonymity may lead to lower-quality or shallow content.

Instead, in high-empathy, topic-specific environments, anonymity appears to facilitate authenticity and cognitive elaboration rather than degrade quality (Huang et al., 2017). Additionally, from the perspective of social identity and self-presentation theories (Reicher et al., 1995), the ability to post anonymously may allow users to shift away from managing their digital identity and instead focus on delivering honest, emotionally resonant, and content-rich responses. This finding suggests that emotional honesty and informational detail are not necessarily in tension, and that anonymity can serve both ends in emotionally charged environments.

3.9.3 Credential Disclosure Moderates the Impact of Anonymity

Our moderation analysis revealed that users who disclosed more credentials—such as professional background, education, or locations—showed a significantly smaller increase in new answer contributions and a sharper decline in emotional language after the platform shifted to anonymity, compared to users with fewer disclosed credentials.

This finding refines and extends identity signaling theory (Dennen & Arslan, 2022). It suggests that users who previously invested in identity disclosure to build credibility or social capital may experience a loss of incentive to participate once those identity signals become invisible. Rather than simply adapting to anonymity, high-credential users appeared to withdraw engagement and emotional expression, implying that for some users, visibility is a necessary component of motivation and emotional authenticity. This highlights an important boundary condition for the ODE: while anonymity encourages some users to express themselves more freely, it discourages others who relied on identity signaling for status or credibility.

3.9.4 No Moderation Effect on Content Specificity

Our analysis revealed no statistically significant interaction between credential disclosure and anonymity in predicting the specificity of user-generated content. In other words, regardless of whether users disclosed detailed credentials or remained relatively anonymous prior to the policy shift, the level of specificity in their answers—measured by how precisely they addressed mental health symptoms—remained unchanged after anonymity was adopted.

This finding introduces an important nuance to identity signaling theory. While prior research suggests that identity cues can affect content quality by encouraging users to signal expertise or credibility, our results indicate that in high-stakes, sensitive contexts such as mental health, topic seriousness and community norms may override self-presentation motives. The absence of a moderation effect suggests that intrinsic norms of care and informational depth rather than strategic identity may govern management content specificity. This challenges the assumption that higher identity disclosure always correlates with higher informational quality and points to a potential decoupling of credibility signaling from content richness in emotionally sensitive online communities.

3.9.5 Practical Implications

For Platform Designers and Online Community Managers

These findings demonstrate that anonymous posting represents a powerful strategic tool for increasing user engagement without compromising content quality. By lowering participation barriers, anonymity creates pathways for more inclusive and active digital support communities. Platforms should prioritize integrating anonymous participation features, particularly in sensitive topic areas such as mental health, addiction recovery, or stigmatized conditions. However, successful implementation requires robust moderation systems and community guidelines that maintain healthy dialogue while preserving the benefits of anonymous expression. The evidence that anonymity enhances rather than diminishes content quality challenges common assumptions about identity disclosure (B. Qian et al., 2024). Platforms can confidently implement anonymous-by-default posting in high-sensitivity contexts, knowing that users will likely contribute more authentic, emotionally rich, and symptom-specific content. Online community managers should anticipate and prepare for increased emotional intensity in anonymous posts, including expressions of anger and frustration that users may suppress in identified contexts. This requires developing sophisticated response protocols, including automated support prompts, crisis intervention pathways, and specialized training for moderators handling emotionally charged anonymous content.

Balancing Anonymity with Expertise

The challenge of maintaining credibility in anonymous environments requires innovative solutions. Platforms serving expert-driven communities can implement verified anonymous credentials systems—allowing professionals to demonstrate qualifications without identity disclosure. "Verified anonymous expert" badges or similar mechanisms can preserve both privacy and professional accountability.

3.9.6 Future Research Implications

Future researchers should investigate whether the observed increases in specificity and emotional expression represent sustained behavioral changes or temporary responses to anonymity adoption. Longitudinal studies tracking user behavior over months or years could reveal whether users maintain heightened authenticity as they acclimate to anonymous environments, or whether initial effects diminish as anonymity becomes normalized.

Researchers should examine adaptation patterns across different user types, particularly investigating whether high-credential contributors gradually rebuild social capital within anonymous systems or experience permanent disengagement. Understanding these trajectories will inform platform design decisions about balancing anonymity with long-term community sustainability.

The relationship between anonymity and expression likely varies significantly across cultural contexts and community norms. Future researchers should explore how different cultural backgrounds shape what types of emotional expression and specificity are considered acceptable or desirable in anonymous versus identified contexts. Cross-cultural studies could reveal whether anonymity's effects on authenticity are universal or culturally mediated. Additionally, researchers should investigate how community-specific norms influence anonymous behavior. Different support communities (mental health, chronic illness, addiction recovery) may have distinct expectations for appropriate emotional expression and information sharing that moderate anonymity's effects.

3.10 Conclusion

This study contributes to our understanding of how identity design—specifically, anonymity and credential disclosures—shapes user engagement, emotional expression, and informational specificity in online mental health communities. Our findings reveal that anonymity not only increases the quantity of user contributions but also enhances the quality of content, as reflected in greater specificity and emotional expression. This challenges the traditional view of a trade-off between anonymity and content quality and instead suggests that, under the right conditions, anonymity can foster deeper, more meaningful engagement.

These results highlight the need for nuanced approaches to platform governance, where identity policies are not one-size-fits-all but are instead tailored to community goals and content types. As online support spaces continue to grow in scale and influence, understanding how design choices impact behavior is more important than ever. Our study offers a foundation for future research and practice at the intersection of digital identity, mental health, and online community design.

Theoretically, our work deepens the understanding of how identity signals influence user behavior, extending social identity and signaling theories into the digital health context. Practically, the findings inform platform designers and moderators on how anonymity can promote richer engagement and emotional openness, particularly in sensitive domains such as mental health. Methodologically, we employ a rigorous DID framework and leverage language-based LLM metrics to capture subtle behavioral changes, offering a model for evaluating platform-level interventions using computational methods.

Taken together, these contributions advance our knowledge of online behavior and platform design, particularly in contexts where user vulnerability, authenticity, and emotional support are central. As digital platforms increasingly shape the way individuals seek help and connect with others, understanding the role of identity design becomes essential for building healthier, more inclusive online communities.

References

Agne, R. R., Thompson, T. L., & Cusella, L. P. (2000). Stigma in the Line of Face: Self-Disclosure of Patients' HIV Status to Health Care Providers. *Journal of Applied Communication Research*, *28*(3), 235-261.

Afifi, T., & Steuber, K. (2009). The Revelation Risk Model (RRM): Factors that Predict the Revelation of Secrets and the Strategies Used to Reveal Them. *Communication Monographs*, *76*(2), 144-176.

Anonymous. (1998). To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication. *Communication Theory*, *8*(4), 381-407.

Balani, S., & De Choudhury, M. (2015, April). Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1373–1378). Association for Computing Machinery.

Chen, S. (2018). What's in a Name: Facebook's Real Name Policy and User Privacy. *Kansas Journal of Law & Public Policy*, *28*(1), 146-172.

Christopherson, K. (2007). The Positive and Negative Implications of Anonymity in Internet Social Interactions: "On the Internet, Nobody Knows You're a Dog". *Computers in Human Behavior*, *23*(6), 3038-3056.

Cho, D., Kim, S., & Acquisti, A. (2012, January). Empirical Analysis of Online Anonymity and User Behaviors: the Impact of Real Name Policy. In 2012 45th Hawaii International Conference on System Sciences (pp. 3041-3050). IEEE.

De Choudhury, M., & De, S. (2014, May). Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. In Proceedings of the International AAAI conference on Web and Social Media (Vol. 8, No. 1, pp. 71-80). AAAI Press.

Derlega, V. J., Winstead, B. A., Greene, K., Serovich, J., & Elwood, W. N. (2004). Reasons for HIV Disclosure/Nondisclosure in Close Relationships: Testing a Model of HIV–Disclosure Decision Making. *Journal of Social and Clinical Psychology*, *23*(6), 747-767.

Fisher, G. (2019). Online Communities and Firm Advantages. *Academy of Management Review*, *44*(2), 279-298.

Haimson, O. L., & Hoffmann, A. L. (2016). Constructing and Enforcing" Authentic" Identity Online: Facebook, Real Names, and Non-Normative identities. First Monday.

Hogan, B. (2013). Pseudonyms and the Rise of the Real-Name Web. *A Companion to New Media Dynamics*, 290-307.

Holtgraves, T. (1990). The Language of Self-Disclosure. In H. Giles & W. P. Robinson (Eds.), *Handbook of Language and Social Psychology* (pp. 191–207). John Wiley & Sons.

Hutchinson, A. (2020, January 30). Facebook Climbs to 2.5 Billion Monthly Active users but Rising Costs Impede Income Growth. *Social Media Today*. https://www.socialmediatoday.com/news/facebook-climbs-to-25-billion-monthly-active-users- but-rising-costs-imped/571358.

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3), 321–326.

Keipi, T., Näsi M., Oksanen, A., & Räsänen, P. (2017). Online Hate and Harmful Content: Cross-national Perspectives. Routledge.

Kim, J., & Dindia, K. (2011). Online Self-Disclosure: A Review of Research. *Computer-Mediated Communication in Personal Relationships*, 156-180.

Legard, R., Keegan, J., & Ward, K. (2003). In-depth interviews. In J. Ritchie & J. Lewis (Eds.), Qualitative research practice: A guide for social science students and researchers (pp. 138–169). SAGE Publications.

Lim, J., Nicholson, J., Yang, S., & Kim, H. (2015). Online Authenticity, Popularity, and the “Real Me” in a Microblogging Environment. *Computers in Human Behavior*, *52*, 132-143.

Mahaphol, I., & Jetter, A. J. (2014, July). Does Personality Matter for Contributions to Online Communities? In Proceedings of PICMET'14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration (pp. 1966- 1973). IEEE.

Marx, G. T. (1999). What's in a Name? Some Reflections on the Sociology of Anonymity. *The Information Society*, *15*(2), 99-112.

Mcauley, J., & Leskovec, J. (2014). Discovering Social Circles in Ego Networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *8*(1), 1-28.

Qian, H., & Scott, C. R. (2007). Anonymity and Self-Disclosure on Weblogs. *Journal of Computer-Mediated Communication*, *12*(4), 1428-1451.

Rains, S. A. (2014). The Implications of Stigma and Anonymity for Self-Disclosure in Health Blogs. *Health communication*, *29*(1), 23-31.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *MIS Quarterly*, 841-864.

Rizwan, M., Shabbir, A., Javed, A. R., Srivastava, G., Gadekallu, T. R., Shabir, M., & Hassan, M. A. (2022). Risk Monitoring Strategy for Confidentiality of Healthcare Information. *Computers and Electrical Engineering*, *100*, 107833.

Ruesch, M., & Märker, O. (2012). Making the Case for Anonymity in E-Participation: An Evaluation of Real Name Policy in Gütersloh’s Second Participatory Budget. *eJournal of eDemocracy and Open Government*, *4*(2), 301-317.

Schlosser, A. E. (2020). Self-Disclosure versus Self-Presentation on Social Media. *Current Opinion in Psychology*, *31*, 1-6.

Slade, P., O'Neill, C., Simpson, A. J., & Lashen, H. (2007). The Relationship Between Perceived Stigma, Disclosure Patterns, Support and Distress in New Attendees at an Infertility Clinic. *Human Reproduction*, *22*(8), 2309-2317.

Steuber, K. R., & Solomon, D. H. (2011). Factors that Predict Married Partners' Disclosures about Infertility to Social Network Members. *Journal of Applied Communication Research*, *39*(3), 250-270.

Taggart, T., Grewe, M. E., Conserve, D. F., Gliwa, C., & Isler, M. R. (2015). Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. *Journal of Medical Internet Research*, 17(11), e4387.

Warner, C. H., Appenzeller, G. N., Grieger, T., Belenkiy, S., Breitbach, J., Parker, J., Warner, C. M., & Hoge, C. (2011). Importance of Anonymity to Encourage Honest Reporting in Mental Health Screening after Combat Deployment. *Archives of General Psychiatry, 68*(10), 1065–1071

Zick Rubin. (1975). Disclosing Oneself to a Stranger: Reciprocity and its Limits. *Journal of Experimental Social Psychology*, *11*(3), 233–260.

Chapter 4: Conclusion

This dissertation advances our understanding of how digital technologies—specifically telehealth and online platform design—reshape the mental health landscape at both community and individual levels. Through two complementary essays, it offers a nuanced, multilevel perspective on how digital interventions operate within broader social, economic, and infrastructural contexts to influence mental health outcomes.

The first essay contributes to the literature by conceptualizing telehealth as a form of digital capital and examining its role in enhancing community mental health outcomes. Using a quasi-experimental design and causal forest modeling on a panel dataset of over 2,800 U.S. counties, this essay reveals that telehealth adoption improves mental health at the county level, but that these effects are highly heterogeneous. Specifically, the most pronounced benefits emerge in under-resourced communities—particularly those with low broadband access but high economic or cultural capital. This finding challenges the assumption that infrastructure alone is sufficient and underscores the need for context-aware telehealth strategies that align with existing forms of capital. The methodological contribution lies in applying causal machine learning to uncover treatment effect heterogeneity in health IT interventions—offering both precision and policy-relevance.

The second essay shifts focus to the individual level and explores how anonymity affects user engagement and content quality in online mental health discussions. Leveraging a natural policy change on Quora and employing a matched difference-in-differences approach combined with LLM-based text analysis, this study finds that transitioning from real-name to anonymous posting significantly increases participation, content specificity, and emotional expression, particularly among users with low credential disclosure. These results provide novel insights into how platform identity policies shape engagement in sensitive health contexts and extend signaling theory to digital mental health communities. Theoretically, the work contributes to literature on online disinhibition, identity management, and user vulnerability, and highlights the empowering role of anonymity for marginalized voices.

Collectively, this dissertation contributes to both theory and practice in several important ways. Conceptually, it expands the application of social capital theory and signaling theory to digital health contexts, demonstrating how digital interventions—such as telehealth and platform design—interact with structural and psychological factors to shape mental health outcomes. Methodologically, it introduces innovative analytical approaches to the digital health literature by combining causal forest modeling, policy tree visualizations, and LLM-based content analysis. These tools enable a more granular and flexible understanding of treatment heterogeneity and user behavior in both structured and unstructured data. Empirically, the dissertation offers robust, large-scale evidence that telehealth adoption improves mental health outcomes most in structurally disadvantaged communities, and that anonymity empowers vulnerable users to engage more actively and authentically in online mental health discussions. In terms of policy relevance, the findings provide actionable insights for both health care policymakers and digital platform designers. For policymakers, the results highlight the importance of tailoring telehealth expansion strategies to local resource constraints and social capital configurations. For platform designers, the research underscores the value of anonymity features in fostering inclusive, supportive digital environments for stigmatized health topics. Together, these contributions help reframe digital health not merely as a technological solution, but as a contextually embedded, socially responsive tool for improving mental health at both community and individual levels.

As digital technologies continue to permeate health care delivery and support systems, this dissertation underscores the importance of designing equitable, context-sensitive interventions. It challenges one-size-fits-all assumptions and highlights the need to tailor digital mental health strategies to community conditions and user characteristics. Ultimately, this research contributes to a more inclusive vision of digital health—one that not only expands access but also empowers users to engage meaningfully in their mental health journey.