Towards Achieving Mental Health Equity for the Underserved Population: Evaluating the Potential of Mobile Apps

Yi Tang
Carlson School of Management
University of Minnesota
tangx688@umn.edu

Kingshuk K. Sinha
Carlson School of Management
University of Minnesota
ksinha@umn.edu

Adam Moen
Real EmPowerment Solutions
adam@reptech.io

Necati Ertekin
Carlson School of Management
University of Minnesota
nertekin@umn.edu

Under 3rd round review at *Production and Operations Management*
(Previous round: Minor Revision)

Version: October 1, 2023

Towards Achieving Mental Health Equity for the Underserved Population: Evaluating the Potential of Mobile Apps

Abstract

The underserved population – i.e., individuals who identify themselves as non-female, non-heterosexual, or non-White – have long faced inequities when accessing traditional clinic-based mental health services. Individuals from the underserved population tend to use and benefit from these services less than their better-served counterparts (i.e., those who identify themselves as female, heterosexual, and White). This study investigates whether similar inequities exist within the context of mobile apps specifically designed to provide self-support and peer-support. To this end, we conduct an empirical analysis using longitudinal user-level data collected from a mental health mobile app (MHMA). Our results indicate that: (i) in contrast to the traditional clinic-based mental health services, in a MHMA setting, users from the underserved population engage with the app services and derive similar benefits as much as their better-served counterparts; and (ii) there is a positive relationship between app usage frequency and the mental condition of app users. Our post-hoc analysis uncovers that the MHMA promotes equitable usage and benefit for users: (i) from the underserved population via self-management functions that enable self-support; and (ii) from the better-served population via online community functions that facilitate peer-support. These findings suggest that MHMAs have the potential to achieve equity through both user engagement and the benefits derived from such engagement. We conclude by discussing the implications for mobile app firms, policymakers, and organizations in their efforts to achieve mental health equity for the underserved population.

Key words: Mental health equity, mobile app, self-support, peer-support, underserved population

1. Introduction

Affecting more than 700 million people around the world (WEF 2019), mental illnesses[1] have become a leading cause of disability for decades (Murray et al. 2018). In the U.S. alone, mental illnesses are estimated to generate a $193.2 billion loss in annual earnings (NAMI 2021a). The COVID-19 pandemic has further exacerbated the already-severe mental illnesses due to fear and stress related to COVID-19 death tolls, working from home, reduced social contact, and unemployment (Moreno et al. 2020, Wang et al. 2020). Recognizing that preventing mental illnesses is essential for sustainable development on a global scale, the United Nations has officially established Goal 3.4, aiming to reduce mental illness-related mortality by one third before 2030 (United Nations 2020a). Similarly, the United Nations has responded to the mental health challenges posed by the COVID-19 pandemic by issuing comprehensive guidance (United Nations 2020b).

Despite the increasing global efforts to address mental illnesses, surprisingly, only around 40% of adults with any mental illness used mental health services between 2013 and 2018 (SAMHSA 2019) mainly for two reasons. First, on the supply-side, mental health services suffer from insufficient resources (e.g., shortages of psychiatrists and mental health clinics, and limited insurance coverage) (Mnookin 2016).

[1] A mental illness is a mental, behavioral, or emotional disorder that can vary in impact, ranging from no impairment to mild, moderate, or severe impairment (NAMI 2021b). Our study particularly focuses on mild to moderate mental illnesses (e.g., anxiety disorder, depression, eating disorder, etc.) and does not consider serious mental illnesses (i.e., those that interfere or limit major life activities).

Second, on the demand-side, unlike the patients with physical health conditions (e.g., patients with flu or cancer), individuals with mental health conditions often fail to recognize or acknowledge their illnesses due to self-stigma and/or social-stigma associated with seeking professional help at mental health clinics (Eisenberg et al. 2009, Masuda et al. 2009, 2012). Collectively, these observations indicate a significant gap between the supply and demand of mental healthcare delivery, with existing resources for mental health services being constrained and underutilized.

Furthermore, it is concerning that the highlighted gap in mental health service usage varies significantly across populations. The gap widens disproportionately for the underserved population characterized by their socio-demographic characteristics of race-ethnicity and sexual orientation.[2] Individuals from this population are often concentrated in underdeveloped regions, face socio-economic disadvantages (Chow et al. 2003), and/or experience discriminations (Thornicroft 2008). Consequently, essential services, including mental health services, are not equally provided to these groups, making them encounter difficulties in accessing these services due to the *supply-side* challenges. For instance, in 2021 in the U.S., among adults with any mental illness, 52.4% of Whites received mental health services, compared with 39.4% of African Americans, 36.1% of Hispanics or Latinos, and 25.4% of Asians (NAMI 2021a). Similarly, lesbian, gay, bisexual, and transgender (LGBT) individuals are less likely to access mental healthcare services compared to the non-LGBT population (NAMI 2021a, Semlyen et al. 2016). These inequities extend beyond service usage and also persist in the benefits derived from such services. For instance, African-American patients are more likely to discontinue clinic-based treatments prematurely compared to White patients (Fleck et al. 2005, Mongelli et al. 2020, Satcher 2001), suggesting that the underserved population is also less likely to benefit from mental health services even when they use them.

Interestingly, unlike other service domains, the definition of the underserved population within mental healthcare goes beyond sub-populations characterized by their race-ethnicity or sexual orientation, and also includes the male sub-population. The mental healthcare literature consistently demonstrates that males (including White males and heterosexual males) are worse off with respect to their mental health indicators, exhibiting higher rates of suicide, disorders, substance use, and violence compared to their female counterparts (Berger et al. 2005, Davis et al. 2023, Michael 2017, Rice et al. 2018, SAMHSA 2023, Smith 2022, Wong et al. 2017). This arises because compared to females, males: (i) have more difficulty in recognizing emotional issues; and (ii) have greater self- and social-stigma associated with seeking mental health support as they perceive help-seeking as a threat to their masculinity (Berger et al. 2005).

[2] According to the White House and the Department of Health and Human Services (The White House 2021, US Department of Health and Human Services 2019), the underserved population encompasses "*populations sharing a particular characteristic […] that have been systematically denied a full opportunity to participate in aspects of economic, social, and civil life*" and include individuals from several sub-populations such as Latinos, African Americans, LGBTQ+ individuals, and more.

Subsequently, even though mental health services are equally available across gender groups, males face difficulties in utilizing these services due to the *demand-side* challenges stemming from social norms. In the U.S., for instance, among adults with any mental illness, the rate of mental health service usage is 51.2% for females but only 37.4% for males (NAMI 2021a). Similarly, among White adults with any mental illness in the U.S., the rate of mental health service usage between 2015 and 2019 was 54.5% for White females but only 39.7% for White males (SAMHSA 2021). Considering these empirical insights and aligning with the mental healthcare literature, this paper refers to individuals as belonging to the *underserved population* if they identify themselves with one or more of the following socio-demographic characteristics: (i) Gender (i.e., male and other non-female gender identities (Berger et al. 2005, Eisenberg et al. 2009)); (ii) Sexual orientation (i.e., homosexual and other non-heterosexual identities (Hegland and Nelson 2002, NAMI 2021a, Plöderl and Tremblay 2015)); and (iii) Race-ethnicity (i.e., African, Asian, Hispanic/Latino and other non-White identities (Masuda et al. 2009)). Conversely, the *better-served population* refers to individuals who identify as female, heterosexual, and White (Terlizzi and Norris 2021).

Given the significant supply and demand gap, as well as the existing inequities, various stakeholders in the mental healthcare ecosystem are actively seeking solutions to ensure equitable mental health services (Agic 2019, Kirmayer and Jarvis 2019). One solution that has gained traction is the use of *mental health mobile apps* (MHMAs). In recent years, there has been a proliferation of MHMAs offering a range of mental health services to individuals with mental health needs (Donker et al. 2013, Moreno et al. 2020). The demand for MHMAs has been on the rise as people increasingly feel comfortable and secure seeking support for their mental health needs through mobile apps (Gray et al. 2005, Rickwood et al. 2007). Initially, early versions of MHMAs (e.g., DBT Field Coach, CBT MobilWork) were primarily developed for clinical use, empowering mental health professionals to prepare patients for treatments, deliver interventions (e.g., therapies) via the apps, and engage in follow up with patients post-treatment (Ancis 2020, Price et al. 2014). More recent versions of MHMAs (e.g., I Am Sober, Sanvello, and Happify) were designed for non-clinical use by individuals, providing self-support and/or peer-support mental health services (Dorwart 2023). In these MHMAs, mental health professionals do not play a direct role. Instead, app users can: (i) partake in self-support by utilizing app functions such as self-management, skill-training, and symptom tracking, which do not require interactions with other users (NIH 2019); and/or (ii) engage in peer-support with other app users by utilizing app functions that facilitate informational support, emotional support, and companionship (Yan and Tan 2014).[3] Consequently, unlike the traditional clinic-based mental health

[3] MHMAs differ in two aspects from social media platforms that may also foster peer-support (Ancis 2020): (i) MHMAs uniquely offer evidence-based self-support tools such as mindfulness, meditation, anxiety control, attention span improvement, self-awareness, suicide prevention, and self-reflection; (ii) MHMAs facilitate anonymous online communities where users can exchange social support in a psychologically safe environment, while social media platforms encourage disclosure of identities and sharing personal information for connection-building and upkeep.

services that are typically accessed when patients seek treatment from professionals in clinical settings, mental health services within MHMAs are accessed as app users engage with self-support and peer-support functions available to them. Despite the emergence of such apps, to the best of our knowledge, no study has examined whether the well-documented inequity across the underserved and better-served populations in the traditional clinic-based mental healthcare setting also exists within the context of MHMAs designed for providing self-support and peer-support. Towards addressing this gap, our study aims to investigate the following research questions:

(i) **(Equity in Usage)** *Do users from the underserved population use MHMAs less than users from the better-served population?*

(ii) **(Benefit of Mobile Apps Usage)** *Does MHMA usage improve mental condition of the users of MHMAs?*

(iii) **(Equity in Benefit)** *Do users from the underserved population benefit from MHMA usage less than users from the better-served population?*

We investigate our research questions using data obtained from an MHMA, hereafter called Hope (pseudonym). Launched by a mobile app start-up company in 2015, Hope is a free platform available for iOS or Android smartphone users. It offers two distinct functions for self-support and peer-support purposes:

(i) The *self-reflection function* enables users to evaluate their own mental condition, facilitating self-support by enhancing their awareness of personal mental health needs, all without the need to interact with other users.

(ii) The *online community function* enables app users to freely interact with other app users (peers) by sharing their feelings, seeking and/or providing informational and emotional support. This function fosters peer-support among users and requires active engagement with other users.[4]

Our data include extensive app activity records for the two functions, along with data on socio-demographic characteristics, covering 1,843 app users from May 2015 to February 2018. Among those users, 62% belong to the underserved population whereas the remaining 38% belong to the better-served population. With this dataset, we can monitor an individual user's app activities and observe how their mental condition evolves over time.

Using this longitudinal dataset, we evaluate the equity of usage by comparing the app usage frequency between users from the underserved population and users from the better-served population. We measure the benefit derived from app usage by examining the relationship between a user's mental condition at a specific time and their prior app usage frequency. Introducing an indicator that distinguishes users from the

[4] As illustrated in Online Appendix A, with respect to the function offerings, Hope is representative of MHMAs for non-clinical use in the market.

underserved population versus the better-served population as a variable that moderates this relationship, we evaluate the equity in benefits experienced by these two populations. From a methodological standpoint, we assess equity by examining *statistical equivalence* in usage and benefit between users from the underserved population and users from the better-served population.

Our study makes the following contributions to the literature on mental healthcare operations management. First, we find that the well-documented inequities in professional mental health service usage between the underserved and better-served populations (Alegría et al. 2002, Aneshensel 2009, Berger et al. 2005, NAMI 2021a, Plöderl and Tremblay 2015) are unlikely to exist in the context of MHMA designed to provide self-support and peer-support. As such, the app usage among users from the underserved population is statistically equivalent to that of users from the better-served population. Second, MHMAs are likely to complement professional mental health services, as we find a positive association between app usage frequency and users' mental condition. For instance, doubling app usage frequency for users with the median self-reflection score (i.e., the measure we adopt for users' mental condition) in our study sample is estimated to improve their self-reflection score by 4.69%. Third, concerning equity in benefit, we find that users from the underserved population are at least as likely as users from the better-served population to benefit from using Hope. This suggests that MHMAs are likely to promote equity not only in usage but also in the benefits derived from such usage.

Next, we conduct post-hoc analysis to examine: (i) how the two major functions of Hope – i.e., the self-reflection function for self-support, and the online community function for peer-support – contribute to our main results; and (ii) whether the main results are more nuanced across users from various underserved sub-populations. Regarding the two major functions, we find that our main results hold true for the usage of each function. However, users from the underserved population are marginally more inclined to use the self-reflection function than users from the better-served population. This implies that the benefit of Hope is primarily realized through the self-reflection function for users from the underserved population, whereas for users from the better-served population, it is realized more through the online community function. This highlights the importance for app developers to consider the distinct value proposition of each function, since MHMAs can effectively cater to the awareness of mental health support needs for users from the underserved population and facilitate peer-support for users from the better-served population. Regarding heterogeneity across users from different underserved sub-populations, our results indicate that MHMAs are particularly beneficial for female users from the underserved population (e.g., female African-American, female homosexual, etc.). Compared to users from other underserved sub-populations, these female app users not only use the app more frequently but also derive greater benefits from the app usage. This suggests that MHMAs can be particularly valuable in societies where violence and discrimination against females from the underserved population are prevalent. For other underserved sub-populations, while some (e.g., users with other gender identity, other sexual orientation, and White underserved) use Hope significantly more than the better-served population, equity in benefit is present across all underserved sub-populations.

Our results highlight the potential of MHMAs designed for self-support and peer-support to offer equitable services to individuals with mental health needs from the underserved population. This contribution aligns with the calls for research in *Production and Operations Management* on identifying emerging technologies that prioritize inclusion and equity for the underserved population (Kalkanci et al. 2019) and social platforms that enhance healthcare operations and promote inclusive healthcare (Cohen et al. 2022, Qiu et al. 2021). The results have operational implications for stakeholders within the mental healthcare ecosystem. In particular, our findings suggest that to ensure equitable delivery of services, (i) MHMA firms should develop apps that include a variety of self-support and peer-support functions; and (ii) organizations (e.g., United Nations, World Health Organization, governments, firms) should expand the delivery of mental health services via MHMAs to "hard-to-reach" patient population.

The remainder of the paper is organized as follows. In section 2, we present a review of the relevant literature and develop the study hypotheses. In section 3, we discuss the empirical setting, data, and variables. In section 4, we present the econometric model specification and report the main results of the empirical analysis. In section 5, we present the post-hoc analysis results to delve into the more nuanced implications of MHMAs. Finally, in section 6, we conclude with an overview of the key study findings, practical implications, and future research directions.

2. Literature Review and Hypothesis Development

Below, we review the relevant literature and develop study hypotheses pertaining to the equity of usage and benefit of MHMAs.

2.1 Mental Healthcare Operations Management Literature

In contrast to the well-established body of literature on physical healthcare operations management, the literature on mental healthcare operations management is still at a nascent stage and is comprised of only a few published studies. Among these studies, Yan and Tan (2014) find that social support exchange within a mental health online community can improve individuals' mental condition. In a follow-up study, Yan et al. (2019) demonstrate that treatment experiences shared by the online community members can influence how a member perceives her own mental health treatment. Zepeda and Sinha (2016) find that enhancing quality and affordability can particularly benefit the underserved population with socio-economic disadvantages. In a recent study, Li et al. (2021) find that eHealth platforms can enable mental health professionals to better-schedule follow-up visits for patients. Although these studies make valuable contributions to the literature on mental healthcare operations management, we are not aware of any study addressing the inequities associated with socio-demographic characteristics of individuals with mental health needs. To advance this emerging stream of literature, our study focuses on the inequities in usage and benefits of mental health services and examines the potential of MHMAs to provide equitable mental health services for the underserved population.

Inequities in mental health service usage and benefits can be attributed to several factors: First, there are significant variations across populations with different socio-demographic characteristics in terms of affordability, accessibility, and awareness (i.e., 3As that are strongly associated with the consumption of healthcare services (Kohnke et al. 2017, Sinha and Kohnke 2009)) of mental health services (Berger et al. 2005, Beyer 2020, Saxena et al. 2007, Sussman et al. 1987). For example, compared to the Whites, African-Americans and other minority race-ethnic sub-populations face challenges in affording and accessing mental health services due to lower income and education levels, limited insurance coverage, and geographical barriers (Beyer 2020). Furthermore, African-Americans exhibit lower awareness regarding the potential need for such services, often considering symptoms of mental illness as normal aspects of their daily life (Sussman et al. 1987). Similarly, males are less likely than females to be aware of their mental health needs, primarily due to the greater difficulties males experience in recognizing and acknowledging their emotional problems (Berger et al. 2005, Davis et al. 2023).

The second contributing factor is the presence of self-stigma and social-stigma associated with seeking care for mental health conditions, which distinguishes it from seeking care for physical health conditions (Masuda et al. 2012, Saxena et al. 2007). Self-stigma refers to the fear of self-disclosing distressing or potentially embarrassing personal information, whereas social-stigma refers to the societal stereotype that viewing individuals seeking mental health services as unpredictable, permanently damaged, incompetent, or threatening (Masuda et al. 2009, 2012). Such stigmas are particularly prominent within the underserved population. For instance, compared to the White individuals, African-American, Asian, and Hispanic/Latino individuals are less likely to acknowledge their mental health needs (Eisenberg et al. 2009, Lipson et al. 2018, Masuda et al. 2009, 2012). Additionally, even if they acknowledge their needs, they have less trust in mental health services (Cooper et al. 2003, Eisenberg et al. 2009). Similarly, compared to female individuals, male individuals are more inclined to perceive mental health treatment as undermining their social power and control (Berger et al. 2005, Davis et al. 2023). Thus, they are less likely to seek mental health services and openly discuss their emotional conditions during treatment (Good et al. 1989, Good and Wood 1995).

Third, the process of diagnosing and treating mental health conditions primarily relies on verbal communication, making seamless communication between mental health patients and professionals crucial for effective delivery of mental health services (Alegría et al. 2002, Nguyen and Reardon 2013). Thus, mental health patients often prefer consulting with a health professional who shares similar socio-demographics characteristics such as gender, sexual orientation, or race-ethnicity in order to establish rapport and enhance understanding (Cooper et al. 2003). Otherwise, a socio-demographic mismatch between patients and mental healthcare providers can contribute to inequities in care. The underserved population is more likely to experience such mismatches than the better-served population because there are far fewer mental healthcare professionals from the underserved population than from the better-served population. For instance, in 2015, only 14% of clinical psychologists in the U.S. were Asian, Hispanic, or African-American, whereas the majority were White (Lin et al. 2018).

Finally, discriminatory practices, including denial of care or unfair diagnosis and treatment, can also contribute to inequities in mental health services. For instance, despite a significantly higher mental illness rate of 44.1% among sexual minorities (i.e., lesbian, gay, bisexual, transgender, and questioning individuals) compared to the national average rate of 20.6%, they are at least twice more likely to be denied mental health services than their better-served counterparts (i.e., heterosexual individuals) due to discrimination (NAMI 2021a). Such discrimination can foster long-term mistrust within the underserved population, which hinders their willingness to seek support. For instance, African-Americans exhibit a tendency to rely less on mental health services due to historical discrimination and mistreatment (Diala et al. 2000), opting to address their mental health issues independently rather than seeking professional help (Alegría et al. 2002).

2.2 Hypothesis Development

We contend that MHMAs have the potential to mitigate the inequities observed in traditional modes of seeking mental health support. First, MHMAs are often free or have low-cost subscription plans. In addition, access to MHMAs is not restricted by geographical, temporal, or physical constraints (Stephens-Reicher et al. 2011), eliminating the commute time and costs for their users. These features make MHMAs more affordable and accessible than in-person mental health services delivered by professionals. Second, MHMAs offer the benefit of anonymity, creating a psychologically safe environment for app users. This allows users who are hesitant or uncomfortable discussing their mental health conditions to freely share their issues with their peers or seek mental health-related information without disclosing personal details such as gender, race-ethnicity, or sexual orientation (Ancis 2020), thereby minimizing the self-stigma and social-stigma. Third, MHMAs can be designed to support various languages and user interfaces, facilitating effective conversations among users. This design feature allows individuals with similar socio-demographic characteristics and mental health concerns to connect with peers, reducing socio-demographic mismatches on the platform. Lastly, by implementing design features like active monitoring of online communities by human moderators or artificial intelligence algorithms to detect and address discriminatory, harassing, or abusive content, MHMAs can foster high quality interactions, while minimizing the risk of discrimination (Wadden et al. 2021) (i.e., a potentially significant downside that other online platforms such as social media entail (Ancis 2020)). Overall, these features are likely to motivate individuals, irrespective of their socio-demographic characteristics, to seek mental health support more frequently through MHMAs compared to in-person mental health services delivered by professionals. Because smartphone ownership rates are similar between the underserved and better-served populations (e.g., 85% of whites, 83% of African-Americans, and 85% of Hispanics/Latinos own smartphones in 2021 (Pew Research Center 2021)) and people with different socio-demographic characteristics show comparable interest in utilizing MHMAs (Anderson-Lewis et al. 2018, Lipschitz et al. 2020, Samarasekera 2022), we anticipate that the opportunities for accessing mental health support via MHMAs apply equally to both populations. Therefore, we posit the following hypothesis:

Hypothesis 1 (H1): *The usage of a MHMA is equivalent between users from the underserved population and users from the better-served population.*

MHMAs also offer several advantages that can contribute to improved mental condition for users. First, most MHMAs include functions that enable users to track their mental condition over time using visual trend plots. These tools can alert users to any downward trends or significant fluctuations that may indicate a deterioration in their mental condition. By receiving automated alerts, users are likely motivated to seek mental health support expeditiously. Studies have shown that individuals who seek mental health support early are more likely to benefit from the support than individuals who delay seeking support (McGorry et al. 2006, McGorry and van Os 2013). Additionally, MHMAs facilitate peer-support by enabling users to connect with individuals experiencing similar mental health conditions. Specifically, through MHMAs, peers can learn from each other's experiences, offer immediate feedback, and exchange coping strategies (Yan and Tan 2014). In conclusion, MHMA usage has the potential to help users significantly reduce loneliness, enhance stress management, and improve depression management (Ancis 2020), leading to an improved mental condition (Donker et al. 2013). Thus, we posit the following hypothesis:

Hypothesis 2 (H2): *As the usage of a MHMA increases, users' mental condition improves.*

We infer from our review of the relevant literature that improving the affordability of mental health care within clinic-based services yields the greatest benefits in terms of improving mental condition, particularly for socio-economically disadvantaged communities (Zepeda and Sinha 2016). Additionally, peer-support has been shown to offer greater mental health benefits to individuals facing higher levels of stress, such as minorities defined by their sexual-orientation (Meyer 2003) or race-ethnicity (Aneshensel 2009). Along these lines, we contend that MHMAs, by virtue of their affordability and ability to facilitate peer-support among users, are also likely to contribute towards enhanced mental health benefits, particularly for the underserved population. In fact, the positive impact of MHMAs may even be more pronounced, as certain underserved sub-populations (e.g., males or African-Americans) tend to express their mental health concerns more openly within a mobile app (virtual) environment (McCall et al. 2021, Ritterband 2021), which is fundamental to realizing the benefits of mental health services (Pennebaker 1999). As a result, MHMAs can enable individuals from the underserved population to benefit from app usage at least as much as individuals from the better-served population. Thus, we hypothesize:

> ***Hypothesis 3 (H3):*** *As the usage of a MHMA increases, the mental condition of users from the underserved population improves at least as much as that of users from the better-served population.*

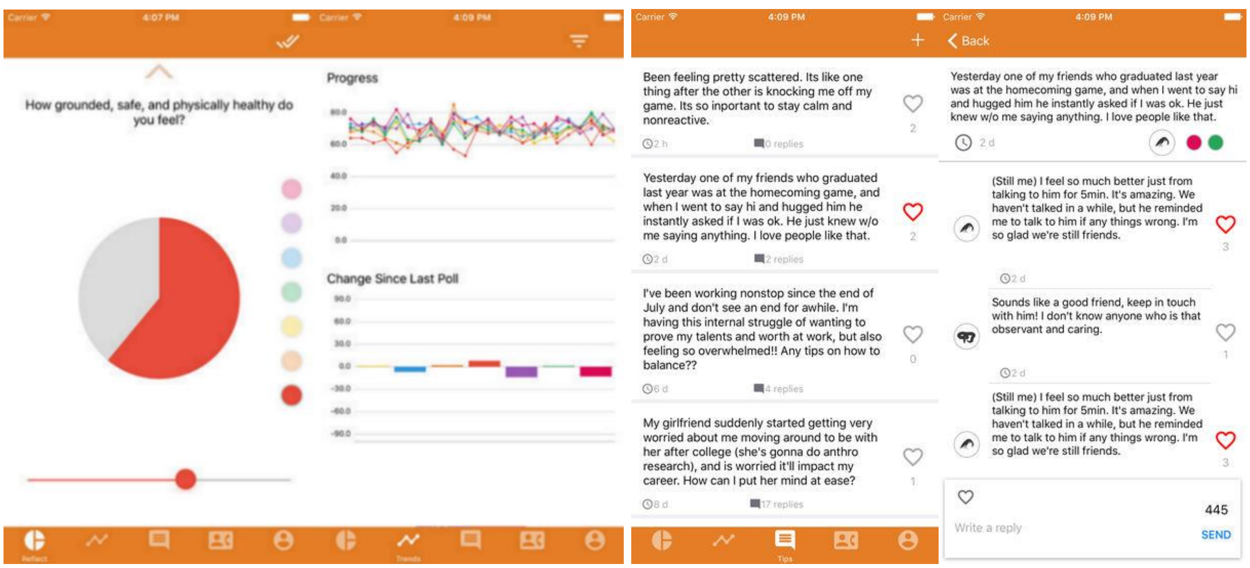
3. Research Design

3.1 Empirical Setting

In this study, we collaborate with a mobile app startup company to understand whether MHMAs can offer equitable self-support and peer-support across populations with different socio-demographic identities. In 2015, the company launched its MHMA, Hope, for public users. Hope was not designed or marketed to target any specific population. Rather, it was accessible by anyone with a smartphone. The Hope user community consists of users who are non-professionals (i.e., Hope is peer-based) and seek mental health support. Hope provides its users with two primary functions for self-support and peer-support:

Figure 1: Screenshot of Hope, a Mental Health Mobile App (MHMA)

(a) Self-reflection Function (b) Online Community Function



(i) **The *self-reflection function.*** This function allows users to self-assess their overall mental condition by responding to seven questions in Hope's Mental Condition Survey (refer to Online Appendix B for the survey questions). Once registered on Hope, users can use this function as frequently as they desire. This feature enables users to not only evaluate their current mental condition when needed, but also track the recorded results over time through a visual chart. This chart helps users understand the variations in their mental condition. Figure 1(a) demonstrates an example of the user interface and visual chart of this function.

(ii) **The *online community function***. This function establishes an anonymous online community within the app. Here, app users can connect and freely engage in peer-support at any time without the need to disclose their physical identities, minimizing the risk of users feeling compelled to present themselves differently for any reason. Within this online community, users can anonymously post to seek or provide help, share their feelings and personal stories, and interact with other users by responding to their posts. Figure 1(b) provides an example of user posts and replies within this function.

3.2 Data

Our study dataset is comprised of user-level socio-demographic and app activity data from 1,935 Hope users between May 2015 (i.e., the launch-month of the app) and February 2018. Hope records information on the socio-demographic characteristics of each user during the registration process. In addition, each time a user uses the self-reflection function or interacts with the online community function, Hope generates an activity record with detailed information such as the timestamp, user id, self-reflection scores, or the post/reply texts. Our sample consists of, on average, 5.91 self-reflection records, 1.08 posts, and 7.60 replies per user.

Using these records, we construct an unbalanced panel in which each user has one observation for every Hope's Mental Condition Survey completed using the self-reflection function. Hence, a row in the panel data includes variables for user i who uses the self-reflection function at time j. An important consideration when analyzing mobile app data is user retention. It is likely that some users that have been on a mobile app for a long period will have more data points than others who have been in the app for a shorter period. This discrepancy can introduce sampling bias, as users with more data points will be overrepresented in the sample. To address this issue and mitigate the sampling bias, we remove 92 outliers from the dataset, resulting in a sample of 1,843 users. These outliers represent users whose cumulative self-reflection function usage exceeds the 95th percentile (i.e., users who completed more than 22 Mental Condition Surveys and thus are overrepresented in the sample). Note that in this setting, a user must complete at least one Hope's Mental Condition Survey to be included in the panel. Among the 1,843 users, 1,688 meet this criterion. Thus, for simplicity, we initially conduct our analysis using app activity data from 1,688 users. Later, in a different specification in Online Appendix E.1, we consider all 1,843 users to evaluate the robustness of our results. Table 1 presents the distribution of the 1,843 users with respect to their socio-demographic identities including gender, sexual orientation, and race-ethnicity.

3.3 Variable Definitions and Measurements

For each observation in our panel, we construct the following variables.

3.3.1 Outcome Variables

Consistent with our hypotheses, we consider two outcome variables:

Table 1: Comparison of Socio-demographic Characteristics of App Users in the Study Sample

		Gender			Sexual Orientation			Race-Ethnicity				
	Total	Female	Male	Other	Heterosexual	Homosexual	Other	White	African	Asian	Latino	Other
N	1,843	1,270	487	86	1,235	153	455	1,502	67	122	131	21
%	100%	68.91%	26.42%	4.67%	67.01%	8.30%	24.69%	81.50%	3.64%	6.62%	7.11%	1.13%

SRS_{ij} *(Self-Reflection Score)* represents a user's mental condition and is operationalized as the score that resulted from the Hope's Mental Condition Survey completed by user *i* at time *j*. The survey includes seven questions, each measured on a continuous scale ranging from 0 to 100. Therefore, we construct *SRS* as a latent variable using those questions. Since these questions are designed and implemented by the partner company, the scale items do not align perfectly with existing scales in the literature. It is important to note, however, that these questions conceptually capture the same underlying constructs as the widely used *PHQ-9* (Patient Health Questionnaire with 9 items) and *GAD-7* (Generalized Anxiety Disorder scale with 7 items) scales, which are self-administered depression or anxiety disorder diagnostic instruments in mental health clinical practice. To evaluate the reliability of *SRS*, we conduct an exploratory factor analysis and a confirmatory factor analysis. The results are presented in Online Appendix C. The factor analysis indicates that the theoretical model between the latent variable and the seven scale items fits the data, and the scales are reliable. Thus, we operationalize the latent variable *SRS* as the weighted average of the seven survey items, where the weights are the standardized loadings in the confirmatory factor analysis.

As an additional step, we conducted a pilot study to assess the face validity of *SRS*. In this study, we recruited 61 college students at a public research university in the mid-western United States and asked the participants to first answer both *PHQ-9* and *GAD-7* questionnaires, and then complete Hope's Mental Condition Survey. As detailed in Online Appendix D, we found that *SRS* is strongly correlated with both *PHQ-9* and *GAD-7* scales, providing face validity for *SRS* in measuring the mental condition of Hope users.

$AppUsage_{ij}$ indicates the frequency of app *usage* and is operationalized as the natural logarithm of the number of times the app functions (i.e., both the self-reflection function and the online community function) are used by user *i* within 15 days prior to the Mental Condition Survey completed at time *j* (excluding the app usage at time *t*).[5] For example, a user who completes the survey twice, writes three posts, and responds to other's posts[6] four times within the last 15 days (i.e., $AppUsage_{ij} = log(1+2+3+4)$) is considered to have

[5] The average time between two consecutive Mental Condition Survey completions is 7.18 days with a minimum of 0 day, a maximum of 623 days, a standard deviation of 29.08, and the 90th percentile corresponding to 11.08 days.
[6] Following the literature that considers both giving and receiving nonprofessional, nonclinical support to be a part of peer-support (Yan and Tan 2014), we categorize any form of online community function engagement, including

a higher frequency of app usage than a user who completes the survey once, writes two posts, and responds to others' posts three times within the same time-period (i.e., $AppUsage_{ij} = log(1+1+2+3)$). We measure $AppUsage_{ij}$ by strictly excluding the app usage on the day the Mental Health Condition Survey is completed to ensure that there is no simultaneity between SRS_{ij} and $AppUsage_{ij}$ (i.e., akin to using lagged variables). We select the threshold of 15 days for the following reason. The effects of a single mental healthcare session such as therapy, tend to diminish over time, necessitating multiple sessions for effective treatment (Austin 2022, Thompson 2022). Therefore, mental healthcare typically follows a treatment course involving weekly, bi-weekly, or monthly sessions, depending on the severity of symptoms, nature of issues, and therapy types (Solara Mental Health 2023). By setting the *AppUsage* threshold at 15 days, we align with the bi-weekly frequency of mental healthcare delivery. Notably, this choice is consistent with our empirical mobile app setting, as 98.8% of app usages before taking a Mental Condition Survey in our panel occur within 15 days prior to taking that survey. Furthermore, we define *AppUsage* using different time-periods and validate the robustness of our findings in Online Appendix E.2.

3.3.2 Key Independent Variable

The key independent variable in our study is $\boldsymbol{Underserved_i}$, which indicates whether user *i* belongs to the underserved population ($Underserved_i=1$) or the better-served population ($Underserved_i=0$). Consistent with the literature documenting inequities in mental health services (Berger et al. 2005, Masuda et al. 2009, NAMI 2021a), we define a user as belonging to the underserved population if they do not identify as a female-heterosexual-White individual. Conversely, female-heterosexual-White users are categorized as representing the better-served population.[7]

3.3.3 Control Variables

We use several control variables that are potentially related to the outcome variables:

$\boldsymbol{Age_{ij}}$ represents the age of user *i* at time *j* and is used to control for the potential impact of technology usage patterns across different age groups on mobile app usage (Pew Research Center 2021) and mental condition (SAMHSA 2022).

$\boldsymbol{InitialSRS_i}$ is the first *Self-reflection Score* recorded for user *i*. We use each user's *InitialSRS* as the proxy for their initial mental condition when starting using the app. Users with better initial mental condition are likely to have higher app usage frequency. We control for this variable to rule out this potential impact of users' mental condition before joining the app on their usage behavior.

$\boldsymbol{Tenure_{ij}}$ indicates a user's membership length and is measured as the number of days between user *i*'s app registration day and time *j* when the Mental Condition Survey is completed.

responding to other peers, as part of a user's *AppUsage*.

[7] $Underserved_i$ is constructed based on user self-reported demographic data. Using self-reported data is a common practice in similar mental health online community studies (e.g., Yan et al. 2019, Yan and Tan 2014).

AppRemindFrequency$_i$ indicates the number of weekly notifications received by user i. This count variable captures the feature where users have the option to set, between 1 and 7, how many times a week Hope should send a notification to their smartphones to remind users that they can engage with the app.

Year-month fixed effect controls for any common trend and seasonality that might be related to app usage or mental condition. We operationalize this fixed effect using year-month indicators.

Table 2 presents descriptive statistics and the correlation matrix for the variables used in the study for users from the underserved population (*Underserved=1*) and users from the better-served population (*Underserved=0*). With a variance inflation factor score mean of 1.05 and a range between 1.02 and 1.09, below the rule-of-thumb cut-off of ten, there is little evidence of multi-collinearity.

Table 2: Descriptive Statistics and the Correlation Matrix

	Underserved = 0		*Underserved* = 1		Correlation Matrix						
Variable	Mean	SD	Mean	SD	1	2	3	4	5	6	7
1. *AppUsage*[ii]	1.21	1.09	1.38	1.10	1.00						
2. *SRS*	41.42	30.05	40.87	28.36	0.02†	1.00					
3. *Age*	23.70	8.15	23.37	7.82	-0.07***	0.12***	1.00				
4. *InitialSRS*	39.57	28.95	33.88	25.48	-0.10***	0.70***	0.19***	1.00			
5. *Tenure*[ii]	22.94	50.28	24.41	58.45	-0.02*	0.14***	0.08***	0.17***	1.00		
6. *AppRemindFrequency*	3.28	2.73	3.94	2.84	0.19***	-0.05***	-0.12***	-0.13***	-0.01	1.00	
7. *Underserved*	-	-	-	-	0.07***	-0.01	-0.02†	-0.10***	0.01	0.11***	1.00
Number of users	648		1,040								
Number of observations	2,628		4,814								

Notes: (i) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1
(ii) The *AppUsage* and *Tenure* variables in this table are repeated observations for users measured at the moment of a specific *SRS* record and are not the final app usage for users or the tenure measured at the last app activity.

4. Empirical Analysis

In this section, we introduce our econometric model specification and present the estimation results.

4.1 Model Specification

In the healthcare literature, the patient outcome is a pivotal outcome variable, holding equal importance for app developers who seek to establish the credibility of their apps towards improving mental condition of app users through usage. Therefore, in our study, we are particularly interested in whether there exist any differences between users from the underserved and better-served populations in: (i) *AppUsage* and (ii) *SRS* as a result of *AppUsage.* This implies that *AppUsage* is expected to influence *SRS*, while *SRS* is not anticipated to affect *AppUsage*. Following this assumption[8], we specify our econometric model as a recursive system of equations:

[App Usage Equation]

$$AppUsage_{ij} = \alpha_0 + \alpha_1 Underserved_i + \boldsymbol{UserControls_{ij}\alpha_2} + \boldsymbol{TimeControls_{ij}\alpha_3} + u_i + \varepsilon_{ij} \quad (1)$$

[8] We later relax this assumption in Online Appendix E.3 and find that our results hold when we allow simultaneity between *AppUsage* and *SRS*.

[SRS (Self-Reflection Score) Equation]:

$$SRS_{ij} = \beta_0 + \beta_1 AppUsage_{ij} + \beta_2 Underserved_i + \beta_3 AppUsage_{ij} \times Underserved_i + \boldsymbol{UserControls_{ij}\beta_4} + \boldsymbol{TimeControls_{ij}\beta_5} + v_i + \tau_{ij} \quad (2)$$

where ε_{ij} and τ_{ij} are random error terms. *UserControls* include *Age*, *InitialSRS*, *Tenure,* and *AppRemindFrequency*. *TimeControls* include year-month fixed effects. Lastly, u_i and v_i denote the random effects for user *i*. Since our interest is to identify the impact of the time-invariant variable *Underserved* on outcome variables, we specify our model as a random effect model. Later in section 4.3, we estimate our model using a fixed effect specification and find that our results related to *SRS* still hold. Note that, in our specification, *AppUsage* is a mediator and *Underserved* moderates the relationship between *AppUsage* and *SRS,* making our model a moderated mediation model. The coefficient of interest to test H1 is α_1 in Equation 1. Similarly, the coefficients of interest to test H2 and H3 are β_1 and β_3 in Equation 2, respectively. We specify Equations 1 and 2 as linear models.

4.2 Estimation Results

We estimate our econometric model using maximum likelihood estimation and present the results in Table 3. Model 1 demonstrates the estimation of Equation 2 (the SRS equation) with only control variables. Model 2 builds on Model 1 by adding *AppUsage* into the SRS equation. Model 3 further enhances Model 2 by including the mediation model for *AppUsage,* estimating both Equation 1 and Equation 2 simultaneously without the interaction term in Equation 2. Lastly, Model 4 represents the moderated mediation model, derived from Model 3 by adding the interaction term *AppUsage* × *Underserved* into the SRS equation.

We examine the significance of the incremental variance due to any added variable in the SRS equation by performing a likelihood ratio test that compares the model with the added variable and the model without the added variable. We use Model 3 to test H1 and H2 because these hypotheses are related to the main effects of *Underserved* (in the App Usage equation) and *AppUsage* (in the SRS equation). We test H3 using Model 4 as it includes the interaction term.

With respect to the equity in MHMA usage, the estimated coefficient of *Underserved* in the App Usage equation in Model 3 ($\alpha_1 = 0.06, p = 0.134$) is not statistically significant. Following the recent published literature that establishes racial equity in various contexts through demonstrating statistical insignificance (e.g., Cui et al. 2020, Ganju et al. 2020), this result may suggest equity in app usage between users from the underserved and better-served populations. However, the statistics literature indicates that statistical insignificance can arise due to not only statistical equivalence, but also statistical indeterminance (Tryon 2001). Therefore, to further examine the source of statistical insignificance in our empirical setting, we conduct a statistical equivalence test (Schuirmann 1987) by comparing the average app usage between users from the underserved population and the better-served population and find support for statistical equivalence ($t_1 = 5.79,\ \ p_1 = 0.00;\ t_2 = 2.21, p_2 = 0.01;\ \ \Delta = 0.2\sigma_{\overline{AppUsage}}$).[9] Combined with the estimated coefficient α_1 in Table 3, this result suggests that users from the underserved population use Hope at least as much as the users from the better-served population, supporting H1.

Table 3: Main Analysis Estimation Results

	MODEL 1	MODEL 2	MODEL 3	MODEL 4
APP USAGE EQUATION				
Underserved			0.06	0.06
			(0.04)	(0.04)
Age			-0.01**	-0.01**
			(0.00)	(0.00)
InitialSRS			-0.00	-0.00
			(0.00)	(0.00)
Tenure			-0.00	-0.00
			(0.00)	(0.00)
AppRemindFrequency			0.07***	0.07***
			(0.01)	(0.01)
SELF-REFLECTION SCORE (SRS) EQUATION				
AppUsage		2.59***	2.59***	1.91**
		(0.39)	(0.39)	(0.72)
Underserved	1.57*	1.39†	1.39†	0.57
	(0.72)	(0.72)	(0.72)	(0.60)
AppUsage × Underserved				1.01
				(0.85)
Age	0.00	0.02	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
InitialSRS	0.81***	0.81***	0.81***	0.81***
	(0.02)	(0.02)	(0.02)	(0.02)
Tenure	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
AppRemindFrequency	0.27†	0.08	0.08	0.09
	(0.14)	(0.14)	(0.14)	(0.14)
AIC	63739.42	63612.48	82825.99	82823.51
BIC	64009.1	63889.07	83372.26	83376.71
Likelihood Ratio Test	-	128.94***	-	4.47*
Number of users	1,688	1,688	1,688	1,688
Number of observations	7,442	7,442	7,442	7,442

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.
(iii) User random effects and year-month fixed effects are included in all models.

With respect to the benefit of MHMA usage, we find that the estimated coefficient of *AppUsage* in the SRS equation in Model 3 ($\beta_1 = 2.59, p < 0.001$) is positive and statistically significant, supporting H2.

[9] There is no universal guideline for determining the difference limit (i.e., Δ) in statistical equivalence tests. The literature often employs a standardized effect size or Cohen's d (i.e., a percentage of the standard deviation of the variable of interest) to determine Δ (e.g., $\Delta = 0.33\sigma$ in Master et al. (2016), and $\Delta = 0.45\sigma$ in Lakens et al. (2020)). Compared to the literature, we use a more conservative Δ (i.e., $\Delta = 0.2\sigma$), which implies that two users are considered to have an equivalent usage if the difference in their number of app usage within 15 days falls within 1.16. As a sensitivity analysis, we gradually decrease Δ (i.e., making it even more conservative) and find that the support for statistical equivalence is no longer present for $\Delta < 0.16\sigma$, which is considered a trivially small Δ to rule out statistical equivalence (Maxwell et al. 2015).

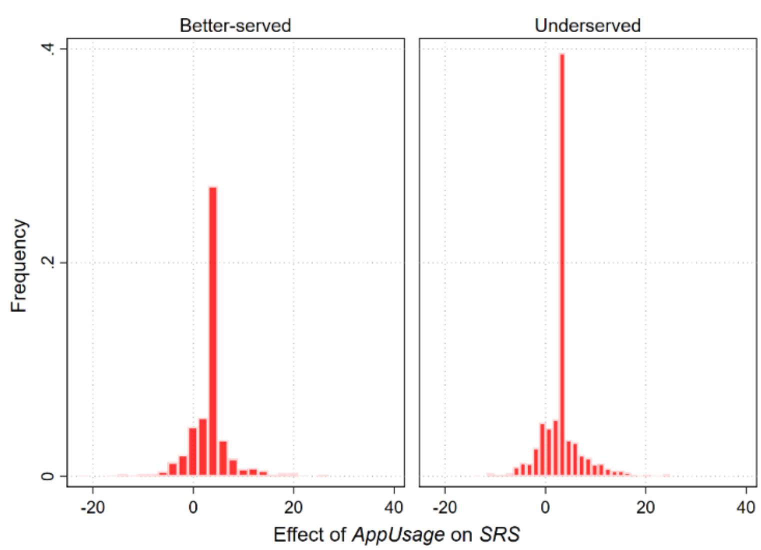
Based on the predictive margins of *SRS*, doubling app usage frequency is associated with an increase of 1.80 unit increase in *SRS.* For instance, such an increase would correspond to a 4.69% increase in self-assessed mental condition of a user with median *SRS* in our study sample, testifying the benefit of MHMA usage.

Finally, Model 4 estimation results indicate that the interaction term $AppUsage$ × *Underserved* ($\beta_3 = 1.01, p = 0.237$) is positive but not statistically significant. Note that unlike with the equity in usage in which we seek statistical equivalence between the two groups with respect to the variable *AppUsage*, the equity in benefit requires statistical equivalence between the two groups with respect to the effect of *AppUsage* on *SRS* (i.e., β_1 in Model 3, which is a slope rather than a variable). Thus, to examine the source of the insignificance of the interaction term coefficient, we first re-specify Equation 2 as a random slope model in which the effect of $AppUsage$ on *SRS* varies between users:

$$SRS_{ij} = \beta_0 + \beta_1 AppUsage_{ij} + \beta_2 Underserved_i + \boldsymbol{UserControls_{ij}\beta_4} + \boldsymbol{TimeControls_{ij}\beta_5} + \zeta_{0i} + \zeta_{1i} AppUsage_{ij} + \tau_{ij} \tag{3}$$

where ζ_{0i} denotes the deviation of user i's intercept from the mean intercept of β_0 and ζ_{1i} denotes the deviation of user i's slope for $AppUsage$ from the mean slope of β_1. With this specification, the effect of $AppUsage$ on *SRS* for user i is estimated to be $\beta_1 + \zeta_{1i}$. The significant log-likelihood ratio test statistic (i.e., $\chi^2(2) = 2833.47$, p=0.00) obtained from the estimated random slope model (not reported) indicates that the effect of $AppUsage$ on *SRS* indeed varies between users.

Figure 2: Estimation Results of the User-specific Effect of *AppUsage* on *SRS* (Self-reflection Score)



The histogram plots in Figure 2 summarize the estimated $\widehat{\beta_1 + \zeta_{1i}}$ for users from the underserved and better-served populations. Next, similarly to testing H1, we conduct a statistical equivalence test by comparing the estimated $\widehat{\beta_1 + \zeta_{1\iota}}$ between the two groups of users and find that there is statistical equivalence between them ($t_1 = 4.43, p_1 < 0.001;\ t_2 = 3.56,\ p_2 < 0.001; \Delta = 0.2\sigma_{\widehat{\beta_1+\zeta_{1\iota}}}$[10]). Combined with the estimated coefficient β_3 in Table 3, this result suggests that using MHMA improves the mental condition equally for users from the underserved population and users from the better-served population, providing support for H3.

We also conduct several robustness tests and demonstrate in Online Appendix E that our main results are consistent across several model specifications, variable operationalization, and alternative explanations.

5. Post-hoc Analysis

Having established the robustness of our results, we now conduct post-hoc analysis to understand: (i) how the two Hope (MHMA) functions (i.e., the self-reflection and online community functions) contribute to equity in usage and benefit; and (ii) whether this equity holds equally across different underserved sub-populations.

5.1 Self-reflection Function vs. Online Community Function

The two MHMA functions offer distinct types of support to app users. The self-reflection function enables users to monitor their mental condition and increase the awareness of their own mental healthcare needs without interacting with other users. In contrast, the online community function facilitates user interactions and the exchange of peer-support within Hope. It provides a platform for users to connect with one another, offering the opportunity to receive from and provide support to fellow users. Hence, the two app functions can be associated with two different user behaviors, with the self-reflection function emphasizing individual introspection and the online community function emphasizing interaction and peer-support.

To investigate whether there are differences in user behaviors between users from the underserved population and users from the better-served population, we conduct post-hoc analysis. This analysis expands on our main analysis by decomposing the *AppUsage* variable into two components. In particular, we operationalize *SRUsage*$_{ij}$ *(Self-Reflection Usage)* and *OCUsage*$_{ij}$ *(Online Community Usage)* as the natural logarithm of the number of times the self-reflection function and the online community function, respectively, are used by user *i* within 15 days prior to the Mental Condition Survey completed at time *j*. We then replace *AppUsage* in Equations 1 and 2 with *SRUsage* and *OCUsage*, and estimate our main model with each decomposed variable in isolation. Further, there is a potential interplay between *SRUsage* and *OCUsage*. Users who engage with the online community through the app may feel happier and more inclined to track their mental condition, leading to increased *SRUsage*. Similarly, users who frequently use

[10] As with H1, we examine the sensitivity of the statistical equivalence test for H3 by gradually decreasing Δ and find that it is no longer supported when $\Delta < 0.1\sigma$, a trivially small Δ to rule out statistical equivalence (Maxwell et al. 2015).

the self-reflection function may be more motivated to engage and assist others, resulting in increased *OCUsage*. To account for this interdependency, we include the lagged *OCUsage* variable as a control variable in the *SRUsage* equation and the lagged *SRUsage* variable as a control variable in the *OCUsage* equation.

Table 4: Estimation Results with Decomposed *AppUsage* Variables:
SRUsage:* Self-reflection Function Usage; *OCUsage: Online Community Function Usage

	Decomposed *AppUsage*			
	SRUsage		***OCUsage***	
	MODEL 3	MODEL 4	MODEL 3	MODEL 4
APP USAGE EQUATION				
Underserved	0.05†	0.05†	-0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)
Lagged SRUsage			0.35***	0.35***
			(0.03)	(0.03)
Lagged OCUsage	0.43***	0.43***		
	(0.02)	(0.02)		
Age	-0.00**	-0.00**	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
InitialSRS	0.00	0.00	-0.00**	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Tenure	-0.00*	-0.00*	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
AppRemindFrequency	0.05***	0.05***	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
SELF-REFLECTION SCORE (SRS) EQUATION				
SRUsage	3.58***	2.41*		
	(0.51)	(0.96)		
OCUsage			1.34**	1.79†
			(0.47)	(0.91)
Underserved	1.34†	0.22	1.54*	1.75*
	(0.71)	(0.60)	(0.72)	(0.69)
SRUsage × Underserved		1.73		
		(1.12)		
OCUsage × Underserved				-0.66
				(1.06)
Age	0.02	0.02	0.01	0.00
	(0.03)	(0.03)	(0.03)	(0.03)
InitialSRS	0.81***	0.81***	0.81***	0.81***
	(0.02)	(0.02)	(0.02)	(0.02)
Tenure	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
AppRemindFrequency	0.04	0.04	0.24†	0.24†
	(0.14)	(0.14)	(0.14)	(0.14)
AIC	78225.29	78218.75	78130.36	78131.21
BIC	78778.48	78778.86	78683.55	78691.32
Likelihood Ratio Test	-	8.54**	-	1.15
Number of users	1,688	1,688	1,688	1,688
Number of observations	7,442	7,442	7,442	7,442

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) User random effects and year-month fixed effects are included in all models.

As is evident in column *SRUsage* Model 3 in Table 4, we find that users from the underserved population use the self-reflection function marginally more frequently than users from the better-served population ($\alpha_1 = 0.05, p = 0.078$). We do not find a similar difference with respect to the usage of the online community function ($\alpha_1 = -0.003, p = 0.928$ in column *OCUsage* Model 3). With respect to the impact of the usage of each function on mental condition, we find that the use of both functions is associated with an increase in *SRS* ($\beta_1 = 3.58,\ p < 0.001$ in column *SRUsage* Model 3 vs. $\beta_1 = 1.34,\ \mathrm{p} = 0.005$ in column *OCUsage* Model 3). However, when we compare these two coefficients in a seemingly unrelated regression model, we find that, for the same frequency of use, the self-reflection function is associated with significantly more increase in *SRS* than the online community function ($\chi^2 = 18.18, p < 0.001$). This suggests that MHMAs are likely to offer value particularly through increasing awareness of users' mental health needs. When we examine the two functions with respect to the equity in benefit, we find no significant difference between users from the underserved and better-served populations $\beta_3 = 1.73,\ p = 0.121$ in column *SRUsage* Model 4 vs. $\beta_3 = -0.66,\ p = 0.536$ in column *OCUsage* Model 4). This suggests that the equity in benefit is present for both functions. However, considering that the coefficients of the two interaction terms have opposite signs, we find in a seemingly unrelated regression model that β_3 in the *SRUsage* model is significantly greater than β_3 in the *OCUsage* ($\chi^2 = 4.77, p = 0.03$). In other words, users from the underserved population benefit more from using the self-reflection function whereas users from the better-served population benefit more from using the online community function. This suggests that MHMAs are likely to offer value to: (i) users from the underserved population mainly through increasing their awareness of mental health needs and (ii) users from the better-served population more through facilitating peer-support among app users.

In summary, our post-hoc analysis of the two app functions reveals that the variety in mental health support offerings through MHMAs is likely to be key towards ensuring equity in usage and benefits. Such variety allows users from both underserved and better-served populations to leverage each of the two MHMA functions in ways that best align with their specific mental health needs.

5.2 Equity Across Users from Different Underserved Sub-populations

In our main analysis, we treat all users belonging to the underserved population (e.g., African-American, Asian, Hispanic/Latino, etc.) as similar in terms of the inequities they may face regarding usage and benefit of mental health services, without distinguishing between different socio-demographic characteristics. However, there is anecdotal evidence suggesting that the degree of inequity may vary across different underserved sub-populations. For intance, it is reported that African-American and Asian patients experience more discriminiation in mental health services than Hispanic/Latino patients (Horowitz et al. 2019). In order to obtain nuanced insights, we now extend our analysis to examine potential variations across users from different underserved sub-populations within a MHMA setting.

To identify users from different underserved sub-populations, we use three socio-demographic variables corresponding individually to: gender (*male, female, other gender*), sexual orientation

Table 5: Estimation Results for Different Underserved Sub-Populations

	Identification of constituent underserved sub-populations					
	Gender (Column 1)		Sexual Orientation (Column 2)		Race-ethnicity (Column 3)	
APP USAGE EQUATION						
Socio-demographic Characteristics:						
	Female-Underserved	0.07	*Heterosexual-Underserved*	-0.02	*White-Underserved*	0.08†
		(0.05)		(0.05)		(0.04)
	Male-Underserved	0.02	*Homosexual-Underserved*	0.06	*African-Underserved*	0.09
		(0.05)		(0.08)		(0.13)
	Other-Underserved	0.25*	*Other-Underserved*	0.15**	*Asian-Underserved*	0.03
		(0.10)		(0.05)		(0.08)
					Latino-Underserved	0.01
						(0.09)
					Other-Underserved	-0.15
						(0.18)
Age		-0.01*		-0.01*		-0.01**
		(0.00)		(0.00)		(0.00)
InitialSRS		-0.00		-0.00		-0.00
		(0.00)		(0.00)		(0.00)
Tenure		0.00		0.00		-0.00
		(0.00)		(0.00)		(0.00)
AppRemindFrequency		0.07***		0.07***		0.07***
		(0.01)		(0.01)		(0.01)
SELF REFLECTION SCORE (SRS) EQUATION						
AppUsage		1.91**		1.91**		1.91**
		(0.72)		(0.72)		(0.72)
Socio-demographic Characteristic Main Effects		Estimated		Estimated		Estimated
AppUsage × Socio-demographic Characteristics:						
	Female-Underserved	1.90†	*Heterosexual-Underserved*	1.04	*White-Underserved*	1.18
		(1.00)		(0.99)		(0.90)
	Male-Underserved	0.18	*Homosexual-Underserved*	1.47	*African-Underserved*	1.91
		(0.95)		(1.67)		(1.61)
	Other-Underserved	0.35	*Other-Underserved*	0.86	*Asian-Underserved*	-0.33
		(1.62)		(0.98)		(1.83)
					Latino-Underserved	0.65
						(1.68)
					Other-Underserved	-2.49
						(2.57)
Age		0.01		0.02		0.01
		(0.03)		(0.03)		(0.03)
InitialSRS		0.81***		0.81***		0.81***
		(0.02)		(0.02)		(0.02)
Tenure		0.01		0.01		0.01
		(0.01)		(0.01)		(0.01)
AppRemindFrequency		0.10		0.09		0.09
		(0.14)		(0.14)		(0.14)
AIC		82818.76		82826.28		82838.94
BIC		83413.44		83420.96		83475.11
Number of users		1,688		1,688		1,688
Number of observations		7,442		7,442		7,442

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) User random effects and year-month fixed effects are included in all models.

(*heterosexual, homosexual, other sexual orientation*), and race-ethnicity (*white, african, asian, latino, other race-ethnicicty*). In Equations 1 and 2, we substitute the variable *Underserved* with a categorical variable indicating different underserved sub-populations and the better-served population. For instance, when identifying users from underserved sub-populations based on gender, the categorical variable indicates whether a user belongs to the better-served population (i.e., female, heterosexual, White individual), the female-underserved sub-population (e.g., a female African-American), the male-underserved sub-population, or the other-gender-underserved sub-population (e.g., Hispanic individuals who do not associate themselves with any gender). We handle sexual orientation and race-ethnicity in similar ways. We then estimate our recursive system of equations (i.e., Model 4) with each of the three categorical variables (with users from the better-served population being the reference level) and present the results in Table 5.

From these results, the following two insights stand out. First, regardless of how we identify underserved sub-populations, we consistently find across all three model estimation results that none of the users from these sub-populations use and benefit from the MHMA significantly less than users from the better-served population. This suggests that the inequity reported in the prior literature for clinic-based mental health sevices (NAMI 2021a, Terlizzi and Norris 2021) is unlikely to be present in the context of MHMAs for users from any of the underserved sub-populations examined. Second, the results reported in the first column of Table 5 indicate that MHMAs might be particularly effective for female users within the underserved population (e.g., female African-American, female homosexual, etc.). The relationship between app usage and self-reflection score is marginally stronger for these users compared to users from the better-served population. We do not find such nuanced variation for the equity in benefit for users from other underserved sub-populations, despite some of them (i.e., users from underserved sub-populations identified by other gender identity, other sexual orientation, and White race-ethnicity) using MHMAs more frequently than users from the better-served population. In summary, our analysis reveals variations in the usage and benefit of MHMA-enabled mental health support across users from different underserved sub-populations. Nonetheless, these variations are unlikely to result in inequities in usage and benefit for users from any underserved sub-populations when compared to users from the better-served population.

6. Conclusion

6.1 Overview

In this paper, we evaluate the potential of MHMAs to provide equitable self-support and peer-support mental health services across users from the underserved and better-served populations. It is well-documented in the literature that the traditional clinic-based mental health services cannot be systematically accessed by certain sub-populations based on their socio-demographic characteristics such as gender, sexual orientation, and race-ethnicity, leading to inequities in usage of and benefit from mental health services

(NAMI 2021a, Terlizzi and Norris 2021). COVID-19 has further highlighted these inequities (Wang et al. 2020). Anecdotal evidence suggests that MHMAs that facilitate self-support and peer-support have the potential to offer equitable mental health support to the underserved population (Aziz et al. 2022, Emerson et al. 2021). Our study examines this potential using longitudinal user-level data on socio-demographic characteristics and app activity data collected from a MHMA. By way of research design, we estimate a recursive system of equations to assess whether equity in MHMA usage and benefit exists between users from the underserved and better-served populations. Further, we investigate: (i) whether users from the underserved and better-served populations use the two MHMA functions – namely, the self-reflection function and the online community function – similarly or differently; and (ii) whether there exists any difference in equity in MHMA usage and benefits from such usage across users from different underserved sub-populations.

6.2 Contributions to the Literature

By way of contributions, our study provides, to the best of our knowledge, the first theoretically-grounded empirical analysis that highlights the potential of MHMAs in delivering equitable mental health services – in the form of self-support and peer-support – to the underserved population. Specifically, we examine equity with respect to two user outcomes, namely the frequency of app usage (i.e., equity in usage) and mental condition (i.e., equity in benefit). With respect to equity in usage, we find that users from the underserved population exhibit MHMA usage frequencies statistically equivalent to those from the better-served population. This suggests that the inequity in clinic-based mental health service usage reported in the literature (Berger et al. 2005, Masuda et al. 2009, NAMI 2021a) is unlikely in MHMAs designed particularly for self-support and peer-support. In addition, our results indicate that MHMA usage frequency has a positive impact on users' mental condition. For instance, considering users with a median self-reflection score from our sample, doubling their app usage frequency (i.e., from 0.2/day to 0.4/day) can lead to a 4.69% increase in self-reflection score. With respect to equity in benefit, we find that MHMAs are equally beneficial for users from both underserved and better-served populations, as our results indicate that the association between usage frequency and mental condition is statistically equivalent between the two populations. It is worth noting that these novel insights with regards to equity in mental health support offered by MHMAs are relevant even when we consider specific underserved sub-populations (e.g., African-American underserved, female underserved, homosexual underserved, etc.) in isolation. As such, our post-hoc analysis demonstrates that, when considered in isolation, users from none of the underserved sub-populations experience lower usage and benefit from MHMAs compared to users from the better-served population. Conversely, users from certain underserved sub-populations (e.g., female African-American, female homosexual, etc.) tend to derive greater benefit from MHMAs compared to users from the better-served population.

To shed light on how MHMAs uphold equity in usage and benefit between users from the underserved and better-served populations, we conduct a post-hoc analysis to compare the usage of the app functions (i.e., the self-reflection and the online community functions) between the two groups of users. The results suggest that equity in MHMA usage and benefit is likely to be driven by the heterogeneity in users' preferences for different app functions. Specifically, we find that users from the underserved population use the self-reflection function marginally more than users from the better-served population. Therefore, users from the underserved population derive benefit from MHMAs more through the use of the self-reflection function, whereas users from the better-served population derive benefit from MHMAs more through the use of the online community function. The self-reflection function enables users to develop self-awareness of their mental health needs and can be used individually without the need for interaction with other users. The online community function allows users to exchange peer-support with each other and requires virtual interactions. Given the distinct nature of these two functions, it is conceivable that MHMAs will likely offer value to the underserved population primarily by enhancing their self-awareness of mental conditions and needs. In contrast, MHMAs will likely offer value to the better-served population primarily by facilitating a psychologically safe environment for peer-support interactions.

6.3 Practical Implications

A significant implication of our study findings is that MHMAs designed to deliver self-support and peer-support mental health services have the potential to address the well-documented inequity across the underserved and better-served populations in the traditional clinic-based mental healthcare settings. As detailed in Online Appendix F, fundamental to this implication is the identification of two types of relationship between the app users and the traditional mental health patient population. Both relationships can shed light on our results and suggest distinct implications of our study findings for stakeholders within the mental healthcare ecosystem.

First, MHMA users can inherently be different from the population seeking traditional clinic-based professional support. Hence, MHMA users can be those who do not seek traditional clinic-based professional support, yet seek peer-support and self-support help on their own in the app. This implies that MHMAs may expand the mental health services (through self-support and peer-support) to a new population segment who otherwise would not seek traditional clinic-based professional support. This explanation highlights the value of MHMAs in providing access to the "hard-to-reach" mental health patient population in the traditional clinic-based mental healthcare delivery settings, which can have several practical implications when combined with our study results. For instance, several global organizations such as United Nations and WHO consider improving access and advancing equity as the core guiding principle for achieving sustainable mental healthcare delivery (World Health Organization 2013). Our study findings suggest that by way of supporting initiatives directed towards enhancing access and advancing equity, global organizations and local governments should consider promoting and investing in enhancing the use and access to MHMAs within the underserved population, particularly in societies where there is high level of discrimination and violence against underserved females. Next, notwithstanding the potential for significant lost earnings on account of mental health-related disabilities of employees, many organizations are still hesitant to provide mental healthcare benefits and resources to their employees due to high upfront costs (Brodey 2021). In organizations that do provide such benefits and resources, employees with most needs (e.g., those from the underserved population) tend to avoid using mental health services due to social stigma and fear of losing their jobs (Brodey 2021), resulting in significant wastage of resources. Our study suggests that MHMAs can be an alternative value-added resource for organizations due to their relatively low upfront costs and the anonymity that MHMAs can offer to employees regardless of their socio-demographic characteristics. By expanding mental health services via MHMAs to employees who might not use mental health benefits in the traditional clinic-based delivery settings otherwise, organizations can enhance employee welfare, reduce wastage of investments in providing mental health benefits, and minimize lost earnings due to mental health-related disabilities.

Second, MHMA users can inherently be the same as the population seeking traditional clinic-based professional support. This implies that the self-support and peer-support functions of MHMAs can complement the delivery of mental health services for those seeking traditional clinic-based professional support. This scenario highlights the value of MHMAs in providing an alternative channel for seeking help and potentially facilitating behavioral changes among underserved population with the following implications. Specifically, our results demonstrate that MHMAs are likely to achieve equity by providing variety in app functions such as self-support and peer-support functions, which can induce different user behaviors. Therefore, mobile app firms should design apps with a variety of functions including both self-support and peer-support. Self-support functions such as Hope's self-reflection function can enable users to track their mental condition and conduct skill-training functions (i.e., to teach users coping or thinking skills) without interacting with other users (NIH 2019). Peer support functions such as Hope's online community function enable users to seek informational support or emotional support by virtually connecting and interacting with other app users. Next, to ensure safety and effectiveness of the app functions, governmental regulatory agencies such as the United States Food and Drug Administration have formed new divisions to develop protocols and guidelines for mobile app firms. Our study suggests that, to ensure the delivery of equitable mental healthcare services to the underserved population, in addition to safety and effectiveness, regulatory agencies should also consider developing market standards for designing MHMAs that encompass functions aligning with diverse behavioral needs.

All in all, regardless of these two likely relationships between the app users and the traditional mental health patient population, our results highlight that MHMAs may have the potential to address equity in mental healthcare delivery to the underserved population. Understanding how this potential can be realized will require further research. In particular, there is a need to investigate the interaction between MHMAs and the traditional clinic-based mental healthcare setting to shed light on questions such as: (i) how MHMAs can advance equity in mental health service delivery and (ii) whether adopting MHMAs (versus not adopting MHMAs) changes mental health patients' access to and use of the traditional clinic-based mental health services. Being the first to highlight the potential of MHMAs designed to provide self-support and peer-support services, we believe our study will motivate future inquiries addressing the above two questions that have consequential health, social, and economic implications for populations at large.

6.4 Limitations and Future Research Directions

As with any paper, our paper, too, has limitations. However, notwithstanding the limitations, our paper provides the motivation and lays the groundwork for future research aimed at sustaining equitable mental health service delivery, as discussed above. First, our empirical setting consists of a MHMA operated by a mobile app firm that does not involve mental health professionals. By involving professionals, MHMAs may offer more benefits by enabling them to detect at-risk patients based on their app activities and provide timely interventions and treatments. Yet, MHMAs operated by professionals may also provide users with the feeling of close monitoring, and thus exacerbate their feelings of stigma and inequity. Therefore, future research should: (i) examine how the impact on equity in usage and benefit of MHMAs changes with or without involving professionals; and (ii) identify conditions, including detailed patient characteristics, under which MHMAs should be integrated into the traditional mental health services while continuing to improve equity in usage and benefit.

Second, while we find empirical support for a positive relationship between MHMA usage and a user's mental condition, we do not investigate how mobile app firms can increase MHMA usage. A direction of future research could be to explore the effectiveness of different strategies (e.g., monthly subscription, sending notifications, using certain incentives) to improve usage of both the self-support and peer-support functions.

Third, in addition to mental condition tracking and peer engagement, it is conceivable that MHMAs can be leveraged to offer proven therapies such as cognitive behavioral therapy or acceptance commitment therapy to treat mental health disorders. A direction for future research could be to investigate whether these MHMA-enabled therapies improve mental conditions as well as advance equity and inclusion in care delivery to the underserved population.

Fourth, users from the underserved population account for more than 50% of all users in the MHMA that serves as the empirical setting of our study. This proportion is typically greater than the proportion of patients from the underserved population in the traditional clinic-based mental healthcare setting (NAMI 2021a) potentially for the following reasons: (i) the underserved population consider MHMAs to be attractive, given their effectiveness in providing access to "hard-to-reach" mental health patients; and/or (ii) the better-served population are likely to be less interested in seeking help via MHMAs. Future research could explore why the underserved population is overrepresented in MHMA settings.

Finally, as with many observational studies on online communities (e.g., Yan et al. 2019, Yan and Tan 2014), our study relies on self-reported data on socio-demographic characteristics. Since MHMA users do not have to disclose their physical identities because the online communities in MHMA settings are anonymous, they are likely to report their socio-demographic characteristics accurately. Despite this conjecture, future research could explore what factors (e.g., distrust in any online platform, need for role playing, etc.) may motivate (if any) MHMA users to report their socio-demographic characteristics inaccurately.

Acknowledgements

The authors gratefully acknowledge the financial support of: (i) Cisco's Tech for Healthcare Research Grant and (ii) University of Minnesota's Leadership in Equity, Inclusion and Diversity (LEID) Fellowship for conducting this study. The authors would like to express their sincere gratitude to: (i) Kalyan Singhal, editor, for his support and guidance throughout the review process and (ii) the three anonymous reviewers for their detailed and developmental comments on earlier versions of the paper.

References

Agic B (2019) Promising practices in equity in mental healthcare: Health equity impact assessment. *Healthcare Papers* 18(2):42–47.

Alegría M, Canino G, Ríos R, Vera M, Calderón J, Rusch D, Ortega AN (2002) Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino whites. *Psychiatric Services* 53(12):1547–1555.

Ancis JR (2020) The age of cyberpsychology: An overview. *Technology, Mind, and Behavior* 1(1).

Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S (2018) mHealth technology use and implications in historically underserved and minority populations in the united states: Systematic literature review. *Journal of Medical Internet Research mHealth and uHealth* 6(6):e8383.

Aneshensel CS (2009) Toward explaining mental health disparities. *Journal of Health and Social Behavior* 50(4):377–394.

Austin J (2022) 5 reasons why therapists ask you to attend multiple therapy sessions. *MedicallyInfo*. Retrieved (June 20, 2023), https://medicallyinfo.com/health/5-reasons-why-therapists-ask-you-to-attend-multiple-therapy-sessions/.

Aziz M, Erbad A, Almourad MB, Altuwairiqi M, McAlaney J, Ali R (2022) Did usage of mental health apps change during COVID-19? A comparative study based on an objective recording of usage data and demographics. *Life* 12(8).

Berger JM, Levant R, McMillan KK, Kelleher W, Sellers A (2005) Impact of gender role conflict, traditional masculinity ideology, alexithymia, and age on men's attitudes toward psychological help seeking. *Psychology of Men and Masculinity* 6(1):73–78.

Beyer D (2020) *The economic state of Black America in 2020* United States Congress Joint Economic

Committee, ed.

Chow JCC, Jaffee K, Snowden L (2003) Racial/ethnic disparities in the use of mental health services in poverty areas. *American Journal of Public Health* 93(5):792–797.

Cohen MC, Huang N, Li M, Pavlou PA (2022) Special issue of Production and Operations Management on "Responsible Data Science" . *Production and Operations Management* 31(6):2760–2761.

Cooper LA, Gonzales JJ, Gallo JJ, Rost KM, Meredith LS, Rubenstein L V., Wang NY, Ford DE (2003) The acceptability of treatment for depression among African-American, Hispanic, and white primary care patients. *Medical Care* 41(4):479–489.

Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* 66(3):1071–1094.

Davis K, Mwamuka R, Magnúsdóttir E, Brent G (2023) How men's mental health resources can help serve all communities. *Stanford Social Innovation Review*. Retrieved (July 1, 2023), https://doi.org/10.48558/34C9-B118.

Diala C, Muntaner C, Walrath C, Nickerson KJ, LaVeist TA, Leaf PJ (2000) Racial differences in attitudes toward professional mental health care and in the use of services. *American Journal of Orthopsychiatry* 70(4):455–464.

Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H (2013) Smartphones for smarter delivery of mental health programs: A systematic review. *Journal of Medical Internet Research* 15(11):1–13.

Dorwart L (2023) Best mental health apps. *Very well mind*. Retrieved (June 20, 2023), https://www.verywellmind.com/best-mental-health-apps-4692902.

Eisenberg D, Downs MF, Golberstein E, Zivin K (2009) Stigma and help seeking for mental health among college students. *Medical Care Research and Review* 66(5):522–541.

Emerson MR, Harsh Caspari J, Notice M, Watanabe-Galloway S, Dinkel D, Kabayundo J (2021) Mental health mobile app use: Considerations for serving underserved patients in integrated primary care settings. *General Hospital Psychiatry* 69:67–75.

Fleck DE, Keck PE, Corey KB, Strakowski SM (2005) Factors associated with medication adherence in African American and white patients with bipolar disorder. *Journal of Clinical Psychiatry* 66(5):646–652.

Ganju KK, Atasoy H, McCullough J, Greenwood B (2020) The role of decision support systems in attenuating racial biases in healthcare delivery. *Management Science* 66(11):5171–5181.

Good GE, Dell DM, Mintz LB (1989) Male role and gender role conflict: Relations to help seeking in men. *Journal of Counseling Psychology* 36(3):295–300.

Good GE, Wood PK (1995) Male gender role conflict, depression, and help seeking: Do college men face double jeopardy? *Journal of Counseling & Development* 74(1):70–75.

Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA (2005) Health information-seeking behaviour in adolescence: The place of the internet. *Social Science and Medicine* 60(7):1467–1478.

Hegland JE, Nelson NJ (2002) Cross-dressers in cyber-space: Exploring the internet as a tool for expressing gendered identity. *International Journal of Sexuality & Gender Studies* 7(2/3):139–161.

Horowitz J, Brown A, Cox K (2019) Race in America 2019. Retrieved (October 1, 2021), https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2019/04/Race-report_updated-4.29.19.pdf.

Kalkanci B, Rahmani M, Toktay LB (2019) The role of inclusive innovation in promoting social sustainability. *Production and Operations Management* 28(12):2960–2982.

Kirmayer LJ, Jarvis G (2019) Culturally responsive services as a path to equity in mental healthcare. *Healthcare Papers* 18(2):11–23.

Kohnke EJ, Mukherjee UK, Sinha KK (2017) Delivering long-term surgical care in underserved communities: The enabling role of international NPOs as partners. *Production and Operations Management* 26(6):1092–1119.

Lakens D, McLatchie N, Isager PM, Scheel AM, Dienes Z (2020) Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology: Series B* 75(1):45–57.

Li M (Manqi), Nassiri S, Liu X, Ellimoottil C (2021) How Does Telemedicine Shape Physician's Practice in Mental Health? *Available at SSRN*. Retrieved (June 28, 2021), https://papers.ssrn.com/abstract=3766619.

Lin L, Stamm K, Christidis P (2018) How diverse is the psychology workforce? Retrieved (September 1, 2021), https://www.apa.org/monitor/2018/02/datapoint.

Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE (2020) Patient interest in mental health mobile app interventions: Demographic and symptom-level differences. *Journal of Affective Disorders* 263:216–220.

Lipson SK, Kern A, Eisenberg D, Breland-Noble AM (2018) Mental health disparities among college students of color. *Journal of Adolescent Health* 63(3):348–356.

Master A, Cheryan S, Meltzoff AN (2016) Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology* 108(3):424.

Masuda A, Anderson PL, Edmonds J (2012) Help-seeking attitudes, mental health stigma, and self-concealment among African American college students. *Journal of Black Studies* 43(7):773–786.

Masuda A, Anderson PL, Twohig MP, Feinstein AB, Chou YY, Wendell JW, Stormo AR (2009) Help-seeking experiences and attitudes among African American, Asian American, and European American college students. *International Journal for the Advancement of Counselling* 31(3):168–180.

Maxwell SE, Lau MY, Howard GS (2015) Is psychology suffering from a replication crisis?: What does "failure to replicate" really mean? *American Psychologist* 70(6):487–498.

McCall T, Ali MO, Yu F, Fontelo P, Khairat S (2021) Development of a mobile app to support self-management of anxiety and depression in African American Women: Usability study. *Journal of Medical Internet Research Formative Research* 5(8):e24393.

McGorry P, van Os J (2013) Redeeming diagnosis in psychiatry: Timing versus specificity. *The Lancet* 381(9863):343–345.

McGorry PD, Hickie IB, Yung AR, Pantelis C, Jackson HJ (2006) Clinical staging of psychiatric disorders: A heuristic framework for choosing earlier, safer and more effective interventions. *Australian and New Zealand Journal of Psychiatry* 40(8):616–622.

Meyer IH (2003) Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence. *Psychological Bulletin* 129(5):674–697.

Michael L (2017) White male suicide: The exception to privilege. *Yale Global Health Review*. Retrieved (September 25, 2023), https://yaleglobalhealthreview.com/2017/05/14/white-male-suicide-the-exception-to-privelege/.

Mnookin S (2016) *Out of The Shadows: Making Mental Health a Global Development Priority* (Washington, D.C.).

Mongelli F, Georgakopoulos P, Pato MT (2020) Challenges and opportunities to meet the mental health needs of underserved and disenfranchised populations in the United States. *Focus* 18(1):16–24.

Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, Cannon M, et al. (2020) How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry* 7(9):813–824.

Murray CJL, Mokdad AH, Ballestros K, Echko M, Glenn S, Olsen HE, Mullany E, et al. (2018) The state of US health, 1990-2016: Burden of diseases, injuries, and risk factors among US states. *Journal of the American Medical Association* 319(14):1444–1472.

NAMI (2021a) Mental health by the numbers National Alliance on Mental Illness, ed. Retrieved (August 28, 2021), https://www.nami.org/mhstats.

NAMI (2021b) Mental health information. Retrieved (December 1, 2021), https://www.nimh.nih.gov/health/statistics/mental-illness.

Nguyen D, Reardon LJ (2013) The role of race and english proficiency on the health of older immigrants. *Social work in health care* 52(6):599–617.

NIH (2019) Technology and the future of mental health treatment. Retrieved (October 1, 2021), https://www.nimh.nih.gov/health/topics/technology-and-the-future-of-mental-health-treatment.

Pennebaker JW (1999) The effects of traumatic disclosure on physical and mental health: the values of writing and talking about upsetting events. *International Journal of Emergency Mental Health* 1(1):9–18.

Pew Research Center (2021) Mobile fact sheet. Retrieved (September 1, 2021), http://www.pewinternet.org/fact-sheet/mobile/.

Plöderl M, Tremblay P (2015) Mental health of sexual minorities. A systematic review. *International Review of Psychiatry* 27(5):367–385.

Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, Ruggiero KJ (2014) mHealth: A mechanism to deliver more accessible, more effective mental health care. *Clinical Psychology and Psychotherapy* 21(5):427–436.

Qiu L, Hong Y (Kevin), Whinston A (2021) Special issue of Production and Operations Management: Social technologies in operations. *Production and Operations Management* 30(5):1555–1556.

Rice SM, Purcell R, McGorry PD (2018) Adolescent and young adult male mental health: Transforming system failures into proactive models of engagement. *Journal of Adolescent Health* 62(3):S9–S17.

Rickwood DJ, Deane FP, Wilson CJ (2007) When and how do young people seek professional help for mental health problems? *Medical Journal of Australia* 187(S7):S35.

Ritterband V (2021) How men are finding mental health support with digital tools. *Harvard Pilgrim Health Care*. Retrieved (September 1, 2021), https://www.harvardpilgrim.org/hapiguide/how-men-are-finding-mental-health-support-with-digital-tools/.

Samarasekera U (2022) The rise of racial minority health apps. *The Lancet Digital Health* 4(4):e218–e219.

SAMHSA (2019) *Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health* (Rockville, MD).

SAMHSA (2021) Racial/ethnic differences in mental health service use among adults and adolescents (2015-2019). *Publication No. PEP21-07-01-002*. Retrieved (September 28, 2023), https://www.samhsa.gov/data/sites/default/files/reports/rpt35327/2021NSDUHMHChartbook.pdf.

SAMHSA (2022) *Key substance use and mental health indicators in the United States: Results from the 2021 national survey on drug use and health* (Rockville, MD).

SAMHSA (2023) Lesbian, gay, and bisexual behavioral health: Results from the 2021 and 2022 National Surveys on Drug Use and Health. *Publication No. PEP23-07-01-001*. Retrieved (September 29, 2023), https://www.samhsa.gov/data/sites/default/files/reports/rpt41899/2022_LGB_Brief_Final_06_07_23.pdf.

Satcher D (2001) Mental health: Culture, race, and ethnicity—A supplement to mental health: A report of the surgeon general. *US Department of Health and Human Services*. Retrieved (October 15, 2021), https://pubmed.ncbi.nlm.nih.gov/20669516/.

Saxena S, Thornicroft G, Knapp M, Whiteford H (2007) Resources for mental health: scarcity, inequity, and inefficiency. *Lancet* 370(9590):878–889.

Schuirmann DJ (1987) A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 15(6):657–680.

Semlyen J, King M, Varney J, Hagger-Johnson G (2016) Sexual orientation and symptoms of common mental disorder or low wellbeing: Combined meta-analysis of 12 UK population health surveys. *BMC Psychiatry* 16(1).

Sinha KK, Kohnke EJ (2009) Health care supply chain design: Toward linking the development and delivery of care globally. *Decision Sciences* 40(2):197–212.

Smith DG (2022) How to get more men to try therapy. *The New York Times* (December 9) https://www.nytimes.com/2022/12/09/well/mind/men-mental-health-therapy.html.

Solara Mental Health (2023) How often should I go to therapy? Is once a week enough? Retrieved (June 20, 2023), https://solaramentalhealth.com/how-often-should-i-go-to-therapy/.

Stephens-Reicher J, Metcalf A, Blanchard M, Mangan C, Burns J (2011) Reaching the hard-to-reach: How information communication technologies can reach young people at greater risk of mental health difficulties. *Australasian Psychiatry* 19(SUPPL. 1):S58--S61.

Sussman LK, Robins LN, Earls F (1987) Treatment-seeking for depression by black and white Americans.

Social Science and Medicine 24(3):187–196.

Terlizzi EP, Norris T (2021) Mental health treatment among adults: United States, 2020. *NCHS data brief* (419):1–8. Retrieved (October 15, 2021), https://www.cdc.gov/nchs/data/databriefs/db380-H.pdf.

The White House (2021) Exec. Order 13985: Advancing racial equity and support for underserved communities through the federal government. *Federal Register* 86(14):7009–7013. Retrieved (June 20, 2023), https://www.govinfo.gov/app/details/DCPD-202100054/.

Thompson S (2022) Why you need to attend multiple therapy sessions to improve your mental health. *The Art of Healthy Living*. Retrieved (June 20, 2023), https://artofhealthyliving.com/why-you-need-to-attend-multiple-therapy-sessions-to-improve-your-mental-health/.

Thornicroft G (2008) Stigma and discrimination limit access to mental health care. *Epidemiologia e Psichiatria Sociale* 17(1):14–19.

Tryon WW (2001) Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological methods* 6(4):371.

United Nations (2020a) Goal 3: Ensure healthy lives and promote well-being for all at all ages. Retrieved (August 28, 2021), https://www.un.org/sustainabledevelopment/health/.

United Nations (2020b) Policy brief: COVID-19 and the need for action on mental health. Retrieved (June 1, 2022), https://unsdg.un.org/resources/policy-brief-covid-19-and-need-action-mental-health.

US Department of Health and Human Services (2019) Serving vulnerable and underserved populations. Retrieved (June 20, 2023), https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/006_Serving_Vulnerable_and_Underserved_Populations.pdf.

Wadden D, August T, Li Q, Althoff T (2021) The Effect of Moderation on Online Mental Health Conversations. *Proceedings of the International AAAI Conference on Web and Social Media* 15:751–763.

Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, Ho RC (2020) Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health* 17(5):1729.

WEF (2019) The global risks report. *World Economic Forum*. Retrieved (August 28, 2021), http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf.

Wong YJ, Ho MHR, Wang SY, Miller ISK (2017) Meta-analyses of the relationship between conformity to masculine norms and mental health-related outcomes. *Journal of Counseling Psychology* 64(1):80–93.

World Health Organization (2013) Mental health action plan 2013-2020. Retrieved (October 1, 2021), https://apps.who.int/iris/handle/10665/89966.

Yan L, Tan Y (2014) Feeling blue? Go online: An empirical study of social support among patients. *Information Systems Research* 25(4):690–709.

Yan L, Yan X, Tan Y, Sun SX (2019) Shared minds: How patients use collaborative information sharing via social media platforms. *Production and Operations Management* 28(1):9–26.

Zepeda ED, Sinha KK (2016) Toward an effective design of behavioral health care delivery: An empirical analysis of care for depression. *Production and Operations Management* 25(5):952–967.

E-Companion for

Towards Achieving Mental Health Equity for the Underserved Population: Evaluating the Potential of Mobile Apps

Appendix A: A Review of MHMAs for Non-clinical Use

To assess the representativeness of Hope (i.e., the focal MHMA we study), we review all the MHMAs for non-clinical use that are listed in Dorwart's "Best Mental Health Apps" as of September 2023 (Dorwart 2023) with respect to the functions they offer and their scale in terms of # of downloads (i.e., Small scale: # of downloads is less than 10,000; Large scale: # of downloads is greater than 10,000). We expand this list by adding three more MHMAs for non-clinical use that, as with Hope, are not in Dorwart's list. Table A1 summarizes the function offerings across our focal app (highlighted in grey) and the 11 popular MHMAs reviewed. As seen in the table, all the listed 11 MHMAs offer self-support and/or peer-support functions. Among the 11 MHMAs, 10 offer a self-support function for different uses (e.g., mental health self-reflection, journals, knowledge sharing etc.) and 4 offer an online community function (e.g., chat groups, virtual groups, and anonymous online community). Similar to Hope, 3 MHMAs, all in Dorwart's list, offer both self-support and peer-support functions. In sum, we conclude that Hope is representative of MHMAs for non-clinical use in the market as of September 2023.

Table A1: A Listing of MHMAs with their Functions and Scale

App Name	Self-support Functions	Peer-support Functions	Scale (based on # of downloads)	Dorwart's List
Focal MHMA: Hope	Mental health self-reflection	Anonymous online community	Small	No
Huddleverse	N/A	Anonymous online community, messaging	Small	No
Iona: Mental Health Support	Mental health self-reflection, cognitive-behavioral therapy (CBT) exercises through chatbots	N/A	Small	No
Alkeme: Black Mental Health	Mental health knowledge sharing	N/A	Small	No
MoodKit	Mental health self-reflection, evidence-based activities rooted in CBT principles	N/A	Small	Yes
Worry Watch	Mental health self-reflection, guided anxiety journal, guided coping techniques, and guided positive reinforcement	N/A	Small	Yes
Breath, Think, Do with Sesame	Child problem-solving skill building and self-soothing strategy training	N/A	Small	Yes
I am Sober	Sobriety and recovery milestone tracking, sobriety calculator	Virtual groups	Large	Yes
Headspace	Mental health self-reflection, guided meditation, guided physical exercises	N/A	Large	Yes
Sanvello	Mental health self-reflection, guided meditation, CBT	Themed chat groups	Large	Yes
Calm	Mental health self-reflection, gratitude/sleep check-ins, daily calm reflection, guided meditations, grown-up bedtime stories, physical exercises	N/A	Large	Yes
Happify	Mental health self-reflection, strength assessment, games, meditation, and exercises	Anonymous online community	Large	Yes

Appendix B: Hope's Mental Condition Survey Items

Question	Scale
S1. (Sense of Self & Belonging) How connected, supported, comfortable, and included do you feel?	
S2. (Purpose & Emotional Clarity) How clear, purposeful, intentional, and intuitive do you feel?	
S3. (Decisions, Commitment & Reliability) How trusting, reliable, decisive, and committed do you feel?	
S4. (Relationships & Contentment) How open, accepting, and content do you feel?	**0-100[a]**
S5. (Work, Academics & Motivation) How influential, valuable, and capable do you feel?	
S6. (Stress & Emotional Well-being) How sparked, energized, and inspired do you feel?	
S7. (Sleep, Exercise & Nutrition) How grounded, safe, and physically healthy do you feel?	

[a] Larger number stands for a better self-evaluation on the corresponding question.

Appendix C: Factor Analysis for *SRS* (Self-reflection Score)

Based on the *PHQ-9* and *GAD-7* questionnaires, the social technology firm that developed the mental health mobile app, Hope, developed the Mental Condition Survey presented in Appendix A. We measure the latent variable *SRS* using those seven questions. We first perform an exploratory factor analysis using 10% of the sample and present the results in Table C1. We use principal axis factoring. All seven standardized loadings are above 0.7. The eigenvalue (5.38) and the Cronbach's α (0.9667) indicate one clear factor with high reliability, providing support for unidimensionality and convergent validity.

Table C1: Exploratory Factor Analysis for *SRS* (Self-reflection Score) Scale

Latent variable	Measurements	Standardized Loading	Eigenvalue	Cronbach's α
SRS	S1	0.7949	5.38442	0.9667
	S2	0.9871		
	S3	0.8924		
	S4	0.8262		
	S5	0.8573		
	S6	0.9290		
	S7	0.8375		

Note: Sample size = 742

Then, to examine the fit of the theoretical model, we perform a confirmatory factor analysis using the remaining data and present the results in Table C2. All goodness-of-fit statistics (*RMSEA=0.083, CFI=0.988, TLI=0.083 SRMR=0.027*) indicate good fit. The construct reliability (0.9665) for the latent variable SRS is greater than the cutoff value of 0.7. Average variance extracted (0.8048) is greater than 0.5, indicating reliability. The Cronbach's α (0.9663) suggests high scale reliability. All seven standardized loadings are above 0.8 and statistically significant at *p=0.000,* indicating convergent validity. In conclusion, the measures are reliable, valid, and support the theoretical model.

Table C2: Confirmatory Factor Analysis for the *SRS* (Self-reflection Score) Scale

Latent variable	Measurements	Standardized Loading	z-values	Cronbach's a	CR	AVE
SRS	S1	0.8271	282.23	0.9663	0.9665	0.8048
	S2	0.9112	544.74			
	S3	0.9278	655.51			
	S4	0.8944	463.38			
	S5	0.9028	500.3			
	S6	0.9240	627.92			
	S7	0.8888	411.44			

Notes: Sample size = 6,698. All loadings are significant at *p=0.000*. Goodness-of-fit measurements: *RMSEA=0.083, CFI=0.988, TLI=0.083 SRMR=0.027. CR:* construct reliability (suggested cutoff value 0.7); *AVE:* average variance extracted (suggested cutoff value 0.5); *RMSEA:* root mean squared error of approximation; *CFI:* confirmatory fit index; *TLI:* Tucker-Lewis index; *SRMR:* standardized root mean squared residual

Appendix D: Pilot Study to Assess the Face Validity of *SRS* (Self-reflection Score)

To further establish face validity for *SRS*, we conducted a pilot study between February 2017 and April 2017 to investigate how the mental condition measured by the Hope's Mental Condition Survey is comparable to that measured by the clinically approved *PHQ-9* and *GAD-7* questionnaires. For this study, we recruited 61 college students at a public research university in the mid-western United States. The participants were asked to first complete both *PHQ-9* and *GAD-7* questionnaires, and then register for the Hope app to complete Hope's Mental Condition Survey. We use the data from the pilot study to operationalize mental condition using *SRS*, *PHQ-9*, and *GAD-7*. Table D1 reports the descriptive statistics and correlation matrix for the three variables. We found that both *PHQ-9* and *GAD-7* scores have negative and significant correlations with *SRS.*

Table D1: Descriptive Statistics and the Correlation Matrix for Hope's *Self-reflection Score* and *PHQ-9 / GAD-7*

			Correlation Matrix		
Variable	Mean	SD	1	2	3
1. *SRS (Self-reflection Score)*	58.37	27.25	1.00		
2. *PHQ-9*	16.54	6.08	-0.65***	1.00	
3. *GAD-7*	13.84	5.18	-0.46***	0.75***	1.00
Number of observations	61				

Note: *** p<0.001, ** p<0.01, * p<0.05, † p<0.1

We also estimated two OLS regressions in which PHQ-*9* and *GAD-7* scores are regressed on *SRS*. As demonstrated in Table D2, the results indicate that Hope's *Self-reflection Score* is significantly and negatively associated with both *PHQ-9* ($\beta = -0.14, p < 0.001$) and *GAD-7* ($\beta = -0.09, p < 0.001$) scores. This is expected because an increase in *PHQ-9* or *GAD-7* scores represents a more severe condition related to depression and anxiety disorder, respectively, whereas an increase in *SRS* represents a better mental condition. Overall, the pilot study results provide face validity for Hope's *Self-Reflection Score* as a measure of mental condition.

Table D2: OLS Model Estimation Results for Evaluating the Association of Hope's *Self-reflection Score* with the *PHQ-9* and *GAD-7* Scores

VARIABLES	*PHQ-9 Score*	*GAD-7 Score*
SRS (Self-reflection Score)	-0.14***	-0.09***
	(0.02)	(0.02)
Constant	24.95***	18.89***
	(1.54)	(1.47)
Observations	61	61
R-squared	0.42	0.21

Robust standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05, † p<0.1

Appendix E: Robustness Checks

E.1 Endogeneity due to Sample Selection

Sample selection bias is a potential concern for studies on mobile apps (Bateman et al. 2011, Li and Hitt 2008) since registering to a mobile app is voluntary and not randomized. We use the Heckman model (Heckman 2013, Van de Ven and Van Praag 1981) to address this concern. For this model, in addition to our main sample (i.e., data from 1,688 users), we also use data from users who registered but never used Hope (i.e., 155 users who never completed Hope's Mental Condition Survey during the data period) with the assumption that those users can be seen as proxies for people who are in need for mental health support but do not register for Hope.[1] To provide face validity for this assumption, in Table E1, we statistically compare the 155 Hope users to the mental health patient population in the U.S. with respect to gender and race/ethnicity (i.e., the two socio-demographics statistics that are publicly reported).[2] We find that the socio-demographic differences between underserved and better-served groups among the 155 Hope users are not statistically different from those among the U.S. mental health patient population (the Pearson's χ^2 test p-value>0.14 for the two variables). We also report the sexual orientation information for the 155 Hope users.

We explore the impact of the potential sample selection on our results in three steps. First, we re-specify our SRS equation as a non-linear model since our selection model does not include an exclusion restriction and the Heckman model can be estimated without an exclusion variable for non-linear models (Ichino et al. 2008). We achieve this by: (i) dichotomizing the continuous *AppUsage* and *SRS* variables into binary variables $AppUsage'_{ij}$ (equals 1 if user i has used the app at least once within 15 days prior to the

[1] A similar approach is heavily used in the literature to address non-response bias in survey based research by assuming that late respondents in a survey can be seen as proxies for non-respondents (Etter and Perneger 1997, Siemiatycki and Campbell 1984).

[2] We are not able to compare the 155 Hope users to the mental health population with respect to sexual orientation statistic since this statistic is not available for the mental health population.

Table E1: A Comparison of the Statistics for Socio-demographic Characteristics

	Status	Definition	155 Hope users	U.S. Mental Health Patients %[i]	Pearson's χ^2 test
Gender	Better-served	Female	67.74%	63.65%	*p-value = 0.278*
	Underserved	Male, other	32.26%	36.45%	
Race-Ethnicity	Better-served	White	72.90%	67.39%	*p-value = 0.143*
	Underserved	Asian, African, Hispanic/Latino, other	27.10%	32.61%	

Mental Condition Survey completed at time j, 0 otherwise) and SRS'_{ij} (equals 1 if SRS_{ij} is greater than the median SRS (i.e., high SRS) in our entire sample and 0 otherwise (i.e., low SRS)); and (ii) and estimating Equations 1 and 2 as probit models with $AppUsage'_{ij}$ and SRS'_{ij} being the outcome variables. The first two columns in Table E2 demonstrate the estimation results for Models 3 and 4 with the probit specification. The estimated coefficients and their significance levels in models with probit specification are qualitatively similar to those in Models 3 and 4 in Table 3 of the main paper with one noticeable difference – the coefficient of the interaction term $AppUsage' \times Underserved$ $(\beta_3 = 0.28, p = 0.025)$ is positive and statistically significant. Consistent with H3, this indicates that users from the underserved population benefit more from having any app usage than users from the better-served population. Considering that our hypotheses continue to be supported by the non-linear probit specification, we use it as a benchmark to assess the potential impact of sample selection.

Second, we estimate the Heckman probit selection model using data from all 1,843 users and present the results in the last two columns in Table E2. In the first stage of the Heckman model, using all the covariates in our main model, we estimate whether a user is active (i.e., belongs to 1,688 users) or not (i.e., belongs to 155 users) and obtain the inverse Mills ratios. Then, in the second stage, we estimate Equations 1 and 2 by including the inverse Mills ratios obtained from the first stage. Comparing the last two columns to the first two columns in Table E2, there is a negligible difference in the coefficient estimates.

Third, because the selection model does not include an exclusion restriction, following Ichino et al. (2008), we conduct a sensitivity analysis to understand whether and how including a potential exclusion restriction in the selection model could have changed our results. To do so, we estimate the Heckman probit selection model by fixing the correlations between the $AppUsage'$/SRS' models and the selection model at various levels between -0.9 to 0.9. The premise of this approach is that including a potential exclusion restriction in the selection model might provide different correlations and thus can change the estimated coefficients. Hence, examining the sensitivity of estimated coefficients with respect to correlations fixed at different values within a plausible range (i.e., -0.9 to 0.9) enables us to understand the maximum effect that

Table E2: Heckman Probit Selection Model Estimation Results

	Ordinary Probit Model		Heckman Selection Model	
	MODEL 3	MODEL 4	MODEL 3	MODEL 4
APP USAGE EQUATION				
Underserved	-0.07	-0.07	0.06	0.06
	(0.06)	(0.06)	(0.06)	(0.06)
Age	-0.01**	-0.01**	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
InitialSRS	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Tenure	-0.00	-0.00	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
AppRemindFrequency	0.10***	0.10***	0.10***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)
SELF-REFLECTION SCORE (SRS) EQUATION				
AppUsage'	0.38***	0.20†	0.38***	0.20†
	(0.06)	(0.10)	(0.06)	(0.10)
Underserved	0.17*	0.01	0.17*	0.01
	(0.08)	(0.08)	(0.08)	(0.08)
AppUsage' × Underserved		0.28*		0.28*
		(0.13)		(0.13)
Age	0.01	0.01	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)
InitialSRS	0.06***	0.06***	0.06***	0.06***
	(0.00)	(0.00)	(0.00)	(0.00)
Tenure	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
AppRemindFrequency	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
				-0.69
SELECTION MODEL	NO	NO	YES	(0.56)
AIC	13827.57	13815.33	13823.2	13817.11
BIC	14214.8	14181.82	14348.74	14342.64
Likelihood Ratio Test	-	-6.24	-	-6.09
ρ			-1***	-1***
ϱ			-.71	-.69
Number of users	1,688	1,688	1,688	1,688
Number of observations	7,442	7,442	7,442	7,442

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) User random effects and year-month fixed effects are included in all models. (iv) ρ and ϱ are the correlations between the App Usage and the Selection models, and between the Self-reflection Score and the Selection models, respectively. (v) ρ for the App Usage equations are truncated at -1.

any theoretical exclusion restriction might have on our results if included in the selection model. Table E3 reports the results from this sensitivity analysis. We find that the coefficients of interest for H1 in the *AppUsage* equation and H2 in the *SRS* equation remain the same across various correlation values and are consistent with those estimated in Table E2. Overall, the Heckman selection model estimation and the sensitivity analysis suggest that the potential sample selection bias has a negligible impact on our qualitative insights.

Table E3: Sensitivity Analysis for the Heckman Probit Selection Model

App Usage Equation ρ	-0.9	-0.7	-0.5	-0.3	-0.1	0.1	0.3	0.5	0.7	0.9
Underserved	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Self-Reflection Score (SRS) Equation ϱ	-0.9	-0.7	-0.5	-0.3	-0.1	0.1	0.3	0.5	0.7	0.9
AppUsage'	0.20†	0.20†	0.20†	0.20†	0.20†	0.20†	0.20†	0.20†	0.20†	0.20†
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
AppUsage' × *Underserved*	0.28*	0.28*	0.28*	0.28*	0.28*	0.28*	0.28*	0.29*	0.29*	0.29*
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) For brevity, only the coefficients of interests are displayed.

E.2 Operationalization of *AppUsage* with Different Time Thresholds

In the data, 98.8% of the app usages (used to operationalize *AppUsage*) before taking a Mental Condition Survey occur within 15 days prior to taking that survey. It might be possible that the remaining 1.2% not considered when operationalizing *AppUsage* can change our results. To explore this possibility, we construct two alternative *AppUsage* variables by considering two different thresholds; one month and two months corresponding to 99.3% and 99.6%, respectively, of the app usages occurred prior to taking a Mental Condition Survey. Table E4 present the estimation results for Model 4 with the two alternative operationalizations of *AppUsage*. We observe that our insights from the main analysis remain the same even when we consider longer time spans for app usage before completing the Mental Condition Survey, implying that missing out on significant periods of activity has a negligible impact on our results.

Table E4: Estimation Results for Operationalization of *AppUsage* with Different Time Thresholds

Alternative Operationalization of *AppUsage*:	**Number of Observations**	**APP USAGE EQUATION**	**SELF-REFLECTION SCORE (SRS) EQUATION**		
		Underserved	*AppUsage*	*Underserved*	*AppUsage* × *Underserved*
1 Month	7,442	0.07	1.94**	0.31	1.17
		(0.05)	(0.70)	(0.59)	(0.82)
2 Months	7,442	0.06	1.92**	0.19	1.24
		(0.05)	(0.66)	(0.59)	(0.79)

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) For brevity, only the coefficients of variables of interests estimated in Model 4 are included in the table.

E.3 Simultaneity between *AppUsage* and *SRS*

As discussed in Section 4.1, our main model assumes that *AppUsage* is expected to influence *SRS*, while *SRS* is not anticipated to affect *AppUsage*. However, it is conceivable that users with higher *SRS* (i.e., better mental conditions) may be more willing to use MHMAs as they find MHMAs helpful, which is likely to imply simultaneity between *AppUsage* and *SRS,* and thus to violate our model assumption. To address this, we relax the model assumption by including the *Lagged SRS* variable in the *AppUsage* equation (Equation 1 of our main model) to capture the potential impact of previous *SRS* on the current *AppUsage*. Note that by adding the *Lagged SRS* variable, we lose observations (5,754 vs. the original 7,442 data points) because the values for the *Lagged SRS* variable cannot be obtained for the first time a user takes the survey, and thus all the observations related to the first-time usage are dropped. Table E5 presents the results. The estimated coefficient of *Lagged SRS* is positive and significant ($\alpha = 0.003, p < 0.001$), confirming that the previous *SRS* does have a significant and positive association with the current app usage. Nevertheless, our hypotheses still hold after considering the simultaneity between *AppUsage* and *SRS*, suggesting a negligible impact of the simultaneity between *AppUsage* and *SRS* our results.

Table E5: Estimating Results for Simultaneity between *AppUsage* and *SRS*

	MODEL 4
APP USAGE EQUATION	
Underserved	0.11*
	(0.05)
Lagged SRS	0.003***
	(0.00)
Age	-0.00
	(0.00)
InitialSRS	-0.00**
	(0.00)
Tenure	-0.00***
	(0.00)
AppRemindFrequency	0.03***
	(0.01)
SELF-REFLECTION SCORE (SRS) EQUATION	
AppUsage	1.83*
	(0.86)
Underserved	1.60
	(1.57)
AppUsage × Underserved	0.02
	(1.02)
Age	0.02
	(0.07)
InitialSRS	0.68***
	(0.03)
Tenure	0.01†
	(0.01)
AppRemindFrequency	0.08
	(0.21)
Number of users	942
Number of observations	5,754

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

E.4 Diminishing Benefits of App Usage

One can argue that as users become more familiar with the app and learn more from other users' experiences, the benefit of the app may diminish over time (akin to the learning effect). Subsequently, a newly registered user with a certain usage frequency may benefit from the app more than a relatively old user with the same usage frequency. If the aforementioned diminishing marginal benefits vary between users from the underserved and better-served populations, our result regarding the equity in benefit might be attributed to the learning effect, rather than the app itself. Subsequently, the equity in benefit may no longer be present after controlling for the learning effect. To explore this alternative explanation, we re-operationalize $AppUsage_{ij}$ as a cumulative variable (i.e., the natural logarithm of the number of times the app functions are used since app registration by user i prior to the Mental Condition Survey completed at time j) and estimate Models 3 and 4 with a quadratic term for this cumulative variable in the SRS equation. Table E6 presents the results. The estimated coefficient of the qudratic term $AppUsage^2$ in the SRS equation ($\beta = -0.65, p < 0.01$) is negative and statistically significant in Model 3 and insignificant ($\beta = -0.39, p > 0.1$) in Model 4. These results suggest that the significant quadratic term coefficient in Model 3 is indeed driven by the difference in benefits from app usage between users from the underserved and better-served populations, suggesting that diminishing benefit is unlikely in our sample. Regardless, our hypotheses continue to hold with these specifications, ruling out diminishing benefit as an alternative explanation.

Table E6: Estimation Results for Diminishing Benefits of *App Usage*

Model	APP USAGE EQUATION	SELF-REFLECTION SCORE (SRS) EQUATION				
	Underserved	*AppUsage*	*Underserved*	*AppUsage × Underserved*	*AppUsage²*	*AppUsage² × Underserved*
Model 3	0.06	4.97***	1.38†		-0.65**	
	(0.05)	(0.72)	(0.71)		(0.21)	
Model 4	0.06	3.14**	-0.60	2.85†	-0.39	-0.41
	(0.05)	(1.14)	(0.44)	(1.46)	(0.34)	(0.42)

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) For brevity, only the coefficients of variables of interests included in the table.

E.5 Specification with Fixed Effects

As discussed in Section 4.1, our main model includes user random effects instead of fixed effects to enable the testing of H1. To examine whether a fixed effect specification would change our findings, we re-specify Equation 2 as a fixed effect model. Note that in a fixed effect model estimation, all time-invariant variables are dropped. Therefore, we estimate the fixed effect model for Equation 2 (reported in the first row of Table E7) and compare the estimated coefficients to those in the random effect model in Table 3 of the main paper.

We find that while the coefficients of interest (i.e., *AppUsage* and *AppUsage* × *Underserved*) are slightly different in magnitude between the two models, the qualitative insights remain the same.

Table E7: Estimation Results for Various Robustness Checks

Robustness Check	Number of Observations	APP USAGE EQUATION	SELF-REFLECTION SCORE (SRS) EQUATION		
		Underserved	*AppUsage*	*Underserved*	*AppUsage × Underserved*
Fixed Effect Model	7,442		2.32*** (0.69)		0.33 (0.83)
Negative Binomial Model	7,442	0.11 (0.09)	0.14† (0.08)	1.66* (0.68)	-0.04 (0.10)
Alternative Operationalization of *AppUsage:* Excluding Responding to Other Peers	7,442	0.06† (0.03)	2.18* (0.89)	0.25 (0.60)	1.58 (1.04)
Alternative Operationalization of *Underserved:* Excluding gender	7,442	0.09* (0.04)	2.15*** (0.55)	-0.40 (0.60)	0.88 (0.77)
Alternative Operationalization of *SRS*	7,442	0.06 (0.04)	1.74* (0.73)	0.58 (0.60)	1.01 (0.86)
Alternative Panel Structure: Fixed Time Window (1 month)	2,748	0.09† (0.04)	1.19* (0.53)	0.95† (0.57)	1.11† (0.65)

Notes: (i) Robust standard errors clustered at the user level in parentheses. (ii) *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. (iii) For brevity, only the coefficients of variables of interests estimated in Model 4 are included in the table.

E.6 Specification with Negative Binomial Model

In our data, the count variable app function usage prior to completing the Mental Condition Survey has a right-skewed distribution, which is the reason why we operationalized *AppUsage* as the natural logarithm of the number of times the app functions. For such count dependent variables, one can argue that fitting a negative binomial regression would be more appropriate. Therefore, in this robustness check, we operationalize *AppUsage* as the number of times the app functions are used within 15 days prior to completing the Mental Condition Survey and re-specify Equation 1 as a negative binomial model. As is evident from the results presented in the second row of Table E7, our qualitative insights from the main analysis still hold with this alternative specification.

E.7 Operationalization of *AppUsage* with Different Peer-support Activities

When constructing *AppUsage*, following the literature (Yan and Tan 2014) we consider peer-support usage activities related to both receiving support (i.e., writing original posts) and giving support (i.e., responding to other users' posts). Despite this classification in the literature, one may argue that responding to other users' posts may not necessarily benefit the user who responds, and thus should not be considered a part of the peer-support app usage activities. To examine whether our results change under this alternative classification of app usage activities, we exclude all peer-support activities related to responding to other peers, reconstruct *AppUsage* using only activities related to self-reflection function usage and writing a post, and estimate Model 4 with this alternative *AppUsage* variable. As evident in the third row of Table E7, we find that our insights regarding the equity in usage and benefit remain the same with this alternative *AppUsage* variable.

E.8 Operationalization of *Underserved* without Gender

In this robustness test, following the traditional definition of the underserved population in other service domains that consider only supply-side challenges, we operationalize *Underserved* using only socio-demographic identities of race-ethnicity and sexual orientation, ignoring the gender identity. With this alternative classification, we refer to individuals as belonging to the *underserved population* if they identify with one or more of the following socio-demographic identities: (i) Sexual orientation (i.e., homosexual, and other non-heterosexual identities); and (ii) Race-ethnicity (i.e., African, Asian, Hispanic/Latino and other non-White identities). Conversely, the *better-served population* refers to individuals who identify as heterosexual and White. The results for Model 4 with this alternative *Underserved* variable are presented in the fourth row of Table E7. As is evident from the results, excluding gender when operationalizing *Underserved* does not impact our insights. In fact, we find that users from the underserved population would use MHMA more than their better-served counterparts, a desirable outcome for the underserved population. These results imply that our results are not driven by considering gender in classifying underserved users.

E.9 Alternative Operationalization of *SRS*

In our main analysis, we operationalize *SRS* using the estimated standardized loadings in the confirmatory factor analysis. As an alternative, following Siemsen et al. (2009), we re-operationalize the dependent variable *SRS* as the scale average of the seven survey items. As is evident from the results presented in the fifth row of Table E7, our qualitative insights are also robust to this alternative operationalization.

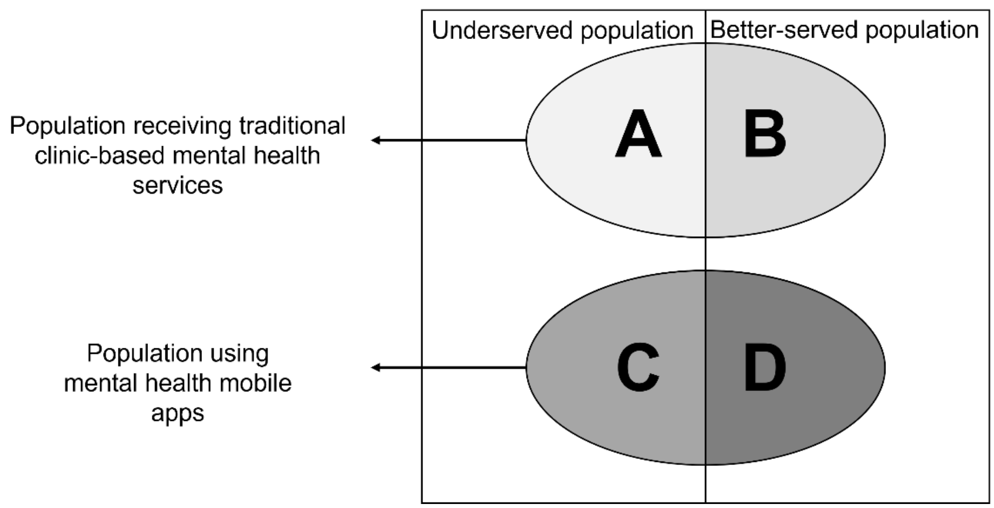
E.10 Alternative Panel Structure

In our panel data, each observation corresponds to a completed Mental Condition Survey. Due to variations in the number of survey completions, it is possible that individuals who use the survey more frequently are overrepresented in our panel compared to those who use it less frequently. To examine the potential impact of this overrepresentation, we construct an alternative panel where the panel time intervals are set to one month (i.e., representing the 98^{th} percentile of the distribution of the time between two consecutive SRSs). Since the average time between two consecutive Mental Condition Survey completions is 7.18 days, it is likely that individuals who use the survey more frequently may complete more than one survey within a month period. Therefore, in this panel, we operationalize *SRS* as the average score of (multiple) surveys completed by user i in time t and *AppUsage* as the total app usage of user i in time $t\text{-}1$. With this data, for users who complete multiple surveys within one month, obtaining the average SRS from those survey scores and creating only one observation for that month (as opposed to having multiple observations as we do in our main analysis), we ensure that individuals using the survey more frequently are not over-represented. We estimate our main model using this alternative panel and present the results in the last row of Table E7. We find that our insights from the original panel regarding equity in usage and benefit of MHMAs remain the same with this alternative panel. In addition, users from the traditionally underserved population use MHMA marginally more and benefit marginally more from MHMA usage than their better-served counterparts, supporting H1 and H3. Overall, this robustness test indicates that the over-representation of individuals who use the survey more frequently has a negligible effect on our results.

Appendix F: MHMA Users vs. the Mental Health Patient Population

In this appendix, we illustrate in detail the two likely relationships between the MHMA users and the traditional mental health patient population, and discuss how these relationships can map with our empirical findings in the main analysis. Consider the following figure:

Figure F1: Illustration of MHMA Users vs. Traditional Mental Health Patients



The figure demonstrates the underserved population (i.e., the left rectangle) and the better-served population (i.e., the right rectangle), who are classified based on exogenously determined socio-demographic characteristics (i.e., gender, sexual orientation, and race-ethnicity). Within these populations, the two circles represent the population seeking professional support delivered through the traditional clinic-based mental health services (i.e., the circle on top) and the population seeking help through MHMAs (the circle at the bottom). The well-documented inequity in the traditional clinic-based mental health services between the underserved and better-served populations in the extant literature implies that $A \neq B$ with respect to usage and benefit. The results of our study imply that $C = D$ with respect to usage and benefit in MHMAs. Within this illustration, we can think of two scenarios for leveraging MHMAs in the traditional clinic-based mental healthcare delivery.

In the first scenario, the underserved and/or better-served populations in the MHMA vs. professional settings (i.e., C vs. A and/or D vs. B) can inherently be different, i.e., the population using MHMAs is different from the population seeking professional support. This implies that MHMA users can be those that do not seek traditional clinic-based professional help, yet seek peer-support and self-support help on their own in the app. In other words, the mobile app may expand the mental health services (through self-support and peer-support) to a new population segment who otherwise would not seek traditional clinic-based professional support, creating a market expansion effect. It is plausible that, if the underserved and better-served populations in this new segment (i.e., C and D) had used professional services, they would not have experienced any inequity, implying an already equity between them before using the app. It is also plausible that the underserved new segment (i.e., C) would have experienced inequity in the professional setting if they had used the professional services, and thus using the mobile app advances equity for them. Regardless of these two possible explanations, this scenario may imply that MHMAs may expand the mental health services (through peer-support and self-support) to a new segment that either: (i) already has equity between the underserved and better-served populations or (ii) achieves equity between the underserved and better-served populations with the help of the mobile app.

In the second scenario, in contrast to the first one, it is possible that the underserved and better-served populations in the MHMA vs. professional settings (i.e., C vs. A and D vs. B) can inherently be the same population, i.e., the population using MHMAs is the same as the population seeking professional support. This implies that MHMA users can be those that also seek traditional clinic-based professional support. In other words, the self-support and peer-support functions of MHMAs can complement the delivery of mental health services for those seeking traditional clinic-based professional support, creating a multi-channel mental health help-seeking effect. In this case, considering the well-documented inequity in the mental healthcare literature (i.e., $A \neq B$), our results (i.e., $C = D$) may imply that MHMAs advance the equity between the underserved and better-served populations through: (i) providing an alternative channel for seeking help for those who experience inequity in the professional settings and (ii) potentially facilitating some behavioral changes among those populations.

References

Bateman PJ, Gray PH, Butler BS (2011) The impact of community commitment on participation in online communities. *Information Systems Research* 22(4):841–854.

Dorwart L (2023) Best mental health apps. *Very well mind*. Retrieved (June 20, 2023), https://www.verywellmind.com/best-mental-health-apps-4692902.

Etter JF, Perneger T V. (1997) Analysis of non-response bias in a mailed health survey. *Journal of Clinical Epidemiology* 50(10):1123–1128.

Heckman JJ (2013) Sample selection bias as a specification error. *Applied Econometrics* 31(3):129–137.

Ichino A, Mealli F, Nannicini T (2008) From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics* 23(3):305–327.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Information Systems Research* 19(4):456–474.

Siemiatycki J, Campbell S (1984) Nonresponse bias and early versus all responders in mail and telephone surveys. *American Journal of Epidemiology* 120(2):291–301.

Siemsen E, Roth A V., Balasubramanian S, Anand G (2009) The influence of psychological safety and confidence in knowledge on employee knowledge sharing. *Manufacturing and Service Operations Management* 11(3):429–447.

Van de Ven WPMM, Van Praag BMS (1981) The demand for deductibles in private health insurance. A probit model with sample selection. *Journal of Econometrics* 17(2):229–252.

Yan L, Tan Y (2014) Feeling blue? Go online: An empirical study of social support among patients. *Information Systems Research* 25(4):690–709.