Stigma and the Distribution of Information: A Mental Health Care Experiment in Refugee Networks

Emma C. Smith*

April 2026

Abstract

Many services available to help the most disadvantaged are associated with stigma, creating a consequential information friction: people may withhold telling others about services they do not want to be associated with. In a field experiment with over 3500 Syrian refugees, I show that this friction is concentrated among those best positioned to spread information. Community members share mental health information with only a quarter of intended recipients despite agreeing to participate, knowing many are in need, and believing services are effective. Encouraging senders to use social cover, by disclosing they are paid to share, increases sharing rates by over a third. But image concerns are far higher among prior mental health care users. These prior users more than double their sharing rates under social cover, and without it, they share at similarly low rates to non-users despite holding higher beliefs about service efficacy. Because prior users are also more connected to high-need peers through homophily, their sending reaches recipients 11 percentage points more likely to be experiencing depression or anxiety than non-users' sending. Inducing initial engagement also activates informal support, leading to a 0.37 standard deviation gain in social connectedness for recipients. The findings help reconcile a puzzle in the take-up of stigmatized services, where stigma constraints appear modest relative to information frictions at the individual level. This paper shows that stigma can operate at the network level, suppressing information flows from the most knowledgeable people to those most in need.

JEL Classification: D83, I12, C93, D91
Keywords: Stigma and Social Image, Learning, Targeting, Mental Health, Refugees

*Georgetown University. Email: emma.smith@georgetown.edu. I am grateful to Emily Breza, Rema Hanna, Edward Miguel and Gautam Rao for guidance and support. For helpful comments I thank Layane Alhorr, Jishnu Das, Eliana La Ferrara, Edward Glaeser, Ronak Jain, Asim Khwaja, Guillermo Palacios-Diaz, I. Bailey Palmer, Benjamin Roth, Simone Schaner, Abdulrazzak Tamim, Lise Vesterlund, and seminar, workshop and conference participants at Georgetown University, Harvard University, the University of Pittsburgh, NEUDC, UNHCR-WB Joint Data Center Research Conference on Forced Displacement, Y-Rise, and CEPR. The project would not have been possible without support from Rachit Shah, Eva Kaplan and the International Rescue Committee, Laura Rodriguez Takeuchi and Matthew Wai-Poi, Noor Daoud, Majd Masannat, and other members of Mindset. This study received generous funding and support from the Innovations for Poverty Action Peace and Recovery Fund, Weiss Fund, SurveyCTO, Weatherhead Institute, Institute for Humane Studies, and Institute for Quantitative Social Sciences. The study received IRB approval from Harvard University protocol number IRB20-1023. The experiments were pre-registered under RCT IDs AEARCTR-0010702 and AEARCTR-0011928.

1 Introduction

Often, services meant to help vulnerable people are associated with stigma. A puzzle in the literature is why well-documented stigma [1] appears to play only a minor role in individual take-up decisions of stigmatized services, relative to information constraints (Anders and Rafkin, 2022; Bhargava and Manoli, 2015; Currie, 2004; Ko and Moffitt, 2024). This paper shows that stigma can substantially shape information constraints, with important implications for who learns about services.

Whether thinking of mental health care, "food stamps" and other social services, or HIV testing and treatment, people who use these services may not want others to know. This can create a consequential information friction: when people do not want others to know they used a service, they may withhold information about it. This can suppress information specifically from the people most informed about a service, who may also be most connected to people in need.

Reaching potential service users, and particularly those most vulnerable, with the right information, is a major challenge across high and low income settings. Even in high income countries where electronic systems and comprehensive online information might theoretically resolve information frictions, large knowledge and access gaps persist and are often concentrated among the most disadvantaged (Bettinger et al., 2012; Finkelstein and Notowidigdo, 2019). The significant impacts of community outreach and information on take-up of social and health services (Aizer, 2003; Dupas and Miguel, 2017; Finkelstein and Notowidigdo, 2019; Kimbrough et al., 2009) underscore the prevailing norm of imperfect information and high transaction costs (Castell et al., 2025). In lower capacity settings like developing countries as well as migrant communities, it is even more common for individuals to rely on peers and local networks as a trusted source of information (Banerjee et al., 2018; Beaman et al., 2021; Bertrand et al., 2000). Such settings increase the appeal of community-based targeting and peer referrals (Alatas et al., 2016; Berg et al., 2017; J. Goldberg et al., 2023), but these informal systems may amplify the consequences of information withholding (Cefalà et al., 2024). With social protection services growing in lower-income countries, identifying barriers to awareness and take-up is of increasing importance (Banerjee et al., 2024).

I study influential peers' willingness to circulate information about one of the main phone counseling helplines funded by the Jordanian government, through the Jordan River Foundation, focusing on the high-need population of Syrian refugees living in Jordan. Peer-to-peer marketing

[1] For examples of documented stigma see: mental health (Bharadwaj et al., 2015; Ridley, 2022), safety net programs (Gustafsson, 2002; Stuber and Schlesinger, 2006), HIV (Derksen et al., 2022; Musheke et al., 2013; Yu, 2023.

and referrals are recognized as uniquely effective for matching consumers to services they value (Aral and Walker, 2011; Luca, 2016). Just as homophily in tastes and needs can increase the benefits of peers' product recommendations (BenYishay and Mobarak, 2019), correlated vulnerability as well as friends' knowledge of one another's needs could improve the match to services (J. Goldberg et al., 2023). Besides learning about the existence of services, conversations between between potential users, who suffer from similar problems, may have value in themselves. Hearing from a friend that they too struggled with mental health may be directly helpful, regardless of whether the person takes up formal mental health care. Yet this virtuous exchange depends crucially on users being willing to share initial information that could signal their user status.

Despite being the leading cause of disability, mental health remains under-treated and stigmatized across the globe (Bloom et al., 2012, WHO et al., 2004 and Pescosolido et al., 2013). Refugees experience particularly high rates of mental health problems – estimates suggest one in three refugees has depression, anxiety or post-traumatic stress disorder (PTSD) (Blackmore et al., 2020). Rates are even higher among Syrian refugees in Jordan, where roughly half the adult population likely has depression or anxiety. Yet, fewer than 10% of households are seeking regular care.

I identify over 800 "central" or well-regarded individuals through peer nominations by a representative sample from the confidential UN refugee registry in Jordan. Through network elicitation surveys and mental health assessments extending to each individual's friend group I uncover not only homophily in mental health need, but also significant knowledge that the core "senders" have about relative mental health need within their social network. This contrasts prior research that has found limited scope for community based targeting of sensitive vulnerabilities (Agüero et al., 2020).

I then use an information sharing experiment to both identify and overcome image concerns that prevent knowledge of services from circulating. As noted before, senders may worry about how they will be perceived if they share the information. I show in a simple signaling model that sharing the information could indicate that the sender is associated with, or used, the service themselves, and similarly could indicate that the recipient is in-need (Chandrasekhar et al., 2018). To identify these image concerns I randomly vary the senders' introduction used when sharing the information over WhatsApp. In the "disclosure" treatment, the study encourages some senders to reveal they are being paid to share the campaign. Doing so provides an excuse for why the sender is sharing. In fact all senders are paid, so disclosing this in some friend groups and not others creates variation only in the "social cover" that senders have.[2] A second source of variation comes from encouraging senders in their introduction either to say that they are trying to share with all of their friends, or instead say they are trying to share specifically with friends who they think will benefit from the information. Privately, the study asks senders to share information with all of their friends (listed by name). But to the recipients, the framing creates variation in whether they think they were targeted, which, as shown in the model below, could dampen the sender's willingness to share.

The experiment's average first stage results reaffirm the status quo of low information sharing: Despite all senders agreeing to share when enrolled, over half of recipients likely having depression or anxiety, and senders often having knowledge of their friends' need, senders share with only about a quarter of recipients.

However, encouraging senders to disclose that they are paid led to significantly more sending. Consistent with the hypothesis that disclosing payment creates a stigma-alleviating excuse, senders are more than 7 percentage points (34%) more likely to share using an introduction that discloses they are paid, relative to the introduction saying they "want to share" but are still compensated privately (p-value$= 0.013$). On the other hand I find little evidence that sending behavior is different on average when the introduction suggests the recipient was targeted. Because the treatment arms differ in no other way than disclosing that the sender is paid, the difference in sending rates can be attributed to the signal-dampening effect of disclosure, and from this I infer that sharing mental health information carries a costly signal.

The total consequences of these image concerns on learning depends on how image-sensitivity and need are distributed in the population. I show that image concerns are in fact highly heterogeneous, and prior mental health care users are the most sensitive by far (Chernozhukov et al., 2018). These individuals respond to the disclosure treatment by more than doubling their sharing rates relative to the "non-disclosed" framing (p-value $= 0.001$). Without the disclosure, these senders share at similarly low rates to non-users, despite believing that the benefits from mental health care are larger, consistent with perceiving higher image costs.

Not only are prior users more image-sensitive, but they are also more connected to vulnerable people in need of mental health services. As a result of this homophily, prior users reach recipients who are 11 percentage points more likely to be experiencing anxiety or depression than those reached by non-users (p-value < 0.05). This stands in notable contrast to many information interventions on sensitive services that have found compliers to be systematically less in need

[2] This approach was developed through pilot observation in the local context, contemporaneously to multiple studies in economics on the signal dampening effects of "justifications" (Bursztyn et al., 2023; Raisaro, 2023; Walsha, 2024).

(Castell et al., 2025; Finkelstein and Notowidigdo, 2019).

A simple signaling framework in the spirit of Chandrasekhar et al., 2018 unifies these findings and explains why the people best positioned to spread information are also the most deterred from doing so. First, sharing information about a stigmatized service carries a social image cost, and that cost is highest for people for whom observers have moderate priors about user status. The logic of Bayesian updating implies that a signal is most informative, and therefore most costly, when uncertainty is highest. This captures the intuition that both a clearly unhoused person and a wealthy philanthropist will worry little about being seen leaving a food pantry, while a lower middle income person may worry a lot about what other community members would infer from seeing them.

Second, because prior users face higher image costs yet are more connected to high-need peers through homophily, image concerns suppress information flows in precisely the direction that most harms targeting.

The intervention also impacted message recipients. A novel aspect of the intervention studied here is that it has the potential to operate not only through delivery of information, but also by triggering informal support between friends. There is remarkably little evidence to-date on the effects of kick-starting sensitive conversations between close ties (Jain and Khandelwal, 2024). Externally facilitating social interaction often improves mental health and social connectedness (Hansen et al., 2024), yet little is known about how to generate these exchanges within existing social groups, where people may hesitate to engage due to heightened relationship stakes (Small, 2017; Webb, 2024).

Inducing engagement generated further conversations and activated informal support for recipients. Treated recipients experienced a 0.37 standard deviation increase in an index of social connectedness, driven by a 0.45 standard deviation increase in the number of times the recipient spent time helping or being helped by a network member, both significant at the 95 percent confidence level. These impacts on social interaction are strikingly comparable to meta-analysis estimates of the impacts across a range of more intensive social interventions (Zagic et al., 2022), perhaps because this intervention engaged existing friends. Treated recipients also held face-to-face or phone conversations about mental health 16 percentage points more (excluding the campaign messages themselves), more than doubling the rate relative to control (p-value < 0.05). This deeper engagement was driven primarily by the "disclosed, targeted" framing, suggesting that recipients responded positively to the sender noticing their need.

There is noisy evidence that recipients take-up mental health treatment, but not the advertised helpline, after six months. In the most-shared treatment arm (disclosed, non-targeted) there is a marginally significant 7.2 percentage point (31%) increase in having ever used mental health treatment, which can include in-person counseling and taking medication (p-value= 0.082, ITT estimate). The pooled IV estimate is statistically insignificant but represents a 51% increase in treatment use (p-value= 0.222). For the phone counseling service specifically, the results rule out even low take-up (including any possibility of spillovers) since less than 2% of both the treatment and control groups used the service. Even lowering the fixed cost of take-up by offering to have the helpline contact the person directly did not show any treatment effects for those exposed to the campaign. However over 60% of participants agreed to have the helpline contact them, underscoring information and transaction costs as dominant constraints rather than stigma.

Together these results show that stigma's consequences extend beyond individual take-up decisions: it shapes who learns about services in the first place. Even when information is present in a community and mutual knowledge of need is high, image concerns cause it to be suppressed most among those with firsthand experience, and rationed away from those most in need. But alleviating image concerns reverses this, leading sharing rates to increase by a third overall and more than double among prior users, whose newly reached recipients are 11 percentage points more likely to be experiencing depression or anxiety than those reached by non-users. The findings point not only to the value of techniques that mitigate image concerns in the moment, but to the broader importance of destigmatization. If knowledge gaps within vulnerable communities are sustained largely by stigma, reducing stigma may deliver far greater informational benefits than previously recognized.

This paper advances four literatures. First, this work advances the literature on take-up of stigmatized services. Across high and low income contexts the literature has studied stigma and information as largely independent constraints, and while information constraints are well documented to bind, identifying effects of stigma have been more elusive (Breza and Kaur, 2026; Dupas and Miguel, 2017; Ko and Moffitt, 2024). Specifically Bettinger et al., 2012 and Finkelstein and Notowidigdo, 2019 show that information is a barrier to take-up of public benefits in the United States. Bertrand et al., 2000 find that the proportion of other benefits users in close proximity influences benefits use, though information and image concerns cannot be fully disentangled. Anders and Rafkin, 2022 document stigma concerns regarding public benefits, but show these concerns fail to explain which people take up, and point instead to information frictions. Osman and Speer, 2024 estimates a small effect of stigma on take-up on job search assistance. In the realm of health, Dupas and Miguel, 2017 review the extensive literature on HIV testing and risky sexual behavior and find mixed but often non-zero impacts of information.

Yang et al., 2023 and Yu, 2023 document both information and social stigma concerns as barriers to HIV testing. In mental health service take-up specifically there is strong descriptive evidence of stigma as a reported barrier, but limited causal evidence (Clement et al., 2015; Lacey et al., 2025; Roth et al., 2024a). This paper helps to reconcile the tension in the literature: stigma can operate at the network learning level, so studies looking at individual take-up choices may underestimate the effect of stigma while large gaps in awareness go unexplained. The paper further shows that, because experienced users face the highest image costs and are most connected to high-need peers, image concerns and need are distributed in precisely the direction that most suppresses learning and community-based targeting (Alatas et al., 2016; Berg et al., 2017; J. Goldberg et al., 2023). Reducing image concerns raises information sharing not just in general, but specifically from the most experienced users to the people most in need.

Second, this paper advances the literature on informal support as a complement to formal care for mental health and social connection. The literature shows that demand for counseling remains stubbornly low (Cronin et al., 2025). Though correcting misperceptions about stigma and sharing information about treatment efficacy can lead to modest and sometimes short-term movement in demand for treatment (Batmanov et al., 2026; Lacey et al., 2025; Roth et al., 2024b), these interventions have not yet closed the gap in care seeking, and sometimes also shift individuals to prefer self-care or non-clinical care (Acampora et al., 2022; Roth et al., 2024a). This highlights the potential value of informal peer support as a complement to formal care. In line with this, efforts to tackle the epidemic of loneliness and social isolation (Banerjee et al., 2023; Office of the Surgeon General (OSG), 2023) often intervene on social interaction directly (Hansen et al., 2024). Those studies show moderate effects, but typically through resource-intensive interventions that involve multiple sessions of in-person activities (Zagic et al., 2022). This paper's findings reveal that triggering engagement within existing social networks can deliver greater social connectedness at low cost, complementing efforts to expand formal services in settings where supply remains scarce.

Third, this paper advances the literature on social learning and reputational concerns in networks. A growing body of work documents that image and reputational concerns shape information flows in social networks (Breza and Chandrasekhar, 2019; Karing, 2018), and in information seeking specifically (Banerjee et al., 2018; Chandrasekhar et al., 2018). Closely related work identifies social cover as a mechanism for alleviating image concerns across a range of settings (Bursztyn et al., 2023; Raisaro, 2023; Walsha, 2024). This paper makes two contributions to the literature on social image and learning. First this paper shows that social cover shapes not just whether people act, but who learns, connecting image concerns to consequences for information distribution and whether the most vulnerable are reached. Second, and most distinctively, this paper shows that firsthand experience, which is typically an asset in social learning (Banerjee et al., 2013; BenYishay and Mobarak, 2019; Conley and Udry, 2010), can become a liability when the experience itself is stigmatized. Prior users are better positioned than anyone to spread useful information, yet they are precisely the people most deterred from doing so. This double-edged sword of experience has implications well beyond mental health: from food banks and unemployment assistance to HIV and substance use treatment, users of stigmatized services may systematically withhold the referrals and recommendations that would most benefit their peers.

Fourth and finally, this paper contributes to a small but growing body of research in economics on the lives of forcibly displaced populations. Displaced people face acute mental health burdens and severely limited access to services, yet remain among the most understudied populations in economics (Alan et al., 2021; Baseler et al., 2023; Harker Roa et al., 2023; Hussam et al., 2022; Tamim et al., 2025). High quality causal evidence in this setting is scarce, and the findings here have wide-reaching relevance for the design of outreach and information programs serving displaced populations.

The remainder of the paper is organized as follows: I describe the context in Section 2. Section 3 presents the data and experimental design. Section 4 outlines the conceptual framework. Section 5 describes the empirical strategy, Section 6 presents the results, and Section 7 concludes.

2 Context and Motivation

I conduct the study with a sample of Syrian refugees living in Jordan. Most of the roughly 660,000 Syrian refugees living in Jordan at the time of the study live outside of camps among the host population, and a majority have been in Jordan for over decade, having been predominantly displaced after the Syrian Civil War began in 2011 (UNHCR, 2024). In this setting there is a large mental health burden, with representative surveys suggesting almost half of the adult population likely has depression (Stillman et al., 2022).

Yet I document that use of mental health services remains very low. Prior to the experiments I collect nationally representative data on depressive symptoms and care-seeking among 1516 Syrian refugees in Jordan. Like prior studies, I find that roughly half of adults may likely have depression based on the PHQ-2 scale, but in only a minority of households is someone seeking regular mental health services. These large treatment gaps are mirrored by substantial knowledge gaps as well. Later in the experiments I see that, in the control group, 70% of recipients cannot name a single organization that provides mental health services (even though several humanitarian and local organizations do so in this context), and likewise over 70% have not spoken to anyone outside their household about mental health in the past 6 months.

A first-order question to some readers may be whether it is reasonable to believe that mental health services can help refugees whose external circumstances are so difficult and outside their control. A majority of Syrian refugees outside of camps in Jordan live below the poverty line and are largely banned from accessing formal employment even after being displaced for over a decade (Portection and Operations, 2022, Erik et al., 2021). As such children have limited employment pathways to aspire to, although school enrollment rates did recover to pre-displacement levels (Krafft et al., 2022). Few refugees expected these conditions to change, with only 20% saying in 2023 that the war in Syrian would likely end in the next two years. In addition to the difficult circumstances while being hosted as refugees, many individuals left Syria in distress and may have experienced trauma before or during their displacement.

Despite these extremely difficult circumstances that Syrian refugees face, a variety of mental health interventions have shown positive effects for this group and similar populations. For example in several refugee and post-conflict settings in-person psychosocial support by non-specialist para-professionals has yielded positive mental health impacts (Rahman et al., 2019, Islam et al., 2021, De Graaff et al., 2023), perhaps most notably through the WHO's widely implemented Problem Management+ intervention. Evidence on lighter-touch interventions (closer to the phone counseling studied in this paper) have yielded smaller but still positive impacts on refugee mental health, often at lower cost. For example different self-guided resources succeeded in reducing distress among Syrian and Ukrainian refugees (Burchert et al., 2024, Khedari, 2020, Asanov et al., 2024). Outside of refugee populations, in both high-income contexts (Cuijpers et al., 2010) and low-income countries, in-person psychotherapy typically leads to robust improvements in mental health for people with symptoms of mood disorders like depression and anxiety (Baranov et al., 2020; Barker et al., 2022; Bhat et al., 2022; Patel et al., 2017), though in a notable exception Haushofer et al., 2020 find no effect of psychotherapy for low-income Kenyans. Lower cost interventions such as those administered by mobile phone or online often have relatively lower but positive efficacy, with evidence that human-guided versions are relatively more effective (S. B. Goldberg et al., 2022; Moshe et al., 2021).

It could be that few people use services because of beliefs – they do not believe they need mental health help, or that services are effective. Yet, a majority of control group recipients state that they likely have depression or anxiety, and in fact on average over-report experiencing this relative to validated assessments. Additionally, recipients at baseline report they believe their distress levels would be significantly lower if they used mental health services such as calling a helpline, visiting a specialist, or receiving medication (Figure D.3). These self-reported measures are only suggestive, because they may be biased by social desirability or demand effects, but nonetheless the responses do not point to perceived lack of need as a driving reason for low service take-up. Instead, stigma toward care-seekers emerges as a likely culprit. Over 40% of a representative sample surveyed says that they would not marry someone who once sought professional mental health services. Strikingly, the rates remain high across different segments of the population, such as male or female, above or under the age of 30, married or unmarried (Figure D.4). People are broadly aware that others hold stigmatizing views, as shown by the fact that at baseline half of respondents say they worry their friends would consider them unreliable if they used mental health services.

3 Data and Experimental Design

In order to study the spread of stigmatized information within friend groups, I conduct two rounds of peer referrals, first to construct a sample of potential “senders", and then to collect data on each sender’s’ friend group and construct the intended “recipient” sample. I then implement the main information sharing experiment in which I “seed” information about the mental health phone counseling service via the senders. Within that experiment I study stigma barriers to sharing the information, by randomizing the framing senders are asked to use when sending information to their friends. From the experiment I study senders’ differential willingness to spread information, and recipients’ interest in taking up a free phone counseling helpline after exposure to information from their friend.

3.1 Sample and Recruitment

3.1.1 Sender Nominations

In order to ensure minimal overlap of social networks and representativeness across the refugee population, sampling began from a roster of households randomly selected directly from the UN Refugee Agency (UNHCR)’s registry of all Syrian refugees registered in Jordan. This sample was identified as part of the World Bank’s COVID-19 monitoring surveys. Respondents in the representative survey were asked to help identify people in their community who may be central or well-regarded. To do so respondents were presented with the three nomination types in random order, and were not made aware of the mental health focus on the intended study. The three types were being “well-regarded or well-know", or “community-minded", or “a common source of news", with the final category informed by the literature on identifying individuals with high diffusion centrality (Banerjee et al., 2019). Out of 1516 surveys of the representative sample, 726 respondents agreed to nominate individuals to help and the average number of nominations was 2.5. The enrolled "sender" sample comprised 849 individuals who agreed to participate after being nominated by peers in an otherwise-unrelated representative survey of Syrian refugees across Jordan in late 2021 through early 2022. As a result of sampling beginning from a representative sample of the refugee population, the nominated sample is dispersed across the country within little overlap in peer groups, as discussed when describing the recipient sample below.

3.1.2 Sender Eligibility

The sender sample was surveyed by phone in January and February 2023. The enumerator first asked about demographics and attitudes around mental health, but did not mention an awareness campaign or a mental health focus of the study. The sender next completed a social network elicitation focused on the sender's close social network outside her household, such as people the senders socialize with frequently, borrow from or lend to, go to for advice or give advice to, spend time helping or being helped. The median number of friends named was 3. Next, the senders were asked to share the phone numbers for their friends. Conditional on sharing any phone numbers, senders were informed for the first time of the WhatsApp mental health awareness intervention. Senders were asked if they were willing to share mental health awareness information with their friends over WhatsApp, as part of an NGO campaign. Conditional on saying yes, the sender was enrolled in the study to be randomized. The final sender sample consisted of 849 senders who listed friends, provided the phone numbers for their friends, and expressed willingness to participate in the campaign.

3.1.3 Targeting Data

The sender survey also collected information the senders' perceptions of their friends' mental health need. After completing the social network elicitation, and within the same survey, the sender was asked to rank his or her friends according to their benefit from mental health resources. Though this question is sensitive, less than five percent of the sender sample declined to do this ranking, suggestive of high rates of trust in the survey's confidentiality. The respondent was asked:

> "Existing research shows that over 50% of people in Jordan are living in distress, including ongoing sadness, helplessness, stress, or having trouble sleeping. If we go

back and think of the [number of friends] friends who you listed, which of them do you think suffer from sadness and stress in their lives, and who would benefit the most from receiving information about identifying and managing psychological distress? Please help me list them in order of who will benefit the most and who will benefit the least."

3.2 Recipient Sample and Baseline

The recipient sample comprised of the senders' friends elicited in the sender survey, and and consisted of 2668 individuals. The final friend networks have little overlap, with only 5% of recipients appearing in more than 1 friend group. This is due in large part to the initial sampling strategy described above, which drew from the refugee population across Jordan.

Short baseline phone surveys were attempted with the new recipients immediately following each sender survey. The timing of recipient baselines and the campaign roll-out were scheduled so that recipient baselines were only attempted before those recipients' senders received the campaign. Of the 2668 recipients, 1422 were reached for a baseline survey. In the survey recipients were assessed for likely depression or anxiety, using the 9-item Patient Health Questionnaire (PHQ-9) (a standard screening tool for depression) and the 2-item Generalized Anxiety Disorder tool to screen for anxiety. Recipients were not informed at that stage of the broader mental health campaign.

3.3 Sender Randomization

Randomization was at the sender level and stratified on gender and the sender's original nominator if the nominator identified multiple senders. Senders were randomized either to treatment (N=642) or control (207), with the treatment group senders asked to share the campaign over WhatsApp with the friends listed in the elicitation. Within the treatment group senders there was additional random variation in how the messages were introduced, which is discussed below. Control group senders were not contacted again for the study and did not receive the awareness content during the experiment period.

Treatment and control are balanced on baseline covariates but some small imbalances arose between individual framing arms (up to an F-statistics of 1.54, see appendix). The primary results are highly robust to forcing the inclusion of imbalanced covariates (above and beyond covariates selected by lasso), shown in Appendix Table D.7.

3.4 WhatsApp Messaging Intervention

The content treatment group senders were asked to share was designed by the International Rescue Committee (IRC) in Jordan to increase awareness of mental health need and direct individuals to a free phone counseling service. Specifically the study informed participants of the Jordan River Foundation's free phone counseling helpline.

The Jordan River Foundation (JRF) is a local Jordanian non-profit that is part of the Queen Rania Foundation – a highly respected organization in Jordan. JRF focuses largely on women and children, and operates a free counseling helpline that is available to all. The helpline is operated by professional counselors, and serves as an entry-point to multiple potential services, including one-time assistance, routine counseling by phone, and, for more acute cases, in-person services with a licensed psychiatrist. Additionally the helpline conducts referrals to other providers. JRF partnered with the study to increase awareness and take-up of these services.

The content itself consisted of awareness messages written in text, infographic-type content such as a comic strip, and links to YouTube videos of a Jordanian psychologist discussing how to recognize and manage common symptoms of distress (see appendix for examples). These were developed by the IRC with behavioral science and human-centered design principals in mind, and received extensive input from Syrian refugee community members and the professional and cultural expertise of a Jordanian psychologist. The campaign aimed to help individuals identify whether they are experiencing distress, learn about self-care approaches, introduce the helpline, and highlight that many people in Jordan have used the helpline.

The content was sent in 3 batches over 8 days, and senders additionally received 3 reminders, one each day after a batch of content was shared.[3] The campaign was administered on a rolling basis in weekly batches, such that senders surveyed in a given week typically began receiving the campaign the following week.

Senders were instructed to copy the content and send it to all of their friends who they had named in the original survey. To remind the sender who to message, the recipient friends' names were listed in the instructions each time the sender received new content to share. Senders were incentivized to share screenshots confirming that they sent the campaign to their friends, and could receive \$1.40 if they shared documentation of sending at least one piece of content to at least one person. Sender incentives were delivered as e-wallet transfers or phone credit transfers,

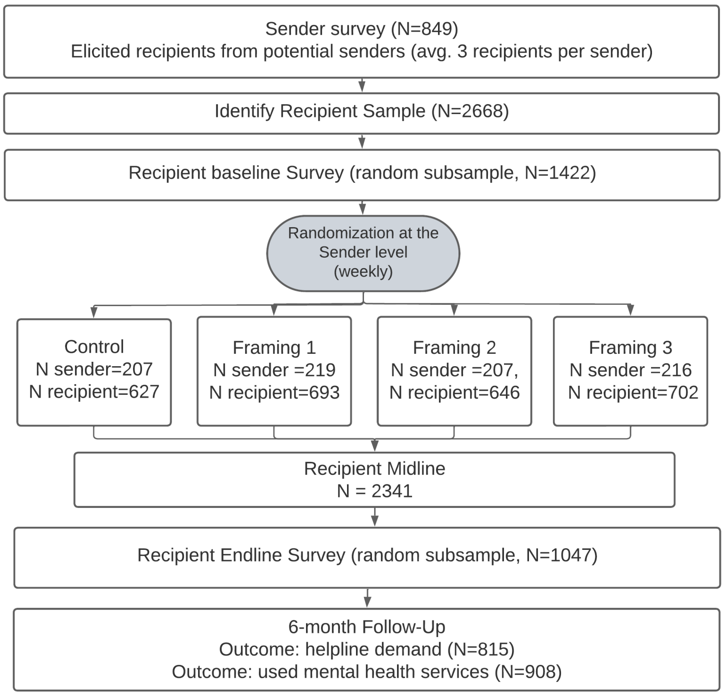
[3] An implementation error caused a random subset of non-disclosed+non-targeted senders to not receive one of the 3 batches of campaign content, and an indicator for this is included in the covariates considered by lasso double selection. The results are qualitatively robust and remain statistically significant if controlling for this or dropping the full affected week, see appendix Table D.7

depending on the respondent's preference.

Follow-up data on recipients collected after the intervention is discussed further below.

The study design is summarized in Figure 1 below.

Figure 1: Experimental Design



3.5 Message Framing Experiment

Senders could worry about what will be signaled by sharing the campaign. Sending the campaign to their friend could signal that they have used the services before themselves, and it could also signal that they think their friend needs mental health services, both of which are stigmatized. Variation in the introduction that senders were asked to share identifies these concerns.

Within the sender treatment group one sentence in the WhatsApp message that introduced the content to the recipient was randomized. The content that the sender was instructed to share began with "Here is some mental health information I received from the International Rescue Committee." This was followed by one of the three sentences below.

> Disclosed Compensation + Non-targeted: An NGO is compensating me to share this with all of my close friends.
>
> Disclosed Compensation + Targeted: An NGO is compensating me to share this with friends who I think can benefit from the information.
>
> Non-Disclosed Compensation + Non-targeted: I want to try to share this with all of my close friend.

In all three conditions the senders were offered the same incentives for participation, and reminded of the incentive each time they received a batch of content or a reminder. And, in all three conditions the sender was told to share the content with all of the friends who she mentioned in the baseline survey, since everyone in this context can benefit from the information, even if only to be able to share it with others. To be clear that all the friends should get the information, all the sender's recipients' names were listed in the instructions that she received with every batch of content. Because the actual compensation and intended recipient group were held constant across senders, the framing conditions vary only the recipient's perception of why the sender shared information.

The randomization achieved balance between treatment and control on 15 of 16 covariates, with an insignificant F-statistic of 1.02. By framing arm, some imbalances arose, with the test of joint significance across covariates for framings 1 and 2 relative to framing 3 being 1.93 and 1.54 respectively, both statistically significant at traditional confidence levels. The imbalanced covariates are included in the list of covariates considered in the pre-specified lasso double selection procedure, and as an additional robustness test I present the primary results when forcing the inclusion of those imbalanced covariates, regardless of whether they are selected by lasso. The robustness test shows that the primary results are not qualitatively different and remain statistically significant when forcing the inclusion of these covariates.

The framings enable me to test for image concerns when sharing this stigmatized information. First, the "disclosed compensation" versus "non-disclosed compensation" comparison tests whether recipients' knowledge of the financial incentive provides social cover that increases senders' sharing. The comparison tests for specifically a social image signaling effect of financial incentives, because the compensation itself is constant across treatments while only visibility of the incentives varies, through disclosure.

The second comparison provided by the framings is that of the "targeted" versus the "non-targeted" framings. This comparison tests whether senders withhold messages that carry a more negative social image signal about the recipient. When the sender tells the recipient that she is trying to send messages to people who may especially benefit, the sender reveals that she think the recipient may be in need. This could be good if it helps the recipient identify that she is good fit for the program. But the sender may worry that her friend will feel insulted or uncomfortable from having her vulnerability revealed. If in fact senders believe and internalize that recipients could feel uncomfortable having their need exposed then senders will be less likely to send the "targeted" framing.

One ex-ante concern could be that revealing that she is paid will be awkward and feel unnatural for the sender. But, on the contrary, the design for the framing arms was borne out of piloting during which, unprompted, many senders informed their friends that an NGO was paying them to share the WhatsApp information. When the behavior persisted even after the research team repeatedly clarified to pilot participants that they did not need to share this information with their friends, it prompted the hypothesis that senders were seeking social cover. Given that some senders were already using the excuse of being paid, the estimated impacts of the framing arms presented in this paper may be underestimates of the effect of social cover.

3.6 Measuring Sender Sharing

Whether the sender shared the campaign with the recipients is measured using all available data collected through sender screenshots and recipient self-reports at midline and endline surveys described below.The primary analysis uses the measure constructed at the recipient-level, where a recipient is recorded to have received a message if they report this in the midline or endline survey, or if their name shows as the message recipient in a screenshot shared by their sender.

The appendix also includes the primary results at the sender-level, where follow-through which is coded as 1 (relative to 0) if the senders shared any screenshots indicating that they shared the content, or any recipients in the sender's friend group report that they received the content. Note that rates of sending and receiving are not perfectly equivalent since, first, a sender might not share with everyone in her friend group, and second, in some sender screenshots it was impossible to conclusively determine the recipient of the WhatsApp message pictured. In these cases the sender was coded as having shared, but no recipient was coded as "1" in return.

Personalized trackable links provide a third way to measure sender sharing. Each of the three batches of content included a personalized trackable link to a YouTube video with mental health awareness content. The link tracking data does not reveal the user's identity, but indicates how many times the link was clicked on by unique devices. Each sender received unique links allowing me to measure which senders' content was engaged with more regardless of whether the sender or recipient self-reported sending or receiving the content.

3.7 Follow-Up Data Collection for Recipients

3.7.1 Midline

Recipients were contacted the week after their sender received the campaign and were asked whether they received the campaign and had used the advertised helpline. 2,341 recipients were reached for this short midline check-in, representing 88% of the total recipient sample.

3.7.2 Endline

An endline phone survey was conducted with a random sample of recipients three weeks after the recipient's last message was scheduled to be received. 1,046 endline surveys were completed. The endline survey collected the recipients' self-reported use of the advertised helpline, as well as their mental health, stigma attitudes and perceptions, and interactions with their social network.

In addition to the endline survey, the helpline conducted a short survey with 98% all first-time callers to the helpline during the study period. With the caller's consent the helpline recorded the caller's phone number in order for it to be matched to the study sample in the analysis.

3.7.3 6 Month Follow-Up

Two distinct samples of data are collected after 6 months. First, the study attempted to reach female recipients and asked if they would like the helpline to contact them directly to receive free phone counseling. 815 female recipients were reached for this question, representing 55% of the female recipient sample. Only female recipients were included due to the helpline's programmatic priorities. Using this data I construct and analyze an indicator variable for helpline demand.

Additionally, the study attempted to survey all recipient households on a series of mental health questions, including whether anyone in the household had ever used mental health services. Because of the helpline's focus on recruiting female users at that time, in households of male recipients the study surveyed a female member of the same household, rather than surveying the original male recipient himself. Because the question on mental health services asked about all household members, the female respondent can be thought of as a proxy respondent for the original male recipient. 908, or approximately a third, of the recipient households were reached for this survey.

3.7.4 Recipient Attrition

Across the recipient surveys attrition was balanced by overall treatment status and by treatment framing arm, with the exception only of the six month mental health follow-up. There, recipients from one of the framing arms were significantly less likely to be found, as shown in the appendix. The results indicate no significant 6-month impacts of that framing arm, and therefore the differential attrition does not meaningfully influence the interpretation of results.

4 Conceptual Framework

A signaling framework in the spirit of Chandrasekhar et al., 2018 demonstrates, first, that when one group systematically expects others to benefit more from information about a stigmatized service, then sharing information carries a costly signal. Second, an observable monetary incentive dampens this. Third, senders will be more image-sensitive when others' priors about them are weak. Fourth and finally, homophily in need can amplify the negative consequences of image sensitivity.

In the stylized model, assume:

Assumption 1. Mental health need (or "vulnerability") V and treatment efficacy beliefs Q are continuously distributed in the population, together yielding an expected benefit $h(V, Q)$ from mental health services, with h increasing in V and Q.

Let prior mental health care users be those who at some point, potentially in the past, faced $h(V_s, Q_s) > \kappa$ where κ is the financial, logistical, and social costs of using services. As such the likelihood of using services is increasing in both need and efficacy beliefs.

A fraction π of the potential sender population are prior users of mental health services (type U) and the remainder are non-users (type N).

Senders choose whether to share information about mental health services with their friends, the potential recipients. Sending is a binary action $S \in \{0, 1\}$ and is chosen to maximize senders' utility which is increasing in the benefit to the recipient and decreasing in the social image loss from sending.

The sender's utility from sending action S is given by

$$U(S) = \underbrace{h(V_r, Q_s)\mathbf{1}_{s=1}}_{\substack{\text{Health benefit}\\ \text{to recipient}\\ \text{(sender's belief)}}} - \underbrace{R(S)}_{\substack{\text{social image cost}\\ \text{to sender}}} - C\mathbf{1}_{s=1} \tag{1}$$

where $h(V_r, Q_s)$ is the sender's expectation of the recipient's health benefit given the recipient's vulnerability V_r and the sender's efficacy beliefs Q_s. $R(S)$ is the image cost of sending, given by the change in the posterior belief that the sender is a mental health care user if the sender shares: $R(S) \equiv \pi_{post,s=1} - \pi_{post,s=0}$.

Let $F_U(\cdot)$ and $F_N(\cdot)$ denote the CDFs of the expected health benefit $h(V_r, Q_s)$ for prior users and non-users respectively, so that $r = 1 - F_N(h^*)$ and $p = 1 - F_U(h^*)$ are the probabilities that each type's perceived benefit exceeds the equilibrium cutoff h^*.

Assumption 2. Prior users have stochastically higher beliefs about recipient benefit, such that $F_N(h) > F_U(h)$ for all h.

The resulting equilibrium is characterized by a cutoff level h^* of expected recipient health benefit above which the senders share and below which they do not (Chandrasekhar et al., 2018). If expected benefit is unrelated to sender's type then the image cost $R(S) = 0$ and senders will share with recipients whose expected health benefit exceeds the fixed cost C, with no inferences made about sender's type.

But the fact that prior users have on average higher efficacy beliefs makes it the case that sharing information conveys a signal about the sender's type. Because $1 - F_U(h^*) > 1 - F_N(h^*)$, prior users send at a higher rate than non-users in equilibrium, so observing that a sender shares will lead observers to update upward their belief that the sender is a prior user. This image cost drives up the equilibrium sending cutoff h^*. In equilibrium prior users will continue to share more than non-prior users (since for the same V_r they believe the recipient benefits more from services), but both groups will face a higher threshold for being willing to share and so less information will be shared overall. See appendix for further details.

4.1 An observable financial incentive mitigates the social image cost

In this situation introducing an observable incentive M can not only directly incentivize sending, but additionally crowd in sending through a secondary effect on the image cost. The sender's utility is then given by

$$U(S) = \underbrace{h(V_r, Q_s)\mathbf{1}_{s=1}}_{\substack{\text{Health benefit}\\ \text{to recipient}\\ \text{(sender's belief)}}} - \underbrace{R(S)}_{\substack{\text{social image cost}\\ \text{to sender}}} - C\mathbf{1}_{s=1} + \underbrace{M\mathbf{1}_{s=1}}_{\substack{\text{monetary incentive}\\ \text{if sends}}} \tag{2}$$

and the equilibrium cutoff will be

$$h^* = R(S) + C - M \tag{3}$$

First, the incentive encourages prior users and non-prior users alike to send more, assuming both types have the same preference for money. This can be seen by the fact that the threshold is decreasing in M:

$$\frac{\partial h^*}{\partial M} = \frac{\partial R(S)}{\partial M} - 1 \tag{4}$$

As long as the image cost does not increase with M, as confirmed next, then the threshold will go down with an increase in the monetary incentive.

Cash indirectly increases sending further by dampening the negative inferences associated with sharing. Since recipients only observe M if the sender shares, $\frac{\partial \pi_{post,s=0}}{\partial M} = 0$, and the change in the threshold for a change in M simplifies to

$$\frac{\partial h^*}{\partial M} = \frac{\partial \pi_{post,s=1}}{\partial M} - 1 \tag{5}$$

This shows that as long as $\frac{\partial \pi_{post,s=1}}{\partial M} < 1$ the monetary incentive will decrease the threshold, and if $0 < \frac{\partial \pi_{post,s=1}}{\partial M} < 1$ then there will additionally be a crowd-in effect, from the image cost decreasing.

Cash induces both users and non-users to share more, which, critically, reduces the proportion of prior users among the people who share. Even when more users than non-users are induced to share, the proportion of non-users will be higher among the marginal senders than those already sharing, and this will dilute the negative signal. Put another way, cash makes it less "telling" that someone who shares is a prior user, by offering a credible excuse for why someone might share while not being a prior-user. The larger the incentive, the more plausible that excuse. As h^* falls, both p and r increase, but the likelihood ratio p/r falls toward 1, reducing $R(S)$. Refer to the appendix for further details.

This closely parallels the familiar intuition that financial incentives might crowd out prosocial actions by dampening the inference about the actor's prosociality (Bénabou and Tirole, 2006). In this setting, the inference is about a stigmatized trait, and so the financial incentive can crowd *in* sharing by dampening the inference about being a prior user.

4.2 Experimentally Identifying Social Image Costs

The secondary effect of the incentive operates purely through the social image channel, which is identified in the experiment through variation in the message framing. The direct effect of the financial incentive, via a preference for money, does not change when the incentive is public or private and rates of sharing are private. But the secondary effect, by which the incentive dampens the signal conveyed by sharing, comes from its observability. Therefore by varying whether others know about the incentive the experiment identifies the social cover effect, $\frac{\partial \pi_{post,s=1}}{\partial M}$.

The same framework can be extended to allow the sender to care not only about their own image but also the recipient's. In the same way that sending is a signal of high efficacy beliefs in senders, it is also a signal of high need in recipients. The experiment tests for evidence of senders attending to recipients' image concerns, by varying the recipient's perception of how targeted sending was. If recipients credibly know there was no targeting, then there is nothing to infer about the recipient's type from receiving the message. Conversely, if recipients are told that they are targeted, then the recipient has greater certainty about the sender's decision criteria, and she will now update her beliefs more conditional on the sender sharing.

This yields two key hypotheses that guide the analysis:

H1: If senders share more when encouraged to disclose that they are paid, while not changing the monetary incentive itself, then image concerns are binding.

H2: If senders share less when assigned to the "targeted" phrasing then recipients' image concerns are binding to the sender.

4.3 Differences in Image Sensitivity

The degree of image sensitivity depends on others' priors about senders' own level of need.[4] As priors get closer to 0 or 1, the signal of sharing becomes less informative, as shown in Chandrasekhar et al., 2018. This follows from the change in the posterior belief about the sender when she shares, given by:

$$R(S) = \pi_{post,s=1} - \pi_{post,s=0} = \frac{\pi(1-\pi)(p-r)}{r+(p-r)\pi} \tag{6}$$

As the prior π that someone is a prior mental health user approaches 0 or 1, the reputation cost $R(S)$ converges to 0.

Importantly this means that different senders can face different image costs, if there are heterogeneous priors about need. People who face more uncertain priors will be more image-sensitive than those who know others already have strong priors about them.

The distribution of these priors is context-specific. In some cases, such as where vulnerability can be partially hidden, it may be easier to infer who has low vulnerability than who has high vulnerability. In other contexts the opposite could be true, and priors about vulnerable people may already be very strong.

This variation in priors can dictate whether senders who have used mental health services before, and therefore have had high mental health need, are more sensitive to image concerns. In the results below I test for heterogeneous priors, and for a correlation between priors and the distribution of need.

4.4 Homophily and Image Sensitivity

Assumption 3. Assume that there is homophily in mental health need, such that the friends of a person with high mental health need are more likely to also have high need than the friends

[4]Technically the relevant object is senders' beliefs about others' priors about the sender's own level of need. Assume for simplicity that senders hold accurate beliefs, such that their beliefs are equivalent to others' priors.

of someone without high mental health need. And, following from that, people who have used mental health services before are on average more connected to others with high mental health need than people who have not used mental health services.

Homophily increases messaging to people in need when image concerns are not positively correlated with need, but this gain in messaging decreases and can become negative when image sensitivity is positively correlated with need. The expected fraction of high-need recipients who receive the message is $P_{HN} = \pi \cdot S_U \cdot f_U + (1-\pi) \cdot S_N \cdot f_N$, where S_U and S_N are the type-specific sending rates and f_U and f_N are the fractions of each type's friends who are high-need. The effect of homophily relative to the no-homophily counterfactual reduces to:

$$\Delta = \delta_U \pi (S_U - S_N) \tag{7}$$

where $\delta_U = f_U - \bar{f} > 0$ is the increase in prior users' high-need connections under homophily. The sign of Δ depends on whether in equilibrium the prior users sending rate S_U is greater than or less than the non-users sending rate S_N. When image costs are not positively correlated with need, prior users' higher efficacy beliefs ensure $S_U > S_N$ and $\Delta > 0$. When image costs are positively correlated with need, the image cost effect can dominate, driving $S_U < S_N$ and $\Delta < 0$. See the appendix for further details.

5 Empirical Strategy

The effect of each framing on sending rates is estimated using first a fully flexible specification and then a specification in which the two disclosure arms are pooled, both at the recipient-level. In all the estimated specifications standard errors are clustered at the sender level.

$$p_r = \alpha_0 + \alpha_1 F_{1s} + \alpha_2 F_{2s} + \alpha F_{3s} + X_s'\beta_1 + X_r'\beta_2 + \Gamma_r + \varepsilon_r \tag{8}$$

$$p_r = \alpha_0 + \psi_1 F_{1or2s} + \psi_2 F_{3s} + X_s'\beta_1 + X_r'\beta_2 + \Gamma_r + \varepsilon_r \tag{9}$$

p_r is a binary indicator of whether the recipient received a message. F_{1s} is an indicator for the "disclosed compensation, non-targeted" framing (Framing 1: "An NGO is compensating me to share this information with all my close friends"). F_{2s} is an indicator for the "disclosed compensation, targeted" framing (Framing 2 : "An NGO is compensating me to share this information with my friends who I think will benefit from the information.") F_{3s} is an indicator for the "non-disclosed, non-targeted" framing (Framing 3: "I want to try to share this information with all my close friends"). F_{1or2s} is an indicator for the sender being assigned to framing 1 or 2.

5% of the recipients appeared in more than 1 sender friend group, and so, because framings were randomized at the sender level, these recipients could receive multiple treatments which were assigned randomly and independently at the sender level. (The results are robust to dropping these recipients with degree greater than 1, shown in the appendix.) X_s' and X_r' are covariate vectors for the sender and recipient respectively and Γ_r are week of survey fixed effects. The final covariates and fixed effects included in each estimation are selected using the lasso double-selection procedure as was pre-specified. Note that the full pre-specified specification tested the same hypotheses using a less easily digestible specification[5], the results of reflect the same pattern of results and are provided in the appendix (see Table D.2). Additionally 8 of the 51 pre-specified covariates and fixed effects were inadvertently dropped from the baseline data collection and therefore not included.

The same specifications are also estimated using the sender-level outcome of link clicks. In those instances the vector of recipient-level covariates takes the median of the recipient outcomes in the sender's friend group, and the error term is at the sender level (ε_s). The three framings were randomized mutually exclusively over the sender sample and together comprise the complete treatment group.

When it comes to recipient impacts, I consider both the pooled effect of receiving the campaign, and the effect of specific message framing arms as pre-specified. The pooled effect estimated using the following two stage least squares specification:

$$T_s = \gamma_0 + \gamma_1 A_s^T + \gamma_2 A_r^T + X_r'\lambda_1 + X_s'\lambda_2 + \Gamma_r + \nu_r \tag{10}$$

$$y_r = \pi_0 + \pi_1 \widehat{T}_s + +X_r'\phi_1 + X_s'\phi_2 + \Gamma_r + \eta_r \tag{11}$$

where, given the low follow-through rates, T_s is an indicator taking 1 if the sender shared any messages (rather than only if the specific recipient was known to receive a message as was originally pre-specified) and y_r are recipient outcomes such as mental health care take-up and social connectedness. Treatment is instrumented using both the initial sender's assignment to treatment A_s^T and assignment for any sender linked to the recipient (A_r^T), to account for the 5% of recipients who appeared in more than 1 sender friend group. As before X_s', X_r', and Γ_r are the vectors of sender covariates and recipient covariates, and survey week fixed effects, respectively. The specific covariates and fixed effects included in each estimation are selected using the lasso

[5] $log\left(\frac{p_r}{1-p_r}\right) = \alpha_0 + \alpha_1 I_s^S + \alpha_2 I_s^R + \delta I_s^S I_s^R + X_s'\beta_1 + X_r'\beta_2 + \Gamma_r + \varepsilon_r$ where I_s^S is an indicator for the sender being assigned to a framing that alleviates the "sender social image concern" (ie. F_1 or F_2 framing), and I_s^R is an indicator for the sender being assigned a framing that alleviates the "recipient social image concern" (ie. F_1 or F_3), and the interaction of the two takes one when both concerns are alleviated (corresponding to F_1 only).

double-selection procedure.

For the effect of the campaign depending on message framing, I focus on the intent to treat estimates, since the different framings have differential sending rates:

$$y_r = \alpha_0 + \alpha_1 F_{1s} + \alpha_2 F_{2s} + \alpha F_{3s} + X_s'\beta_1 + X_r'\beta_2 + \Gamma_r + \varepsilon_r \tag{12}$$

with all variables following the definitions given above and standard errors are clustered at the sender level.

Lastly, for the follow-up experiment I estimate the effects of the message framings themselves, using the intent to treat estimates. Recall from Section 3.8 that the follow-up experiment used roughly the same three introductions as the main experiment, with the addition of a "non-disclosed, targeted" phrasing. And, in order to avoid endogenous selection in who the sender chose to message, in the follow-up experiment respondents gave the study enumerators permission to contact their friends and introduce the helpline using any of four introductions, referencing the sender's name. The primary outcome is whether the (new) recipients agree to be contacted by the helpline. I obtain the intention to treat estimates using the pre-specified specification below where k_r is a binary indicator for whether the recipient consents to be contacted.

$$k_r = \beta_0 + \beta_1 Disclose_r + \beta_2 Target_r + \delta DiscloseXTarget_r + X_s'\phi_1 + X_r'\phi_2 + \Gamma_r + \varepsilon_r \tag{13}$$

$Dislose_r$ takes 1 if the recipient was assigned to the "disclosed, targeted" or "disclosed, non-targeted" framing and 0 otherwise; $Target_r$ takes 1 if the recipient was assigned to the "disclosed, targeted" or non-disclosed, targeted" framing and 0 otherwise, and $DiscloseTarg_r$ takes 1 if the recipient was assigned to the "disclosed, targeted" framing and 0 otherwise. X_s and X_r are vectors of sender-level and recipient-level covariates, Γ_r are survey week fixed effects, and standard errors are clustered at the sender level. Note that because there is no pure control group, the "non-disclosed compensation, non-targeted" group is the omitted reference category.

6 Main Results

The main analysis first establishes individuals' private knowledge of their friends' mental health status. Second I test how social image concerns affect senders' willingness to share information with their friends. I explore mechanisms using heterogeneity analysis, including whether senders utilize their knowledge of who will benefit most from mental health information when deciding whether to share socially uncomfortable information. Lastly I present impacts on recipients' demand for mental health services and other related outcomes.

6.1 Sender Knowledge

A key rationale for involving community members in outreach efforts is that they have private knowledge about who will benefit most from programs. This could be particularly valuable when fit for a program is not easily observable, as in the case of mental health.

Comparing senders' ranking of who would benefit the most from mental health care to recipients' mental health outcomes shows that senders have quite accurate knowledge. Column 1 of Table 1 regresses a recipient-level indicator for being the recipient the sender thinks would benefit the most on an indicator for whether the recipient likely has depression at baseline, given their PHQ-9 score. This predictive analysis reveals that the friend identified as more in need is 10.7 percentage points, or 24% more likely to have depression (p-value= 0.002). The relationship drops only to 9.4 when controlling for demographics, indicating that senders have information above and beyond the standard observable characteristics that could be easily used for targeting (p-value= 0.001). The same analysis with anxiety instead of depression shows that senders' rankings are less strongly predictive of anxiety, but still have a positive and marginally significant correlation (point estimate 5.1 percentage points, p-value= 0.073). Once controlling for demographics this relationship is insignificant. The difference in targeting accuracy may be related to the fact that anxiety was measured using the 2-item GAD-2 and therefore is less precise than the depression indicator which was measured using the 9-item PHQ-9.

6.2 Experimental Effect of Message Framing on Sending Rates

Despite senders agreeing at baseline to share the campaign, knowing their friends are in need, receiving frequent reminders, and being financially incentivized, most recipients never receive the campaign. As shown in Table 2, assignment to treatment led to only a 21 percentage point increase in recipients receiving the campaign, and only 16% of senders had any links clicked on. This first stage is statistically strongly significant, but the low rates of sharing are consistent with the baseline evidence that mental health is a sensitive topic that people rarely discuss.

Table 2 also shows that 6.3% of the control group recipients reporting that they saw the campaign. When only counting recipients who also told the study the name of the person who shared the content the spillovers drop to 1.2% (see appendix). Given these small spillovers the estimated impacts on recipients may be under-estimated.

Table 3 presents the main result of the effect of message framing on sharing rates. Consistent with social cover increasing sharing, encouraging senders to disclose that they were paid led to a 7.2 percentage point increase, equivalent to a 33% increase relative to the non-disclosed group (p-value= 0.047). This estimate comes from comparing the "disclosed, non-targeted" and "non-disclosed, non-targeted" rates of sending presented in column 1. There is a similar 7 percentage point increase in click rates, equivalent an even larger 57% increase, given the low click rates overall (p-value = 0.048).

I find little evidence that senders withhold messages that signaled that the recipient was targeted on need. The difference in sending rates for the "targeted" and "non-targeted" framings (holding disclosed compensation constant) is less than 1 percentage point and has a p-value of 0.85. The difference in click rates is marginally greater at 1.9 percentage points, but still statistically insignificant. This may mean that senders internalize their own image concerns more than those of the recipient, but could also mean that the encouragement design did not induce first-stage differences in which framing the sender used between the two disclosed compensation arms.

Given the negligible differences between the "targeted" and "non-targeted" version of the "disclosed compensation" framing, I combine the two into one pooled "disclosed compensation" framing that I compare to the "non-disclosed, non-targeted" framing. The pooled point estimates shows a similar 7.7 percentage points increase in sharing when providing social cover through the disclosed compensation (p-value = 0.013), and a 6.5 percentage point increase in click rates (p-value = 0.029).

6.3 Heterogeneity in Image Sensitivity

Hereogeneity analysis uncovers large differences in image sensitivity. Using machine learning heterogeneity analysis following Chernozhukov et al., 2018 I first test whether in fact the disclosure effect has heterogeneous effects on the sample, and if yes, then which covariates are associated with the difference in treatment effects. Figure 2 shows the group average treatment effects of disclosure compared to non-disclosure from the least to most affected quantiles of the sample. The comparison of G5-G1 shows that indeed there are large significant differences between the treatment effects for the least affected and most affected quantiles. To understand what characteristics are associated with these heterogeneous treatment effects I follow Chernozhukov et al., 2018 in comparing the average characteristics of participants in the most affected quantile to the characteristics of participants in the least affected quantile after retricting to characteristics that vary significantly between the first and fifth quantiles at the 99% confidence level.

One characteristic dominates any other: whether the sender is a prior mental health care user herself. Senders who are prior users are 25 percentage points (125%) more likely to send the disclosed compensation framing than those who have not used mental health services before (Table 4, p-value$<$ 0.001). If responsiveness to the disclosure compensation were driven by something besides social image concerns, such as salience of the financial incentive, it is unlikely that it would generate this pattern of heterogeneity. Instead, the pattern is consistent with disclosure providing social cover that prior users highly demand.

Data on prior users' beliefs about the benefits of mental health services corroborates that prior users are more image-sensitive. Senders who have used mental health services before, and particularly those who have used them recently, are more likely to believe that mental health services are effective (Table 5). Despite perceiving a higher benefit from services, without social cover these senders share at similarly low rates to non-users (Table 4), consistent with facing higher costs from sharing. Together these results point to prior users perceiving both high practical benefits and also high reputational costs from sharing information about mental health services. This makes them well-positioned to respond to reductions in image costs.

Returning to the overall sender heterogeneity, notably the effect of disclosure is actually negative for one quantile of the sender sample, and this quantile is the least likely to have used mental health services (see Figure 2). This indicates further evidence of the signaling mechanism. Using the disclosure excuse has the benefit of dampening the signal that the sender may be a mental health care user, but comes with the cost of also dampening the signal that the sender is prosocial. For senders who are less worried about others thinking they used mental health services, the cost of the negative signal dominates, and they become less likely to share, demonstrating classic prosocial crowd-out (Bénabou and Tirole, 2006).

The prior-user heterogeneity also offers evidence that some senders may actually also be sensitive to the more targeted framing that could cause the recipient to feel singled out. Among prior-users, senders are 15.8 percentage points (32%) less likely to send the more targeted phrasing, holding constant the disclosure framing (p-value = 0.079). This is only marginally significant, but points to these prior users being more sensitive to image concerns for not only themselves but also their friends.

Other characteristics strong associated with heterogeneity identified by the machine learning approach are also listed in the appendix (see Appendix Table D.10), and include the sender's (lower) stigma beliefs and number of social connections as additional factors. In addition to the machine learning heterogeneity I test 5 pre-specified sender characteristics for heterogeneity: mental health service efficacy beliefs, own stigma views, altruism, gender, and social desirability. Of these the effect of disclosure varies only with stigma views, with lower stigma senders being more responsive to the disclosure treatment, whereas senders with high stigma are unaffected (see appendix Table D.8). This aligns with the reduction in image costs only mattering for senders who are near the threshold of being willing to share, and is consistent with the machine learning heterogeneity results.

6.4 Why prior users face higher image costs

The heterogeneity and efficacy belief results above show that user experience may be a double edge sword. On one hand learning may be most useful from someone with experience. But exactly the qualities that would lead someone to seek care – high efficacy beliefs and high need – can cause them to be the most image-constrained.

As outlined in the conceptual framework earlier, differences in efficacy beliefs are exactly what can make the choice to share a costly signal of type. Because prior users are more likely to think services are effective, they are more likely to share, and therefore people who share are perceived to likely be users. That result alone, though, cannot explain differential image sensitivity. Rather it identifies an image cost that any sender would face when sharing.

The second factor that can characterize prior users – their elevated need –, can also cause them to face higher image sensitivity. The signaling framework predicts that the costliness of the sharing signal is greatest when there is high uncertainty over whether the person is a user or not. Recall the vignette posed in the introduction: seeing a person who appears to be unhoused or a wealthy philanthropist leave a food pantry reveals little. The prior that the unhoused person is vulnerable and in need was already strong, and an observer is unlikely to conclude that the philanthropist is actually financially struggling since they were seen leaving the pantry. (More likely they will be assumed to be donating.) On the other hand a lower/middle-income person may worry about being seen since more might be inferred about them. This captures the mechanics of Bayesian learning in which a signal will be most informative when priors are noisy, and, in contrast, the signal's informativeness is lowest when the prior nears 0 or 1 (Chandrasekhar et al., 2018). This offers the prediction that uncertainty about people's need is high when image concerns hamper information sharing, as seen in this context.

I investigate this intuition in the data using respondents' guesses of what they believe their friends think the respondent's distress is on a scale of 1 to 10.[6] Figure 4 shows that these priors become much noisier at higher levels of need. Because more vulnerable individuals expect that their friends have noisier priors about their distress level, theory predicts that they will be more

[6] I focus here on beliefs about their friends' priors (ie. "second order priors") rather than actual priors, since individuals are making choices based on how they think their friends will update. These outcomes were collected only among the potential recipient sample, but were measured at baseline and therefore likely generalize to the recipients' friends, the senders.

sensitive to an image signal. Moreover, there is a clear perception gap that grows with level of need: high need respondents do not believe their friends are fully aware of their need. This gives them a stake in continuing to protect their reputation. These results confirm the conditions that lead to variation in image sensitivity.

6.5 Homophily and targeting consequences

The results above establish that prior users more image-sensitive. Importantly, these individuals are also more connected to vulnerable people in need of mental health services. Senders who are prior users have recipient friends who are 8 percentage points (15%) more likely to have depression or anxiety at baseline, compared to the friends of non-user senders (p-value< 0.05, see Table 5). This pattern is even stronger when controlling for demographics (9 percentage point difference, p-value < 0.01).

This homophily could be helpful for community based targeting, but instead works directly against it since users are the most image-sensitive. Meanwhile the less image-sensitive senders (non-users) are less connected to the people in need.

As a result, the gains from inducing more sharing from prior users are particularly high. The recipients prior user reach are 11 percentage points (23%) more likely to be experiencing anxiety or depression than those reached by non-users (p-value < 0.10). Controlling for demographics strengthens the result, showing to a 13.7 percentage point higher likelihood of depression or anxiety among prior-users' contacted recipients, relative to the recipients contacted by non-prior users (p-value< 0.05). This finding is particularly promising in the context of prior work on the take-up of social benefits, which points to a trade-off, whereby even when interventions successfully alleviate frictions, the marginal applicants are often systematically less in-need (Castell et al., 2025; Finkelstein and Notowidigdo, 2019). The results presented here, while focused on the information transmission step rather than take-up, show it is possible to increase informational awareness without a targeting trade-off. Doing so, however, relies on leveraging experienced users, who may not want to share information due to image constraints.

6.6 Alternative Explanations of the Disclosure Effect

The key experimental variation comes from encouraging some senders to disclose that they are paid, which was designed to alleviate senders' image concerns by giving them an excuse for sharing. Below I discuss alternate potential mechanisms of the disclosure.

Reminder effect: If the disclosure that the sender is being paid operates through a reminder effect, then it could pose a confound to the social cover mechanism. This risk was minimized by delivering numerous reminders about the payment that all senders. Regardless of treatment arm, every sender was reminded about the incentive on 6 different days across the week of their participation, either via a new bundle of content or a reminder to share the content. This makes the reminder effect less plausible, but does not on its own rule that mechanism out.

However, if disclosure were simply about a reminder effect then there would be no reason that prior mental health care users would be more responsive. This strong heterogeneity cannot be explained by a reminder effect.

Self justification: A sender who is concerned about their friends' reaction to the information might justify their actions to themselves by thinking "I am being paid for this so I may as well do it". This however is cleanly ruled out by the disclosure treatment. Since senders in all treatment arms are paid, the differential effect of treatment cannot be attributed to the sender's self-justification based on payment.

Painful to discuss these topics: Rather than mental health being stigmatized, it could be that mental health is simply painful to talk about, for the sender or the recipient. If that were the underlying reality, then the payment and disclosure would still operate through a justification mechanism, but without being related to stigma per se. However, if the reason that people are reticent to share is specifically because it is painful to talk about these topics, then treatment should only induce one-time willingness to engage, after which people would have no incentive to have more, presumably painful, conversations. Instead, the recipient-level results discussed below show that treatment leads to significantly more conversations, extending beyond the intervention itself. This is consistent with a model where conversations are productive rather than painful, but never start because of the image risks of taking the first step.

"Non-disclosed" group disclosure: A separate concern is that even senders in the "non-disclosed" treatment arm may have disclosed that they were being paid. Indeed 5.6% of senders in the non-disclosed arm sent the study a screenshot showing that they had shared the study instructions with their recipient, which may have revealed that they were being paid. It is possible that even more non-disclosure arm senders shared this information with their recipients and did not convey it to the study via the screenshots.

Empirically, this means that the results may under-estimate the effect of social cover since some comparison group senders also used the excuse. Additionally, the fact that some senders spontaneously disclosed their payment without prompting is itself consistent with the social cover mechanism. These senders, of their own initiative, identified the payment as a useful excuse and chose to leverage it. This mirrors exactly what the treatment encouraged, and suggests that demand for social cover occurs organically and is not purely an artifact of the experimental design.

This raises the question of why all senders would not use the excuse provided by the financial incentive. The interpretation most consistent with the results is that most senders in the non-disclosed compensation group either did not think of the fact that they could leverage the payment as an excuse, or faced some psychological barrier to using the excuse. It is also possible that senders were hesitant to foreground the financial benefit they were receiving. These point to potential cognitive and emotional costs of formulating and using the excuse, and the negative image signal that can also accompany revealing a financial incentive.

It is notable then that when respondents were randomly prompted to use an excuse that they might not have thought of or were not sure was condoned, then their behavior changed and they did share more often. The study's encouragement design was predicated on the assumption that these marginal senders existed. The fact that this excuse is not one that all senders are naturally inclined to use is useful for identifying the social cover mechanism, and also indicates that take-up of social cover might be even higher when used in more naturalistic formulations, such as saying "I committed to the NGO that I would do this," as opposed to saying the NGO is paying the person.

6.7 Recipient Impacts

The main effects of the campaign came through increased social connectedness and conversations with friends about mental health. Receiving the campaign led to a 15.9 percentage points increase in the likelihood that the recipient had any conversations about mental health (not counting the campaign itself) in the past 6 months (p-value = 0.011, Table 6). In this context where, despite an enormous mental health burden, talking about mental health is exceedingly rare, the treatment impact amounts to a 144% increase in conversations. The measurement of this outcome excludes the communication of the campaign itself, showing that the initial interaction precipitated continued discussions between friends. Consistent with this, 80% of recipients who report getting the campaign say the content made them feel supported[7] and 89% were glad someone shared the information with them.

An index of social connectedness also increases by 0.37 standard deviations (p-value = 0.046). This index is based on actual interactions (rather than perceptions or attitudes), such as how many times the respondent socialized with, spoke on the phone with or helped or was helped by people outside their household. One of the largest impacts contributing to the index comes through time spent helping others or being helped, which increases by 0.45 standard deviations

[7] Of the remainder, 15% felt neutral about receiving the information, and only 5% say they felt uncomfortable.

(p-value = 0.026).

These are the only estimated recipient impacts that survive an FDR multiple hypothesis testing adjustment. The full set of outcomes and q-values can be found in the appendix.

The ITT estimates broken out by framing arm show that the positive effects are predominantly driven by the "disclosed, targeted" framing. These differences are not statistically significant, but the pattern of results is notable. If recipients did react more to more targeted phrasing, it could be because of the framing itself or because of differences in who the sender chose to share with depending on the framing. Given that there is no evidence of targeting on baseline depression or anxiety between this study arm and the "disclosed, non-targeted", the most likely explanation is that recipients reacted positively to feeling their need was noticed by their friend.

The evidence on recipient take-up of mental health services is noisy but still offers some insights. First, the study can reject a sharp null hypothesis that the campaign led to take-up of the phone counseling helpline. Receiving the campaign led to an insignificant 1.7 percentage point decrease in the probability of take-up, relative to a control mean of 2 percent. Second, there is no significant effect 6 months later of treatment on the recipient's willingness to have the helpline contact them directly, though rates of agreement are high, with 54% of the control group agreeing.

However, also 6 months after the intervention, there is the suggestion of noisy positive impacts on having used *any* mental health services, such as in-person counseling or medication. The ITT estimates by framing arm show a noisy significant effect of the disclosed, non-targeted framing (which was shared the most). Assignment to this framing led to a 7.2 percentage point (31%) increase in the probability that the recipient ever used any mental health services (p-value= 0.082, ITT estimate). This estimate, which is driven by therapy use as shown in the appendix, does not survive a multiple hypothesis testing adjustment, but suggests at potential positive impacts on care-seeking. The TOT estimate is insignificant at traditional confidence levels but is large, at 11.8 percentage points,equivalent to a 51% increase in use of care. This points to the possibility that awareness campaigns in this context may increase refugees' demand for high touch services more so than low-touch services such as a phone helpline. Furthermore initiating conversations between friends may have helped participants learn about these alternative services.

7 Conclusion

Across a wide range of settings, services designed to help the most vulnerable go underused. Large knowledge gaps persist even when information exists within communities, and efforts to close these gaps through outreach and peer referrals have had mixed success. Understanding why information fails to circulate, and who it fails to reach, is essential for the design of programs serving disadvantaged populations, not just for mental health, but for food banks, unemployment assistance, substance use treatment, and other services.

This paper provides evidence that stigma can be a central part of the answer. By documenting mental health need across both sides of a large number of social ties, drawn from a structured sample of the Syrian refugee population in Jordan, the paper traces the distributional consequences of stigma on learning. This reveals that image concerns do not just suppress information sharing, but that they do so in a pattern that is particularly harmful for learning.

The findings suggest that studies examining stigma through individual take-up decisions may substantially underestimate its true cost. One of stigma's largest effects may be on who learns about services rather than on whether informed individuals choose to use them. Prior service users, who hold the strongest efficacy beliefs and are most connected to high-need peers, are precisely the people most deterred from sharing, turning firsthand experience from an asset into a liability. If experienced users of stigmatized services systematically withhold referrals and recommendations, then persistent knowledge gaps within vulnerable communities may be sustained in large part by stigma itself. Destigmatization efforts may therefore deliver greater informational benefits than previously recognized.

Beyond information transmission, the study documents untapped demand for social connection. Despite the deep reluctance to initiate sensitive conversations, in the study the reception was almost uniformly positive. A large majority of recipients were glad someone shared information, felt supported by it, and then continued to have deeper interactions with their friends, showing that a remarkably light-touch approach can break through the barrier. The findings point to the promise of informal support interventions as a complementary path for tackling the growing global burden of mental distress and loneliness. The results here are an initial step, and suggest the value of further research exploring the potential and limits of these informal support approaches.

That the evidence in this study comes from a population of forcibly displaced people, who face among the highest mental health burdens and lowest rates of service access of any group globally yet remain severely understudied, underscores both the practical stakes and the scope for this type of research to inform policy in settings where it is most needed.

References

Acampora, M., Capozza, F., & Moghani, V. (2022). Mental health literacy, beliefs and demand for mental health support among university students. *Beliefs and Demand for Mental Health Support among University Students (October 29, 2022).*

Agüero, J. M., Aldana, Ú., Field, E., Frisancho, V., & Romero, J. (2020). Is community-based targeting effective in identifying intimate partner violence? *AEA Papers and Proceedings*, *110*, 605–609.

Aizer, A. (2003). Low take-up in medicaid: Does outreach matter and for whom? *American Economic Review*, *93*(2), 238–241.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021). Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics*, *136*(4), 2147–2194.

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2016). Network structure and the aggregation of information: Theory and evidence from indonesia. *American Economic Review*, *106*(7), 1663–1704.

Anders, J., & Rafkin, C. (2022). The welfare effects of eligibility expansions: Theory and evidence from snap. *Available at SSRN 4140433*.

Aral, S., & Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science*, *57*(9), 1623–1639.

Asanov, A.-M., Asanov, I., & Buenstorf, G. (2024). A low-cost digital first aid tool to reduce psychological distress in refugees: A multi-country randomized controlled trial of self-help online in the first months after the invasion of ukraine. *Social Science & Medicine*, *362*, 117442.

Banerjee, A., Breza, E., Chandrasekhar, A. G., & Golub, B. (2018). *When less is more: Experimental evidence on information delivery during india's demonetization* (tech. rep.). National Bureau of Economic Research.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, *86*(6), 2453–2490.

Banerjee, A., Duflo, E., Grela, E., McKelway, M., Schilbach, F., Sharma, G., & Vaidyanathan, G. (2023). Depression and loneliness among the elderly in low-and middle-income countries. *Journal of Economic Perspectives*, *37*(2), 179–202.

Banerjee, A., Hanna, R., Olken, B. A., & Lisker, D. S. (2024). Social protection in the developing world. *Journal of Economic Literature*, *62*(4), 1349–1421.

Baranov, V., Bhalotra, S., Biroli, P., & Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American economic review*, *110*(3), 824–859.

Barker, N., Bryan, G., Karlan, D., Ofori-Atta, A., & Udry, C. (2022). Cognitive behavioral therapy among ghana's rural poor is effective regardless of baseline mental distress. *American Economic Review: Insights*, *4*(4), 527–545.

Baseler, T., Ginn, T., Hakiza, R., & Woldemikael, O. (2023). Can redistribution change policy views? aid and attitudes toward refugees in uganda.

Batmanov, A., Grigoryeva, I., Calderón-Hernández, B., González-Téllez, R., & Ramírez, A. G. (2026). Beliefs, information sharing, and mental health care use among university students. *Journal of Development Economics*, *180*, 103646.

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, *111*(6), 1918–1943.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, *96*(5), 1652–1678.

BenYishay, A., & Mobarak, A. M. (2019). Social learning and incentives for experimentation and communication. *The Review of Economic Studies*, *86*(3), 976–1009.

Berg, E., Ghatak, M., Manjula, R., Rajasekhar, D., & Roy, S. (2017). Motivating knowledge agents: Can incentive pay overcome social distance? *The Economic Journal*, *129*(617), 110–142. https://doi.org/10.1111/ecoj.12544

Bertrand, M., Luttmer, E. F., & Mullainathan, S. (2000). Network effects and welfare cultures. *The quarterly journal of economics*, *115*(3), 1019–1055.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment. *The Quarterly Journal of Economics*, *127*(3), 1205–1242.

Bharadwaj, P., Pai, M. M., & Suziedelyte, A. (2015). *Mental health stigma* (tech. rep.). National Bureau of Economic Research.

Bhargava, S., & Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an irs field experiment. *American Economic Review*, *105*(11), 3489–3529.

Bhat, B., De Quidt, J., Haushofer, J., Patel, V. H., Rao, G., Schilbach, F., & Vautrey, P.-L. P. (2022). *The long-run effects of psychotherapy on depression, beliefs, and economic outcomes* (tech. rep.). National Bureau of Economic Research.

Blackmore, R., Boyle, J. A., Fazel, M., Ranasinha, S., Gray, K. M., Fitzgerald, G., Misso, M., & Gibson-Helm, M. (2020). The prevalence of mental illness in refugees and asylum seekers: A systematic review and meta-analysis. *PLoS medicine*, *17*(9), e1003337.

Bloom, D. E., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L. R., Fathima, S., Feigl, A. B., Gaziano, T., Hamandi, A., Mowafi, M., et al. (2012). *The global economic burden of noncommunicable diseases* (tech. rep.). Program on the Global Demography of Aging.

Breza, E., & Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica*, *87*(1), 175–216.

Breza, E., & Kaur, S. (2026). *Psychology and development: Applications from cognitive and social psychology* (tech. rep.). National Bureau of Economic Research.

Burchert, S., Alkneme, M. S., Alsaod, A., Cuijpers, P., Heim, E., Hessling, J., Hosny, N., Sijbrandij, M., van't Hof, E., Ventevogel, P., et al. (2024). Effects of a self-guided digital mental health self-help intervention for syrian refugees in egypt: A pragmatic randomized controlled trial. *PLoS Medicine*, *21*(9), e1004460.

Bursztyn, L., Egorov, G., Haaland, I., Rao, A., & Roth, C. (2023). Justifying dissent. *The Quarterly Journal of Economics*, *138*(3), 1403–1451.

Castell, L., Gurgand, M., Imbert, C., & Tochev, T. (2025). Take-up of social benefits: Experimental evidence from france. *American Economic Journal: Economic Policy*, *17*(4), 1–29.

Cefalà, L., Irakoze, F., Naso, P., & Swanson, N. (2024). The economic consequences of knowledge hoarding.

Chandrasekhar, A. G., Golub, B., & Yang, H. (2018). *Signaling, shame, and silence in social learning* (tech. rep.). National Bureau of Economic Research.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india* (tech. rep.). National Bureau of Economic Research.

Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J. S., & Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? a systematic review of quantitative and qualitative studies. *Psychological medicine*, *45*(1), 11–27.

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, *100*(1), 35–69.

Cronin, C. J., Forsstrom, M. P., & Papageorge, N. W. (2025). What good are treatment effects without treatment? mental health and the reluctance to use talk therapy. *Review of Economic Studies*, *92*(3), 1699–1737.

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive–behavioural therapy and otherpsychological treatments for adult depression: Meta-analytic study ofpublication bias. *The British Journal of Psychiatry*, *196*(3), 173–178.

Currie, J. (2004). The take up of social benefits.

De Graaff, A. M., Cuijpers, P., Twisk, J. W., Kieft, B., Hunaidy, S., Elsawy, M., Gorgis, N., Bouman, T. K., Lommen, M. J., Acarturk, C., et al. (2023). Peer-provided psychological intervention for syrian refugees: Results of a randomised controlled trial on the effectiveness of problem management plus. *BMJ Ment Health*, *26*(1).

Derksen, L., Muula, A., & van Oosterhout, J. (2022). Love in the time of hiv: How beliefs about externalities impact health behavior. *Journal of Development Economics*, *159*, 102993.

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments* (pp. 3–93, Vol. 2). Elsevier.

Erik, S. S., Aragie, K. T., & Maha, K. (2021). Impact of work permits on decent work for syrians in jordan, reliefweb. *URL: https://reliefweb. int/report/jordan/impact-work-permits-decent-work-syrians-jordan-september-2021-enar.*

Finkelstein, A., & Notowidigdo, M. J. (2019). Take-up and targeting: Experimental evidence from snap. *The Quarterly Journal of Economics*, *134*(3), 1505–1556.

Goldberg, J., Macis, M., & Chintagunta, P. (2023). Incentivized peer referrals for tuberculosis screening: Evidence from india. *American Economic Journal: Applied Economics*, *15*(1), 259–291.

Goldberg, S. B., Lam, S. U., Simonsson, O., Torous, J., & Sun, S. (2022). Mobile phone-based interventions for mental health: A systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS digital health*, *1*(1), e0000002.

Gustafsson, B. (2002). Assessing non-use of social assistance. *European Journal of Social Work*, *5*(2), 149–158.

Hansen, T., Nes, R. B., Hynek, K., Nilsen, T. S., Reneflot, A., Stene-Larsen, K., Tornes, R. A., & Bidonde, J. (2024). Tackling social disconnection: An umbrella review of rct-based interventions targeting social isolation and loneliness. *BMC Public Health*, *24*(1), 1917.

Harker Roa, A., Córdoba Flechas, N., Moya, A., & Pineros-Leano, M. (2023). Implementing psychosocial support models in contexts of extreme adversity: Lessons from a process evaluation in colombia. *Frontiers in psychology*, *14*, 1134094.

Haushofer, J., Mudida, R., & Shapiro, J. P. (2020). *The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being* (tech. rep.). National Bureau of Economic Research.

Hussam, R., Kelley, E. M., Lane, G., & Zahra, F. (2022). The psychosocial value of employment: Evidence from a refugee camp. *American Economic Review*, *112*(11), 3694–3724.

Islam, A., Mozumder, T. A., Rahman, T., Shatil, T., & Siddique, A. (2021). Forced displacement, mental health, and child development: Evidence from the rohingya refugees. *Working Paper*.

Jain, R., & Khandelwal, V. (2024). *Silent networks: The role of inaccurate beliefs in reducing useful social interactions* (tech. rep.). Working Paper.

Karing, A. (2018). Social signaling and childhood immunization: A field experiment in sierra leone. *University of California, Berkeley*, *2*.

Khedari, V. K. (2020). *Out of ivory towers and into refugee camps: Providing refugees with accessible resources about the biological and psychological effects of forced migration* [Doctoral dissertation, The New School].

Kimbrough, L. W., Fisher, H. E., Jones, K. T., Johnson, W., Thadiparthi, S., & Dooley, S. (2009). Accessing social networks with high rates of undiagnosed hiv infection: The social networks demonstration project. *American journal of public health*, *99*(6), 1093–1099.

Ko, W., & Moffitt, R. A. (2024). Take-up of social benefits. *Handbook of Labor, Human Resources and Population Economics*, 1–42.

Krafft, C., Sieverding, M., Berri, N., Keo, C., & Sharpless, M. (2022). Education interrupted: Enrollment, attainment, and dropout of syrian refugees in jordan. *The Journal of Development Studies*, *58*(9), 1874–1892.

Lacey, L., Mishra, N., Mukherjee, P., Prakash, N., Prakash, N., Quinn, D., Sabarwal, S., & Saraswat, D. (2025). Can destigmatizing mental health increase willingness to seek help? experimental evidence from nepal. *Journal of Policy Analysis and Management*, *44*(1), 97–124.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).

Moshe, I., Terhorst, Y., Philippi, P., Domhardt, M., Cuijpers, P., Cristea, I., Pulkki-Råback, L., Baumeister, H., & Sander, L. B. (2021). Digital interventions for the treatment of depression: A meta-analytic review. *Psychological bulletin*, *147*(8), 749.

Musheke, M., Ntalasha, H., Gari, S., Mckenzie, O., Bond, V., Martin-Hilber, A., & Merten, S. (2013). A systematic review of qualitative findings on factors enabling and deterring uptake of hiv testing in sub-saharan africa. *BMC public health*, *13*(1), 220.

Office of the Surgeon General (OSG). (2023). *Our epidemic of loneliness and isolation: The u.s. surgeon general's advisory on the healing effects of social connection and community* (PMID: 37792968). U.S. Department of Health and Human Services. Washington, DC. https://www.ncbi.nlm.nih.gov/books/NBK589732/

Osman, A., & Speer, J. D. (2024). Stigma and take-up of labour market assistance: Evidence from two field experiments. *Economica*, *91*(361), 123–141.

Patel, V., Weobong, B., Weiss, H. A., Anand, A., Bhat, B., Katti, B., Dimidjian, S., Araya, R., Hollon, S. D., King, M., et al. (2017). The healthy activity program (hap), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in india: A randomised controlled trial. *The Lancet*, *389*(10065), 176–185.

Pescosolido, B. A., Medina, T. R., Martin, J. K., & Long, J. S. (2013). The "backbone" of stigma: Identifying the global core of public prejudice associated with mental illness. *American journal of public health*, *103*(5), 853–860.

Portection, E. C., & Operations, H. A. (2022). *Echo factsheet - jordan (19/06/2022)*. Retrieved January 5, 2024, from https://reliefweb.int/report/jordan/echo-factsheet-jordan-19062022

Rahman, A., Khan, M. N., Hamdani, S. U., Chiumento, A., Akhtar, P., Nazir, H., Nisar, A., Masood, A., Din, I. U., Khan, N. A., et al. (2019). Effectiveness of a brief group psychological intervention for women in a post-conflict setting in pakistan: A single-blind, cluster, randomised controlled trial. *The Lancet*, *393*(10182), 1733–1744.

Raisaro, C. (2023). Incentives justifying nonconformity: Experimental evidence from motortaxi organizations in uganda. *Working Paper*.

Ridley, M. (2022). *Essays on the economics of mental illness and belief formation* [Doctoral dissertation, Massachusetts Institute of Technology].

Roth, C., Schwardmann, P., & Tripodi, E. (2024a). Depression stigma.

Roth, C., Schwardmann, P., & Tripodi, E. (2024b). Misperceived effectiveness and the demand for psychotherapy. *Journal of Public Economics*, *240*, 105254.

Small, M. L. (2017). *Someone to talk to*. Oxford University Press.

Stillman, S., Rozo, S. V., Tamim, A., Palmer, I. B., Smith, E., & Miguel, E. (2022). The syrian refugee life study: First glance. *Oxford Review of Economic Policy*, *38*(3), 625–653.

Stuber, J., & Schlesinger, M. (2006). Sources of stigma for means-tested government programs. *Social science & medicine*, *63*(4), 933–945.

Tamim, A., Smith, E. C., Palmer, B., Miguel, E., Leone, S., Rozo, S. V., & Stillman, S. (2025). *Housing subsidies for refugees: Experimental evidence on life outcomes and social integration in jordan* (tech. rep.). National Bureau of Economic Research.

UNHCR. (2024). *Syrian refugee regional response*. Retrieved January 5, 2024, from https://data2.unhcr.org/en/situations/syria/location/36

Walsha, M. (2024). Do social image concerns affect technology diffusion? evidence from mobile banking in pakistan.

Webb, D. (2024). Silence to solidarity: How communication about a minority affects discrimination. *Unpublished manuscript*.

WHO, W. M. H. S. C., et al. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *Jama*, *291*(21), 2581–2590.

Yang, D., Allen Iv, J., Mahumane, A., Riddell Iv, J., & Yu, H. (2023). Knowledge, stigma, and hiv testing: An analysis of a widespread hiv/aids program. *Journal of Development Economics*, *160*, 102958.

Yu, H. (2023). Social stigma as a barrier to hiv testing: Evidence from a randomized experiment in mozambique. *Journal of Development Economics*, *161*, 103035.

Zagic, D., Wuthrich, V. M., Rapee, R. M., & Wolters, N. (2022). Interventions to improve social connections: A systematic review and meta-analysis. *Social Psychiatry and Psychiatric Epidemiology*, *57*(5), 885–906.

8 Tables and Figures

Table 1: Sender ability to target within friend group

	Recipient Depressed at Baseline		Recipient has Anxiety at Baseline	
	No Covariates	With Covariates	No Covariates	With Covariates
	(1)	(2)	(3)	(4)
Highest need recipient in friend group	0.107***	0.094***	0.051*	0.042
	(0.029)	(0.028)	(0.028)	(0.028)
Control Mean	0.436	0.436	0.375	0.375
Covariates	No controls	Demographics Selected by lasso	No controls	Demographics Selected by lasso
N	1330	1330	1330	1330

This table shows the association between senders' indication that a friend (recipient) is or is not the most in need of mental health services, and that recipient's baseline propensity to be depressed or have anxiety. Observations are at the recipient level. The sample is restricted to instances when the sender has more than 1 friend and includes only the recipients that were reached for the baseline survey. The independent variable is a binary variable of the sender having indicated that the recipient would benefit the most from mental health information. The dependent variable in columns 1 and 2 is an indicator for whether the recipient's PHQ-9 score at baseline indicates that the recipient likely has moderate to severe depression (10 or higher). The dependent variable in columns 3 and 4 is an indicator variable for whether the recipient's GAD-2 score at baseline indicates that the respondent likely has anxiety (score 3 or higher). It should be noted that the GAD-2 is only a 2-question screening and thus more imprecise than the depression measure. Columns 2 and 4 includes recipient demographic controls that are selected using the lasso double selection procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Sender Compliance

	(1) Received campaign (recipient-level)	(2) Any clicks (0/1) (sender-level)
Treatment (sender asked to share)	0.210***	0.161***
	(0.020)	(0.015)
Control Mean	0.063	–
Covariates	Lasso Double Selection	Lasso Double Selection
N	2668	849

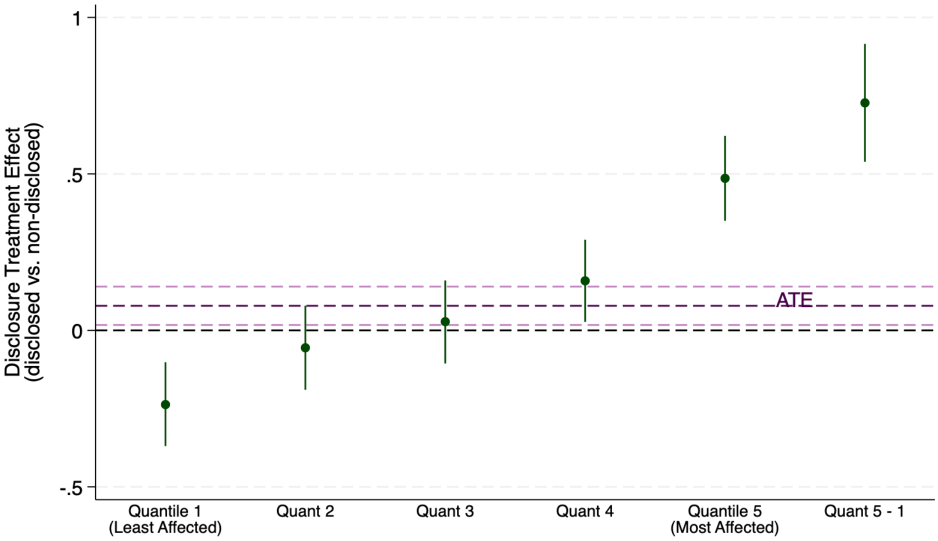
This table shows the rate at which treated senders participated by sending messages to recipients. The dependent variable in column 1 is whether the given recipient received a message from the sender. The dependent variable in column 2 is whether the sender's links were ever clicked on. A recipient is recorded to have received a message if they report this in the midline or endline survey, or if their name shows as the message recipient in a screenshot shared by their sender. Standard errors clustered at the sender level. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Effect of Message Framing on Sender Sharing

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Any clicks (0/1) (sender-level)	(4) Any clicks (0/1) (sender-level)
Disclosed Compensation framing, *non-targeted*	0.230***		0.191***	
	(0.029)		(0.028)	
Disclosed Compensation framing, *targeted*	0.223***		0.172***	
	(0.030)		(0.027)	
Disclosed Compensation framing, *pooled*		0.240***		0.188***
		(0.023)		(0.020)
Non-Disclosed Compensation framing, *non-targeted*	0.158***	0.163***	0.121***	0.123***
	(0.028)	(0.027)	(0.025)	(0.025)
p-values				
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.047]		[.048]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.071]		[.138]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.851]		[.618]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.013]		[.029]
Control Mean	0.063	0.063	–	–
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2660	2660	849	849

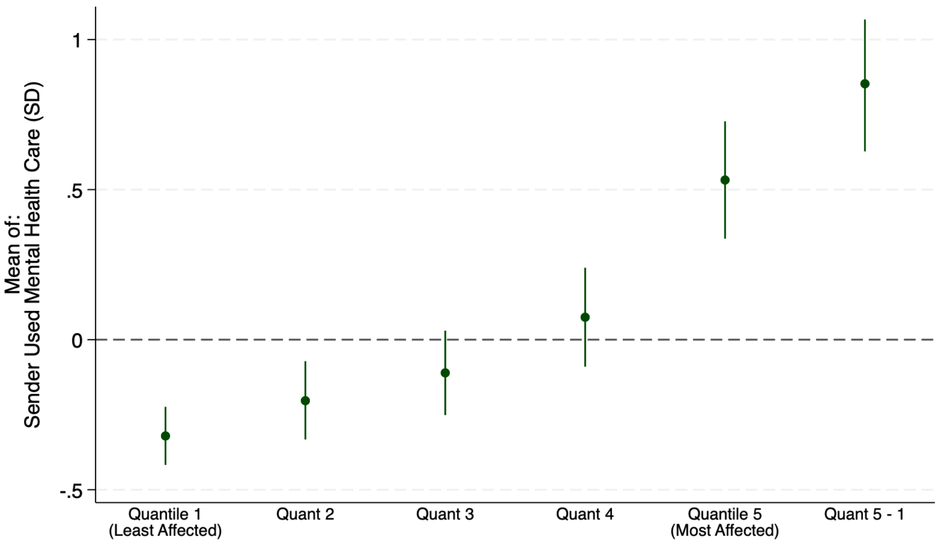
This table shows the rates of sending associated with assignment to each of the framing arms within treatment, relative to the control group which never received the campaign to share. Note that selected control covariates can vary by column. The pooled disclosed compensation framing group comprises the "disclosed compensation, non-targeted" and "disclosed compensation, targeted" groups, which were "An NGO is compensating me to share this *with all of my close friends /friends who I think can benefit from the information.* The non-disclosed compensation framing was always non-targeted, and was "I want to try to share this with all of my close friend." The framing arm coefficients are not additive. P-values are reported in brackets for the differences in point estimates. The dependent variable in columns 1 and 2 is an indicator for whether the recipient received a message from the sender. A recipient is recorded to have received a message if they report this in the midline or endline survey, or if their name shows as the message recipient in a screenshot shared by their sender. The dependent variable in columns 3 and 4 is an indicator for whether there were any clicks to links that were included in the senders' content to the recipients. These clicks may have been by anyone. In the appendix I restrict the variable to take 1 only for instances of more than 1 click from different devices and find a similar pattern of results. Standard errors of the recipient-level analysis in columns 1 and 2 are clustered at the sender level, and standard errors for the sender-level analysis in columns 3 and 4 and are robust to heteroskedasticity. Covariates are selected using the lasso double-selection procedure from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2: Sorted Group Average Treatment Effects (Effect of Disclosure)



This figure shows the group average treatment effects of disclosure, from the least affect quantile to the most affected quantile, obtained from the machine learning heterogeneity analysis following Chernozhukov et al., 2018's "generic ML" procedure. The figure shows that there is statistically significant heterogeneity in the effect of the disclosure treatment, because the difference between the first and fifth quantiles is strongly statistically significant. The purple lines depicts the estimate and confidence interval for the average treatment effect of the disclosure treatment, pooled across quantiles.

Figure 3: Mean of "Sender Used Mental Health Care" by Quantile of Impact of Disclosure



This figure shows that the quantile most affected by the disclosure treatment was much more likely to have used mental health services before compared to the least affected quantile. The figure plots the average value of the standardized variable of whether the the sender has ever used mental health services, for each quantile of the sender distribution. Out of all covariates tested, this variable is by far the most strongly associated with whether the sender is in the top or bottom quantiles of the group average treatment effects of disclosure (shown in Figure 2 above).

Table 4: Message Framing Heterogeneity by Sender Use of Mental Health Services

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Received campaign (recipient-level)
Treatment (sender asked to share)	0.173***		
	(0.021)		
Sender has used mental health services X Treatment (sender asked to share)	0.204***		
	(0.056)		
Sender has used mental health services X Disclosed Compensation framing, *non-targeted*		0.314***	
		(0.077)	
X Disclosed Compensation framing, *targeted*		0.143**	
		(0.070)	
X Non-Disclosed Compensation framing, *non-targeted*		-0.015	-0.002
		(0.080)	(0.078)
Sender used mental health services X Compensation framing, *pooled*			0.246***
			(0.061)
Disclosed Compensation framing, *non-targeted*		0.185***	
		(0.027)	
Disclosed Compensation framing, *targeted*		0.197***	
		(0.032)	
Non-Disclosed Compensation framing, *non-targeted*		0.179***	0.182***
		(0.032)	(0.032)
Disclosed Compensation framing, *pooled*			0.198***
			(0.024)
Used mental health services previously	-0.013	0.013	0.001
	(0.033)	(0.035)	(0.034)
p-values for differences of means for senders who have not used mental health services			
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$		[.88]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$		[.657]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$		[.745]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$			[.645]
p-value for senders who used services			
$\text{Disclosed}_{targeted} - Disclosed_{non-targeted}$		[.081]	
Control Mean	0.063	0.063	0.063
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2662	2662	2662

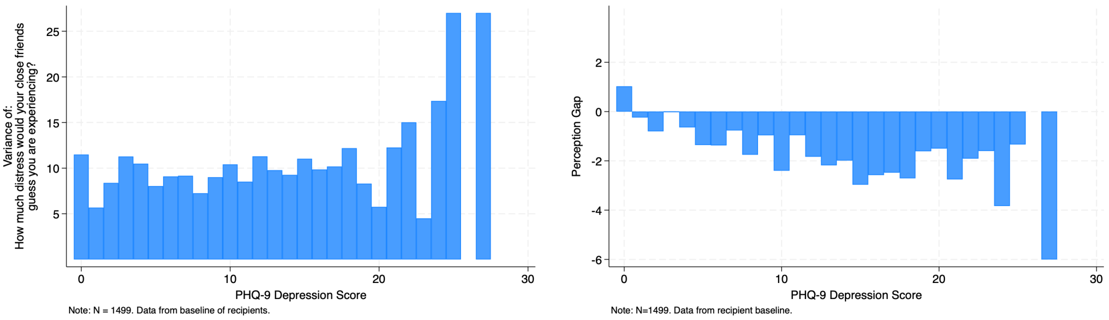
This table shows the interaction of whether the sender has ever used mental health services with assignment to treatment. The first column shows the interaction with assignment to the pooled treatment. Column 2 shows the interaction with each of the 3 framings, and column 3 shows the interaction when pooling the compensation framings. The dependent variable in column 1 is an indicator for whether the sender of the recipient sent any message to anyone in the friend group. See Appendix Table D.9 for the corresponding analysis using link clicks as the outcome. A sender is recorded to have sent any message if the sender shared a screenshot with the study documenting having shared the message, or any of the sender's recipients said in the midline survey or the endline survey that they received messages. Robust standard errors clustered at the sender level. P-values for the difference in means are reported in brackets in the bottom panel. The sample includes all recipients in the experiment. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Efficacy Beliefs, Homophily, and Targeting

	(1)	(2)	(3)	(4)
Panel A. All Baseline Potential Senders	Sender Believes Mental Health Care is Effective at Baseline (0/1)			
Sender has used mental health services (ever)	0.043*		0.043*	
	(0.025)		(0.026)	
Sender used mental health services (in past 30 days)		0.093***		0.086***
		(0.027)		(0.027)
Reference group mean	0.864	0.865	0.864	0.865
Demographic controls	No	No	Yes	Yes
N	1044	1045	1038	1039
Panel B. All Baseline Potential Recipients	Recipient Likely has Depression or Anxiety at Baseline (0/1)			
Sender has used mental health services (ever)	0.080**		0.091***	
	(0.035)		(0.035)	
Sender used mental health services (in past 30 days)		0.112**		0.115**
		(0.056)		(0.058)
Reference group mean	0.528	0.528	0.528	0.528
Demographic controls	No	No	Yes	Yes
N	1425	1427	1425	1427
Panel C. Recipients Who Received the Campaign	Recipient Likely has Depression or Anxiety at Baseline (0/1)			
Sender has used mental health services (ever)	0.117*		0.137**	
	(0.060)		(0.060)	
Sender used mental health services (in past 30 days)		0.162*		0.151*
		(0.089)		(0.085)
Reference group mean	0.510	0.510	0.510	0.510
Demographic controls	No	No	Yes	Yes
N	346	346	346	346

This table shows, first, descriptive baseline data on senders' beliefs about treatment efficacy and connections to recipients who likely have depression or anxiety, and the need level of marginal recipients reached by senders in the campaign, broken out by whether the sender used mental health services. The dependent variable in Panel A takes 1 if the respondent indicated that, if a friend were experiencing psychological distress, mental health services would help the friend be happier, healthier, and able to meet their goals, and 0 otherwise. The Panel A sample is all potential senders surveyed at baseline. Sample size difference across columns are due to question-specific non-response. The dependent variable in Panel B takes 1 if the recipient likely had depression or anxiety based on the PHQ-9 depression screening or the GAD-2 anxiety screening, and 0 otherwise. The Panel B sample is all potential recipients surveyed at baseline, matched to the senders of their friend group. Panel C restricts to the recipients who received the campaign, showing that those reached by prior-user senders are significantly more likely to have had depression or anxiety at baseline. (Note that the recipient-level campaign receipt variable is coded as 1 only when the recipient reported that they received the campaign or the recipient was identifiable from the sender's screenshots. As detailed in the data section, this can differ from the main sender-level variable, which takes 1 if the sender shared with any recipient in their network.) Columns 1 and 2 present results from univariate regressions with no control covariates, and columns 3 and 4 includes demographic controls (gender, age, education, employment status, and, in Panel A on efficacy beliefs, social desirability score.). Standard errors are robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4: Heterogeneous Priors by Recipient Mental Health Need



These figures examine respondents' priors about others' perceptions of their mental health need, by baseline level of distress. The first plot graphs the variance of respondents' priors of how much distress their friends think they (the respondent) are in. The friends' perception was elicited by asking: "*Think of your friends who you are close with and talk to often. On a scale of 1 to 10, how much psychological distress would they guess you've felt on average over the past two weeks? As a reminder, 1 means they think you are in no distress at all, while 10 means they think you are in severe distress almost every day.*" The variance is calculated across respondents within each 1-unit bin of the PHQ-9 depression score. The second plot graphs the average gap between the respondent's own assessment of their distress ((from 1 to 10) and their friends' perception. Negative values of the perception gap therefore mean the respondent believes her friends think she is doing better than she is. These outcomes were collected only among the potential recipient sample.

Table 6: Main Estimated Impacts on Recipients

	(1) Any conversations about mental health	(2) Social Connectedness Index (SD)	(3) Labor Assistance Freq. (SD)
Panel A: Pooled IV Estimates			
Sender shared (to anyone)	0.159**	0.372**	0.448**
	(0.063)	(0.186)	(0.202)
Control Mean	0.110	0.000	-0.010
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	1038	1042	1042
Panel B: ITT Estimates by Message Framing Arm			
Disclosed Compensation framing, *non-targeted*	0.050*	0.107	0.092
	(0.030)	(0.091)	(0.095)
Disclosed Compensation framing, *targeted*	0.070**	0.205**	0.201**
	(0.034)	(0.092)	(0.102)
Non-Disclosed Compensation framing, *non-targeted*	0.041	0.075	0.085
	(0.031)	(0.092)	(0.098)
Control Mean	0.11	0.00	-0.01
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	1038	1042	1042

Notes: This table contains the recipient-level significant impacts that survive a multiple hypothesis testing adjustment. (See appendix for other outcomes and FDR q-values.) Panel A presents the IV estimates pooling the three treatment arms and comparing to control. Panel B presents the ITT estimates for each message framing arm. (IV estimates are not provided by message framing due to experimentally induced differential compliance by arm, per the sender-level results.) The dependent variable in column 1 is an indicator of whether the respondent had conversations about mental health with anyone outside her household in the past 6 months. The dependent variable in column 2 is a standardized index consisting of times that the respondent has socialized with, spoken on the phone with, or helped or been helped by someone in his/her network. The dependent variable in column 3 is a standardized measure of the frequency that the respondent has taken time to help someone outside his/her household with tasks such as childcare, accompanying someone to an appointment, etc., or been helped in similar ways. The sample includes only recipients in the subsample reached to be surveyed at endline. Robust standard errors clustered at the sender level. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimated Impacts on Recipient Take-Up of Mental Health Services

	(1) Called Helpline	(2) Willing to accept call from helpline (after 6 months)	(3) Ever used mental health services (after 6 months)
Panel A: Pooled IV Estimates			
Sender shared (to anyone)	-0.017	-0.045	0.118
	(0.021)	(0.126)	(0.096)
Control Mean	0.020	0.540	0.230
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	1041	812	905
Panel B: ITT Estimates by Message Framing Arm			
Disclosed Compensation framing, *non-targeted*	-0.009	0.036	0.072*
	(0.009)	(0.050)	(0.041)
Disclosed Compensation framing, *targeted*	-0.004	-0.078	0.062
	(0.010)	(0.049)	(0.044)
Non-Disclosed Compensation framing, *non-targeted*	-0.008	0.002	0.042
	(0.009)	(0.051)	(0.044)
Control Mean	0.02	0.54	0.23
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	1041	812	905

Called helpline (column 1) measured at endline roughly 3 weeks after the intervention. Willing to accept call from helpline (column 2) is measured 6 months after implementation among only original female recipients (due to helpline programmatic priorities). Ever used mental health services is an indicator taking 1 if the respondent reports in the endline or 6 month follow-up that anyone in the household has ever used mental health services. In column 3 the "disclosed compensation, targeted" arm, which has no significant treatment effect, exhibited differential attrition; see appendix Tables D.15 and D.16. All specifications include covariates selected by lasso and standard errors are clustered at the sender level. No results in this table survive a multiple hypothesis testing adjustment; see appendix. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A. Intervention Content

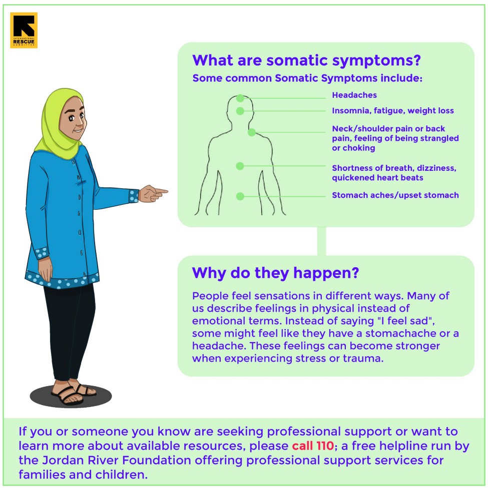


Figure A.1: Example of Campaign Content

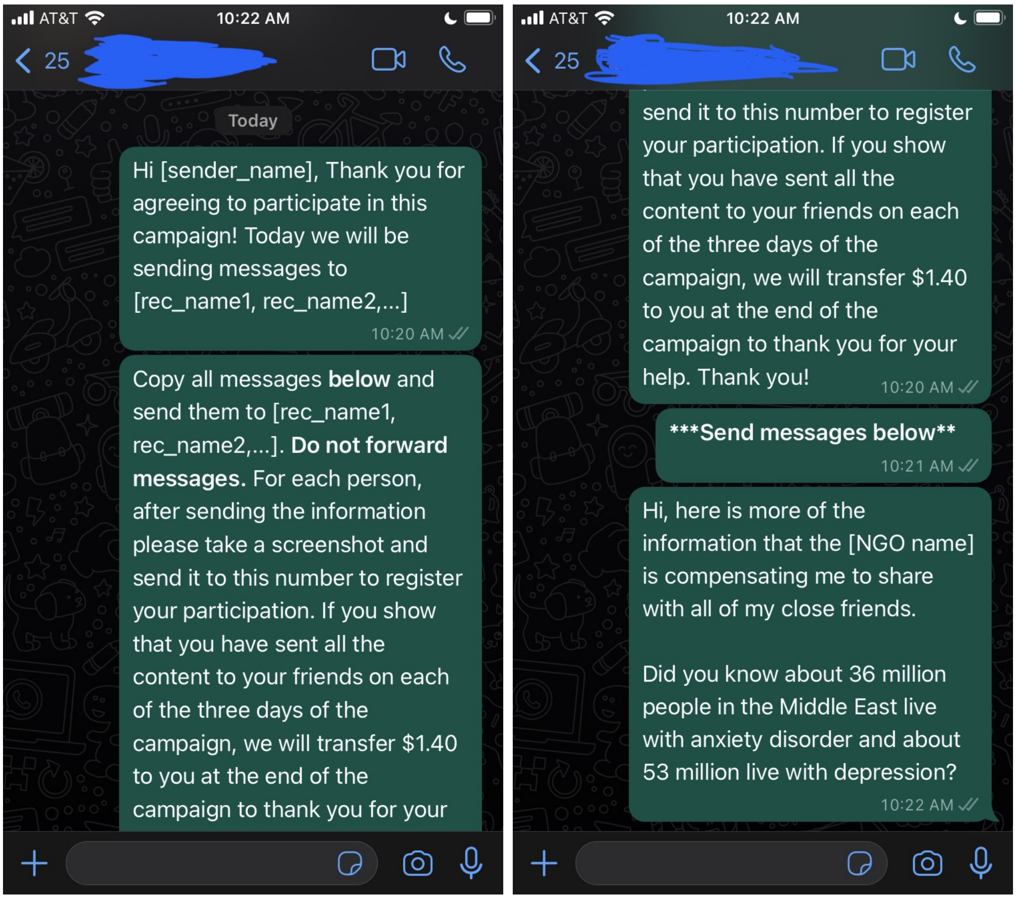


Figure A.2: Example of Campaign Instructions

Appendix B. Sender Elicitation Scripts

- **Well-known or well-regarded**: Think of the people who you know in your community, or the network of people you interact with. From among those people, tell me the name and phone number of one or two people who you know of in your community who are well-known and thought of highly. This could be because their opinions are respected, or simply because they are well-liked.
- **Community-minded**: Now, please tell me the name and phone number of one or two people you know who you believe are community-minded. This could be because they volunteer in an organized way, or they're simply very helpful to others.
- **Good at spreading information**: Now tell me the names and phone numbers of one or two people who, when they share information, many people get to know about it. For example, if they share information about job opportunities, news about Syria, or a wedding, many people would learn about it.
- **Random sample**: Identified through random digit dialing

Appendix C.

Assumption 1. Mental health need (or "vulnerability") V and treatment efficacy beliefs Q are continuously distributed in the population, together yielding an expected benefit $h(V, Q)$ from mental health services, with h increasing in V and Q.

Let prior mental health care users be people who at some point, potentially in the past, faced $h(V_s, Q_s) > \kappa$ where κ is the financial, logistical, and social costs of using services. As such the likelihood of using services is increasing in both need and efficacy beliefs.

A fraction π of the potential sender population are prior users of mental health services (type U) and the remainder are non-users (type N).

Senders choose a binary sending action $S \in 0, 1$ to maximize their utility which is increasing in the benefit to the recipient and decreasing in the social image loss from sending.

$$U(S) = \underbrace{h(V_r, Q_s)\mathbf{1}_{s=1}}_{\substack{\text{Health benefit} \\ \text{to recipient} \\ \text{(sender's belief)}}} - \underbrace{R(S)}_{\substack{\text{social image cost} \\ \text{to sender}}} - C\mathbf{1}_{s=1} \tag{14}$$

where $h(V_r, Q_s)$ is the sender's expectation of the recipient's health benefit given the recipient's vulnerability V_r and the sender's efficacy beliefs Q_s. $R(S)$ is the image cost of sending, given by the change in the posterior belief that the sender is a mental health care user if the sender shares: $R(S) \equiv \pi_{post,s=1} - \pi_{post,s=0}$.

The model intuition can be extended to assume two types of recipients, the vulnerable type and the unvulnerable type who differ in their mental health need, with ω representing the proportion of recipients who are the vulnerable type, and who therefore expereience a social image cost from receiving the message, but here I consider only sender type.

Let $F_U(\cdot)$ and $F_N(\cdot)$ denote the CDFs of the expected health benefit $h(V_r, Q_s)$ for prior users and non-users respectively.

Assumption 2. Assume that prior users have stochastically higher beliefs about recipient benefit, such that $F_N(h) > F_U(h)$ for all h.

Following Chandrasekhar et al., 2018, the rate of sharing will be dictated by an equilibrium cutoff level of expected recipient health benefit, h^*. Above this cutoff, the benefit to the recipient exceeds the image cost of sending. Then $r = 1 - F_N(h^*)$ and $p = 1 - F_U(h^*)$ are the probabilities that each type's perceived benefit exceeds the equilibrium cutoff h^*. The image cost is based on the informativeness of the sending signal, summarized by the log odds ratio:

$$\frac{\mathbb{P}(user = 1|S = 1)}{\mathbb{P}(user = 0|S = 1)} = \frac{\pi}{1-\pi}\left(\frac{1 - F_U(h^*)}{1 - F_N(h^*)}\right) \tag{15}$$

Because $1 - F_U(h^*) > 1 - F_N(h^*)$, prior users send at a higher rate than non-users in equilibrium, so observing that a sender shares will always lead observers to update upward their belief that the sender is a prior user. That image cost drives up the equilibrium sending cutoff h^*.

The change in the posterior belief that the sender is a prior user is captured by $R(S)$.

$$R(S) = \pi_{post,s=1} - \pi_{post,s=0} = \frac{\pi(1-\pi)(p-r)}{r + (p-r)\pi} \tag{16}$$

The expression also shows that image sensitivity is non-monotone in π. As the prior π that someone is a prior mental health user approaches 0 or 1, the reputation cost $R(S)$ converges to 0.

8.1 Effect of an observable financial incentive

Next, an observable monetary incentive will decrease the cutoff. The sender's utility is now

$$U(S) = \underbrace{h(V_r, Q_s)\mathbf{1}_{s=1}}_{\substack{\text{Health benefit}\\ \text{to recipient}\\ \text{(sender's belief)}}} - \underbrace{R(S)}_{\substack{\text{social image cost}\\ \text{to sender}}} - C\mathbf{1}_{s=1} + \underbrace{M\mathbf{1}_{s=1}}_{\substack{\text{monetary incentive}\\ \text{if sends}}} \tag{17}$$

and the equilibrium cutoff will be

$$h^* = R(s) + C - M \tag{18}$$

Since recipients only observe M if sending happens, assume that $\frac{\partial \pi_{post,s=0}}{\partial M} = 0$. Then the change in the threshold for a change in M is given by

$$\frac{\partial h^*}{\partial M} = \frac{\partial \pi_{post,s=1}}{\partial M} - 1 \tag{19}$$

Therefore the observable incentive will decrease the threshold as long as $\frac{\partial \pi_{post,s=1}}{\partial M} < 1$. Noting that $\frac{\partial \pi_{post,s=1}}{\partial M} = \frac{\partial \pi_{post,s=1}}{\partial h^*}\frac{\partial h^*}{\partial M}$, this can be re-stated as:

$$\frac{\partial h^*}{\partial M} = \frac{-1}{1 - \frac{\partial \pi_{post,s=1}}{\partial h^*}} \tag{20}$$

This shows that as long as $\frac{\partial \pi_{post,s=1}}{\partial h^*} < 1$ the monetary incentive will decrease the threshold, and if $0 < \frac{\partial \pi_{post,s=1}}{\partial h^*} < 1$ then there will additionally be a crowd-in effect, from the image cost decreasing.

For tractability, assume the difference in efficacy beliefs between users and non-users creates a constant difference in expected health benefit, so that $F_U(h) = F_N(h - q)$ for some $q > 0$.

The posterior $\pi_{post,s=1}$ will be increasing in h^* if, for any $h < h^*$ and $q > 0$, $\frac{f(h-q)}{f(h)} < \frac{f(h^*-q)}{f(h^*)}$. This holds for normally distributed h. As h^* decreases, the proportion of prior users is smaller among the marginal senders induced to participate than among the inframarginal senders.

8.2 Homophily

Assumption 3. Assume that there is homophily in mental health need, such that the friends of a person with high mental health need are more likely to also have high need than the friends of someone without high mental health need. And, following from that, people who have used mental health services before are on average more connected to others with high mental health need than people who have not used mental health services.

Define the sending rates for type "user" and "non-user" as S_U and S_N respectively, and the fraction of the sender's friends who are high-need as f_U and f_N. Under Assumption 3 (homophily) $f_U > \bar{f} > f_N$, with $\bar{f}$ the baseline fraction without homophily. Assume connections are conserved so that $\pi\delta_U = (1-\pi)\delta_N$, where $\delta_U = f_U - \bar{f} > 0$ and $\delta_N = \bar{f} - f_N > 0$.

The expected fraction of high-need recipients who receive the message is:

$$P_{HN} = \pi \cdot S_U \cdot f_U + (1-\pi) \cdot S_N \cdot f_N \tag{21}$$

The counterfactual without homophily ($f_U = f_N = \bar{f}$) is:

$$P^0_{HN} = \bar{f} \cdot [\pi S_U + (1-\pi) S_N] \tag{22}$$

The effect of homophily given a fixed type-specific sending rate is therefore:

$$\Delta = P_{HN} - P^0_{HN} = \pi S_U \delta_U - (1-\pi) S_N \delta_N = \delta_U \pi (S_U - S_N) \tag{23}$$

where the last equality uses the conservation assumption. The sign of Δ depends on whether S_U is greater than or less than S_N.

Case 1 (image sensitivity not correlated with need, $\pi_U \approx \pi_N$): Prior users send more than non-users due to their efficacy advantage ($Q_U(V) > Q_N(V)$), so $S_U > S_N$ and $\Delta > 0$. Homophily increases messaging to high-need recipients.

Case 2 (image sensitivity positively correlated with need, $\pi_U > \pi_N$): Higher image costs for prior users reduce S_U. If the image cost effect dominates the efficacy advantage, $S_U < S_N$ and $\Delta < 0$. Homophily then reduces messaging to high-need recipients relative to the no-homophily counterfactual.

Appendix D. Supplementary Tables and Figures

List of Tables

List of Figures

Figure D.1: Prevalence of Depression among Recipients at Baseline

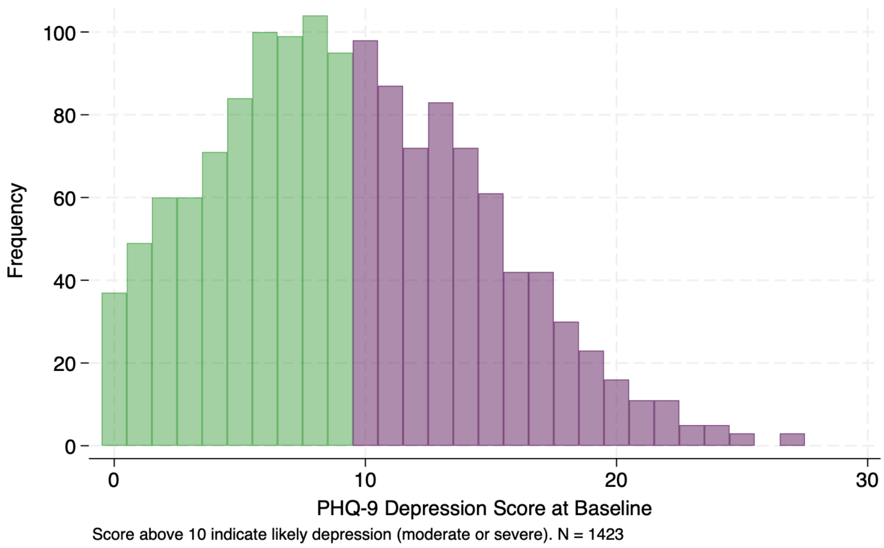


Figure D.2: Prevalence of Likely Anxiety Among Recipients at Baseline

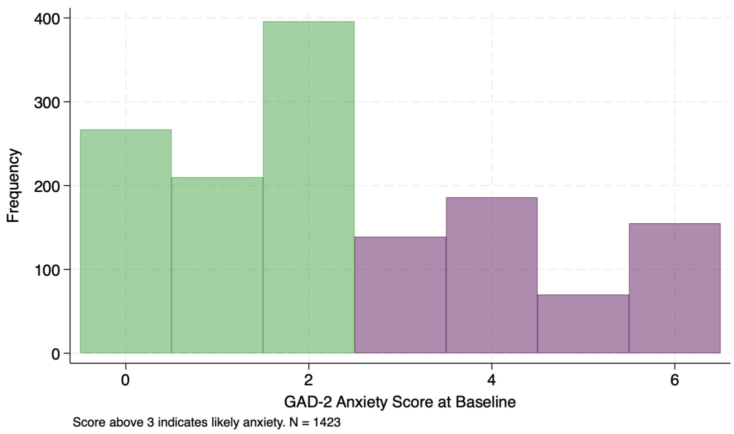


Table D.1: Robustness: Recipient only considered to receive message if shared the name of who they got the message from

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)
Disclosed Compensation framing, *non-targeted*	0.217***	
	(0.025)	
Disclosed Compensation framing, *targeted*	0.219***	
	(0.026)	
Disclosed Compensation framing, *pooled*		0.229***
		(0.019)
Non-Disclosed Compensation framing, *non-targeted*	0.163***	0.166***
	(0.024)	(0.024)
p-values		
$\text{Disclosed}_{non-targeted}$ − $\text{Non-Disclosed}_{non-targeted}$	[.12]	
$\text{Disclosed}_{targeted}$ − $\text{Non-Disclosed}_{non-targeted}$	[.099]	
$\text{Disclosed}_{targeted}$ − $\text{Disclosed}_{non-targeted}$	[.828]	
$\text{Disclosed}_{pooled}$ − $\text{Non-Disclosed}_{non-targeted}$		[.033]
Control Mean	0.012	0.012
Covariates	Lasso Double Selection	Lasso Double Selection
N	2668	2668

This table shows the main treatment effects when restricting the outcome variable "received campaign (recipient-level)" to 0 if the recipient does not share the name of who they received the message from. This is a more conservative measure of whether the recipient really got the campaign, and was not pre-specified. The table shows the rates of sending associated with assignment to treatment and assignment to each of the framing arms within treatment, relative to the control group which never received the campaign to share. The framing arms are mutually exclusive and together comprise the complete treatment group, therefore coefficients on the framing arms are not additive. P-values are reported for the differences in point estimates. The dependent variable is an indicator for whether the sender of the recipient sent any message to anyone in the friend group, taking 1 if the recipient says they received the campaign and gives the name of the person who shared the campaign with them. The dependent variable also takes 1 if the recipient is identifiable from screenshots shared by the sender documenting their participation. Robust standard errors clustered at the sender level and reported in parentheses. Covariates are selected using the lasso double-selection procedure from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: Pre-Specified Specification for Main Analysis

	(1) Received campaign (recipient-level)	(2) Any clicks (0/1) (sender-level)
Sender's image concerns alleviated	1.813***	7.041***
	(0.238)	(2.055)
Recipient's image concerns alleviated	1.387***	6.538***
	(0.245)	(2.045)
Both sender's and recipients image concerns alleviated	-1.329***	-6.292***
	(0.304)	(2.048)
Estimated effect of sender image concerns $(\alpha_1 + \alpha_2 + \delta) - \alpha_1$	0.484	0.749
p-value	[.013]	[.012]
Estimated effect of recipient image concerns $(\alpha_1 + \alpha_2 + \delta) - \alpha_2$	0.058	0.246
p-value	[.757]	[.342]
Control Mean	0.063	–
Covariates	Lasso Double Selection	Lasso Double Selection
N	2660	848

This table presents the main results using the pre-specified specification. The results are very similar in magnitude and significance, but are less easily digestible. The specification is:

$$log\left(\frac{p_r}{1-p_r}\right) = \alpha_0 + \alpha_1 I_s^S + \alpha_2 I_s^R + \delta I_s^S I_s^R + X_s'\beta_1 + X_r'\beta_2 + \Gamma + \varepsilon_r \tag{24}$$

where I_s^S is an indicator for the sender being assigned to a framing that alleviates the "sender social image concern" (ie. m_1 or m_2 framing), and I_s^R is an indicator for the sender being assigned a framing that alleviates the "recipient social image concern" (ie. m_1 or m_3), and the interaction of the two takes one when both concerns are alleviated (corresponding to m_1 only). Recall that the regression coefficients are rates of sending relative to a pure control group. The reference group for calculating treatment effects is when both concerns are alleviated, ie. $I^S = 1$ and $I^R = 1$, captured by the sum of the three regression coefficients $\alpha_1 + \alpha_2 + \delta$. The estimated effect of alleviating the sender image concern is given by the difference between the rate when both concerns are alleviated versus only recipient image concerns are alleviated: $(\alpha_1 + \alpha_2 + \delta) - \alpha_2$. The estimated effect of alleviating the recipient image concern is given by the difference between the rate when both concerns are alleviated versus only sender image concerns are alleviated: $(\alpha_1 + \alpha_2 + \delta) - \alpha_1$. As in all the analysis, X_s is a vector of sender-level covariates, X_r is a vector of recipient baseline covariates, Γ are week of survey fixed effects, and covariates are selected from the list below using the double post lasso method following Belloni et al. 2014. Differences in rates of sharing between the conditions are reported in the second panel, and p-values are reported in brackets below the associated difference in point estimates. The dependent variable in column 1 is an indicator for whether the sender of the recipient sent any message to anyone in the friend group. In column 2 the dependent variable is a sender-level indicator for whether anyone clicked on any of the sender's links. Robust standard errors clustered at the sender level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.3: Robustness: Sender follow-through with unique click data

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Any clicks (0/1) (sender-level)	(4) Any clicks (0/1) (sender-level)	(5) More than 1 unique click (0/1)	(6) More than 1 unique click (0/1)
Disclosed Compensation framing, *non-targeted*	0.230***		0.191***		0.088***	
	(0.029)		(0.028)		(0.020)	
Disclosed Compensation framing, *targeted*	0.223***		0.172***		0.065***	
	(0.030)		(0.027)		(0.018)	
Non-Disclosed Compensation framing, *non-targeted*	0.158***	0.163***	0.121***	0.123***	0.040**	0.045***
	(0.028)	(0.027)	(0.025)	(0.025)	(0.016)	(0.016)
Disclosed Compensation framing, *pooled*		0.240***		0.188***		0.085***
		(0.023)		(0.020)		(0.014)
p-values						
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.047]		[.048]		[.057]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.071]		[.138]		[.29]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.851]		[.618]		[.41]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.013]		[.029]		[.051]
Control Mean	0.063	0.063	–	–	–	–
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2660	2660	849	849	849	849

This table shows the rates of sending associated with assignment to treatment and assignment to each of the framing arms within treatment, relative to the control group which never received the campaign to share. The framing arms are mutually exclusive and together comprise the complete treatment group, therefore coefficients on the framing arms are not additive. Differences in rates of sharing between framing arms are reported in the second panel. P-values are reported in brackets below the associated difference in point estimates. The dependent variable in columns 1-3 is an indicator for whether the sender of the recipient sent any message to anyone in the friend group. A sender is recorded to have sent any message if the sender shared a screenshot with the study documenting having shared the message, or any of the sender's recipients said in the midline survey (the week after the campaign) that they received messages, or any of the sender's recipients said in the endline survey that they received a campaign message. The dependent variable in columns 4 and 5 is an indicator for whether there were any clicks to links that were included in the senders' content to the recipients. These clicks may have been by anyone. The dependent variable in columns 6 and 7 is an indicator for more than 1 click by different devices. In the appendix I restrict the variable to take 1 only for instances of more than 1 click from different devices and find a similar pattern of results. The last comparison in the second panel comes from running the same specification except that framings 1 and 2 are pooled together. That specification is reported in the appendix but not here to avoid encouraging over-interpretation of the comparison. Robust standard errors clustered at the sender level and reported in parentheses. Covariates are selected using the lasso double-selection procedure from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.4: Robustness: Sender follow-through excluding recipients with duplicate treatments

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)
Disclosed Compensation framing, *non-targeted*	0.245***	
	(0.031)	
Disclosed Compensation framing, *targeted*	0.227***	
	(0.030)	
Disclosed Compensation framing, *pooled*		0.237***
		(0.023)
Non-Disclosed Compensation framing, *non-targeted*	0.144***	0.144***
	(0.028)	(0.028)
p-values		
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.008]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.02]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.812]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.004]
Control Mean	0.063	0.060
Covariates	Lasso Double Selection	Lasso Double Selection
N	2549	2549

4.5% of recipients were linked to multiple treated senders and so may have been exposed to the treatment more than once. This table tests the primary hypothesis that sending rates vary with the framing while excluding those individuals. This robustness check is not relevant for the click rate variable, because that outcome is directly tied to the unique sender. The dependent variable in all columns of this table is an indicator for whether the recipient received the campaign. The recipient is recorded to have received the campaign if the sender shared a screenshot with the study documenting having shared a message with that person, or any of the sender's recipients said in the midline survey that they received messages, or the recipient said in the endline survey that they received messages. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.5: Robustness: Sender follow-through analysis with logistic regression

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Any clicks (0/1) (sender-level)	(4) Any clicks (0/1) (sender-level)
Disclosed Compensation framing, *non-targeted*	1.680***		0.713**	
	(0.209)		(0.289)	
Disclosed Compensation framing, *targeted*	1.641***		0.420	
	(0.221)		(0.303)	
Disclosed Compensation framing, *pooled*		1.833***		0.566**
		(0.206)		(0.264)
Non-Disclosed Compensation framing, *non-targeted*	1.290***	1.393***		
	(0.227)	(0.232)		
p-values				
$\text{Disclosed}_{non-targeted}$ − $\text{Non-Disclosed}_{non-targeted}$	[.048]			
$\text{Disclosed}_{targeted}$ − $\text{Non-Disclosed}_{non-targeted}$	[.075]			
$\text{Disclosed}_{targeted}$ − $\text{Disclosed}_{non-targeted}$	[.851]		[.256]	
$\text{Disclosed}_{pooled}$ − $\text{Non-Disclosed}_{non-targeted}$		[.012]		
Mean click rate in $\text{Non-Disclosed}_{non-targeted}$			.111	.111
Control Mean	0.063	0.063	–	–
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2660	2660	642	643

This table presents the primary analysis using logistic regressions rather than linear regressions. The dependent variable in columns 1 and 2 is an indicator for whether the recipient received the campaign, and in columns 3 and 4 it is whether the sender's link was clicked on by anyone. In columns 3 and 4 the regression is run only on the treatment group and coefficients are relative to the non-disclosed non-targeted framing, because convergence does not occur when including the pure control group. Standard errors are robust in all specifications and clustered at the sender level in columns 1 and 2 where the analysis is at the recipient level. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.6: Robustness: Sender follow-through accounting for imbalance and variation in treatment intensity

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Received campaign (recipient-level)	(4) Received campaign (recipient-level)	(5) Received campaign (recipient-level)	(6) Received campaign (recipient-level)
Disclosed Compensation framing, *non-targeted*	0.237***		0.228***		0.231***	
	(0.030)		(0.035)		(0.035)	
Disclosed Compensation framing, *targeted*	0.233***		0.178***		0.186***	
	(0.029)		(0.035)		(0.035)	
Disclosed Compensation framing, *pooled*		0.247***		0.212***		0.217***
		(0.023)		(0.027)		(0.027)
Non-Disclosed Compensation framing, *non-targeted*	0.174***	0.180***	0.138***	0.142***	0.157***	0.161***
	(0.028)	(0.028)	(0.030)	(0.030)	(0.031)	(0.031)
p-values						
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.083]		[.031]		[.077]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.097]		[.324]		[.48]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.916]		[.277]		[.311]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.028]		[.044]		[.109]
Sample	All Observations	All Observations	Dropping entire affected week	Dropping entire affected week	Dropping entire affected week	Dropping entire affected week
Covariates	Imbalance Covariates And Lasso Double Selection	Imbalance Covariates And Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Imbalance Covariates And Lasso Double Selection	Imbalance Covariates And Lasso Double Selection
N	2660	2660	1764	1764	1764	1764

This table tests alternate ways to address imbalance and an implementation glitch that caused variation in the treatment intensity, by causing a random subset of senders in Framing 3 to not receive one of the three batches of campaign content. In column 1 the specification forces the inclusion of imbalanced covariates regardless of whether they were selected by lasso, and also controls for whether the sender experienced the implementation glitch. In column 2 the entire week affected by the implementation glitch is dropped, which leads to a large loss in power. In column 3 the affected week is dropped and imbalanced baseline covariates are forced to be included regardless of whether lasso selected them. Other covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. Robust standard errors are clustered at the sender level. The dependent variable in all columns of this table is an indicator for whether the recipient received the campaign. The results show that the primary results are robust to accounting for these concerns. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.7: Robustness: Sender follow-through measured at the sender level (extensive margin)

	(1) Sender shared (to anyone)	(2) Sender shared (to anyone)
Disclosed Compensation framing, *non-targeted*	0.331***	
	(0.041)	
Disclosed Compensation framing, *targeted*	0.384***	
	(0.042)	
Disclosed Compensation framing, *pooled*		0.380***
		(0.034)
Non-Disclosed Compensation framing, *non-targeted*	0.290***	0.302***
	(0.043)	(0.043)
p-values		
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.411]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.063]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.282]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.073]
Control Mean	0.058	0.058
Covariates	Lasso Double Selection	Lasso Double Selection
N	849	849

This table replicates the primary result at the sender level rather than the recipient level. The dependent variable takes 1 if the sender was recorded to have shared the content with anyone, and 0 otherwise. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. Standard errors robust to heteroskedasticity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.8: Heterogeneity in Disclosure Effect by Pre-Registered Sender Characteristics

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Received campaign (recipient-level)	(4) Received campaign (recipient-level)	(5) Received campaign (recipient-level)
Disclosed Compensation framing, *pooled*	0.240***	0.239***	0.251***	0.224***	0.240***
	(0.023)	(0.023)	(0.024)	(0.028)	(0.022)
Non-Disclosed Compensation framing, *non-targeted*	0.166***	0.166***	0.163***	0.181***	0.168***
	(0.028)	(0.027)	(0.030)	(0.038)	(0.028)
Sender treatment efficacy beliefs					
X Compensation framing, *pooled*	0.045*				
	(0.024)				
X Non-Disclosed Compensation framing, *non-targeted*	0.017				
	(0.026)				
Mental health service efficacy beliefs (SD)	-0.015				
	(0.019)				
Sender own stigma views above median					
X Compensation framing, *pooled*		-0.035			
		(0.021)			
X Non-Disclosed Compensation framing, *non-targeted*		0.022			
		(0.025)			
Own stigma views (1st order) (SD)		0.014			
		(0.020)			
Sender altruism					
X Compensation framing, *pooled*			-0.002		
			(0.025)		
X Non-Disclosed Compensation framing, *non-targeted*			0.008		
			(0.027)		
Altruism (SD)			-0.003		
			(0.012)		
Sender female					
X Compensation framing, *pooled*				0.033	
				(0.043)	
X Non-Disclosed Compensation framing, *non-targeted*				-0.039	
				(0.054)	
Female				0.047	
				(0.068)	
Sender social desirability					
X Compensation framing, *pooled*					0.027
					(0.023)
Sender social desirability					
X Compensation framing, *pooled*					0.020
					(0.028)
Social desirability score (SD)					-0.006
					(0.015)
p-value: test of heterogeneous effect of disclosure	[.301]	[.05]	[.777]	[.23]	[.801]
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2666	2660	2362	2668	2666

This table shows that the disclosure effect varies by only one of the five pre-specified dimension of sender heterogeneity – sender's own stigma toward mental health care seekers. The p-value at the bottom of the table tests whether the effect of disclosure (Disclosed – Non-disclosed) varies significantly by the given covariate. The pooled disclosed compensation framing group comprises the "disclosed compensation, non-targeted" and "disclosed compensation, targeted" groups, which were "An NGO is compensating me to share this *with all of my close friends /friends who I think can benefit from the information.* The non-disclosed compensation framing was always non-targeted, and was "I want to try to share this with all of my close friend." P-values are reported in brackets for the differences in point estimates. The dependent variable in columns 1 and 2 is an indicator for whether the recipient received a message from the sender. A recipient is recorded to have received a message if they report this in the midline or endline survey, or if their name shows as the message recipient in a screenshot shared by their sender. Standard errors clustered at the sender level and reported in parentheses. Covariates are selected using the lasso double-selection procedure from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.9: Message Framing Heterogeneity by Sender Use of Mental Health Services - Link Clicks

	(1) Any clicks (0/1) (sender-level)	(2) Any clicks (0/1) (sender-level)	(3) Any clicks (0/1) (sender-level)
Treatment (sender asked to share)	0.138***		
	(0.015)		
Sender has used mental health services X Treatment (sender asked to share)	0.146***		
	(0.048)		
Sender has used mental health services X Disclosed Compensation framing, *non-targeted*		0.185**	
		(0.092)	
X Disclosed Compensation framing, *targeted*		0.153*	
		(0.081)	
X Non-Disclosed Compensation framing, *non-targeted*		0.041	0.059
		(0.071)	(0.073)
Sender used mental health services X Compensation framing, *pooled*			0.208***
			(0.064)
Disclosed Compensation framing, *non-targeted*		0.172***	
		(0.028)	
Disclosed Compensation framing, *targeted*		0.136***	
		(0.027)	
Non-Disclosed Compensation framing, *non-targeted*		0.112***	0.108***
		(0.026)	(0.025)
Disclosed Compensation framing, *pooled*			0.148***
			(0.020)
Used mental health services previously	-0.006	-0.005	-0.013
	(0.007)	(0.025)	(0.015)
p-values for differences of means for senders who have not used mental health services			
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$		[.104]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$		[.511]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$		[.341]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$			[.194]
Control Mean	0.000	0.000	0.000
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	848	848	848

This table shows the interaction of whether the sender has ever used mental health services with assignment to treatment. The first column shows the interaction with assignment to the pooled treatment. Column 2 shows the interaction with each of the 3 framings, and column 3 shows the interaction when pooling the compensation framings. The dependent variable takes 1 if any of the sender's links were clicked on. Standard errors are robust to heteroskedasticity. P-values for the difference in means are reported in brackets in the bottom panel. The sample includes all senders in the experiment. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.10: Top 10 Machine Learning Heterogeneity Selected Characteristics

Variable	Q5-Q1 (Standard deviation units)	p-value
Sender used mental health care	0.852711389	1.90106877323843e-18
Sender mental health marriage stigma	-0.785484182	8.223281768911e-18
Sender 1st order stigma beliefs	-0.742727054	7.55659757305122e-16
Sender social connectedness	0.603491127	3.16843001762537e-10
Sender network size	0.569624723	3.48187578445494e-08
Sender 2nd order stigma beliefs	-0.558151781	5.04126464440317e-09
Sender type "well known/recognized"	0.444873426	3.71349290206009e-06
Sender education = secondary	0.440483404	1.58532197690339e-05
Recipient employed	-0.405660572	1.49238807554035e-05
Sender mental health employment stigma	-0.404565685	4.11926441790133e-05
Recipient ranked need	0.392277769	0.000183327

Table D.11: Attrition by Survey Round

	(1) Recipient Surveyed Baseline	(2) Recipient Surveyed Baseline	(3) Recipient Surveyed Baseline	(4) Recipient Surveyed Endline	(5) Recipient Surveyed Endline	(6) Recipient Surveyed Endline
Treatment						
(sender asked to share)	-0.023			-0.022		
	(0.027)			(0.027)		
Disclosed Compensation framing, *non-targeted*		-0.031			-0.021	
		(0.029)			(0.028)	
Disclosed Compensation framing, *targeted*		-0.037			-0.043	
		(0.030)			(0.030)	
Non-Disclosed Compensation framing, *non-targeted*		-0.036	-0.034		-0.029	-0.028
		(0.030)	(0.031)		(0.029)	(0.030)
Disclosed Compensation framing, *pooled*			-0.031			-0.031
			(0.027)			(0.026)
F-Statistic	.76	.734	.8220000000000001	.672	.734	.723
Control Mean	0.551	0.551	0.551	0.409	0.409	0.409
Covariates	No Covariates	No Covariates	No Covariates	No Covariates	No Covariates	No Covariates
N	2668	2668	2668	2668	2668	2668

This table shows that there was no difference in the probability of being treated for recipients reached for the baseline and endline surveys. Note that attrition is not relevant for senders since they are surveyed at the time of enrollment and not again after. Standard errors are clustered at the sender level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.12: Balance, Treatment versus Control

Variable	N	(1) Total Mean/(SE)	N	(2) Control Mean/(SE)	N	(3) Treatment Mean/(SE)	N	(2)-(3) Pairwise t-test Mean difference
Used mental health services previously	848	0.173	207	0.198	641	0.165	848	0.033
		(0.013)		(0.028)		(0.015)		
Female	849	0.420	207	0.435	642	0.416	849	0.019
		(0.017)		(0.035)		(0.019)		
Age quartile	849	2.370	207	2.290	642	2.396	849	-0.106
		(0.037)		(0.075)		(0.042)		
Lives in refugee camp	849	0.153	207	0.130	642	0.160	849	-0.030
		(0.012)		(0.023)		(0.014)		
Education quartile	849	1.910	207	1.860	642	1.927	849	-0.067
		(0.042)		(0.082)		(0.048)		
Working	849	0.353	207	0.353	642	0.354	849	-0.001
		(0.016)		(0.033)		(0.019)		
Own stigma views (1st order) (SD)	846	-0.038	206	-0.190	640	0.011	846	-0.201**
		(0.034)		(0.065)		(0.040)		
2nd order stigma beliefs (SD)	826	-0.012	199	-0.035	627	-0.005	826	-0.031
		(0.034)		(0.071)		(0.039)		
Altruism (SD)	753	-0.023	184	0.002	569	-0.031	753	0.033
		(0.036)		(0.075)		(0.041)		
Social desirability score (SD)	848	-0.017	207	-0.025	641	-0.014	848	-0.012
		(0.034)		(0.063)		(0.040)		
PHQ-2 depression score (SD)	848	0.050	207	0.066	641	0.045	848	0.021
		(0.034)		(0.067)		(0.040)		
GAD-2 anxiety score (SD)	848	0.032	207	0.061	641	0.023	848	0.038
		(0.034)		(0.068)		(0.040)		
Social connectedness (SD)	848	0.041	207	0.102	641	0.021	848	0.081
		(0.034)		(0.065)		(0.040)		
Jordanian	849	0.110	207	0.101	642	0.112	849	-0.011
		(0.011)		(0.021)		(0.012)		
Mental health service efficacy beliefs (SD)	848	0.033	206	0.067	642	0.022	848	0.045
		(0.032)		(0.065)		(0.038)		
Friend group size	849	3.296	207	3.208	642	3.324	849	-0.116
		(0.059)		(0.118)		(0.068)		
F-test of joint significance (F-stat)								1.004
F-test, number of observations								723

Pair-wise regressions and F-test additionally control for governorate and randomization strata which are controls in the main analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.13: Balance by Framing Arm

Variable	N	(1) Total Mean/(SE)	N	(2) Framing: Disclosed, nontargeted Mean/(SE)	N	(3) Framing: Disclosed, Targeted Mean/(SE)	N	(4) Framing: Non-Disclosed, nontargeted Mean/(SE)	N	Pairwise t-test (2)-(3) Mean difference	N	Pairwise t-test (2)-(4) Mean difference	N	Pairwise t-test (3)-(4) Mean difference
Used mental health services previously	641	0.165	219	0.151	207	0.198	215	0.149	426	-0.047	434	0.002	422	0.049
		(0.015)		(0.024)		(0.028)		(0.024)						
Female	642	0.416	219	0.397	207	0.411	216	0.440	426	-0.013	435	-0.043	423	-0.029
		(0.019)		(0.033)		(0.034)		(0.034)						
Age quartile	642	2.396	219	2.461	207	2.367	216	2.356	426	0.094	435	0.105	423	0.011
		(0.042)		(0.071)		(0.075)		(0.072)						
Lives in refugee camp	642	0.160	219	0.155	207	0.140	216	0.185	426	0.015	435	-0.030	423	-0.045
		(0.014)		(0.025)		(0.024)		(0.026)						
Education quartile	642	1.927	219	1.963	207	1.942	216	1.875	426	0.021	435	0.088	423	0.067
		(0.048)		(0.084)		(0.085)		(0.082)						
Working	642	0.354	219	0.329	207	0.391	216	0.343	426	-0.063*	435	-0.014	423	0.049
		(0.019)		(0.032)		(0.034)		(0.032)						
Own stigma views (1st order) (SD)	640	0.011	219	-0.051	206	0.010	215	0.076	425	-0.061	434	-0.128	421	-0.066
		(0.040)		(0.068)		(0.068)		(0.072)						
2nd order stigma beliefs (SD)	627	-0.005	217	-0.054	201	-0.068	209	0.108	418	0.014	426	-0.162*	410	-0.176**
		(0.039)		(0.069)		(0.066)		(0.069)						
Altruism (SD)	569	-0.031	194	0.019	188	-0.092	187	-0.021	382	0.111	381	0.040	375	-0.071
		(0.041)		(0.068)		(0.075)		(0.071)						
Social desirability score (SD)	641	-0.014	218	-0.034	207	0.013	216	-0.019	425	-0.047	434	-0.015	423	0.033
		(0.040)		(0.067)		(0.072)		(0.068)						
PHQ-2 depression score (SD)	641	0.045	218	0.049	207	0.056	216	0.031	425	-0.007	434	0.017	423	0.024
		(0.040)		(0.067)		(0.071)		(0.069)						
GAD-2 anxiety score (SD)	641	0.023	218	0.053	207	0.049	216	-0.031	425	0.003	434	0.084	423	0.080
		(0.040)		(0.070)		(0.072)		(0.065)						
Social connectedness (SD)	641	0.021	219	-0.098	206	0.015	216	0.148	425	-0.113	435	-0.246***	422	-0.134
		(0.040)		(0.067)		(0.071)		(0.069)						
Jordanian	642	0.112	219	0.068	207	0.101	216	0.167	426	-0.033	435	-0.098***	423	-0.065**
		(0.012)		(0.017)		(0.021)		(0.025)						
Mental health service efficacy beliefs (SD)	642	0.022	219	0.081	207	-0.083	216	0.063	426	0.164*	435	0.018	423	-0.146
		(0.038)		(0.058)		(0.073)		(0.063)						
Friend group size	642	3.324	219	3.320	207	3.290	216	3.361	426	0.030	435	-0.041	423	-0.071
		(0.068)		(0.112)		(0.127)		(0.113)						
F-test of joint significance (F-stat)										1.217		1.542*		1.401
F-test, number of observations										370		369		359

Pair-wise regressions and F-test additionally control for governorate and randomization strata which are controls in the main analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.14: Baseline Recipient Attrition by Sender's Ranking of Recipient Need

	(1) Recipient Surveyed Baseline	(2) Recipient Surveyed Baseline
Ranked Recipient Need	-0.003	
	(0.007)	
Highest need recipient in friend group		0.011
		(0.021)
Control Mean	1	1
Covariates	Network Size	Network Size
N	2551	2551

This table shows that recipients who were reached for baseline were not ranked by senders and more or less in need than those recipients who were not reached at baseline. The regression restricts to friend groups of more than 1 person, and controls for the friend group size. Standard errors are clustered at the sender level. * p < 0.10, ** p < 0.05, *** p < 0.01

Table D.15: Recipient Attrition for Mental Health Take-Up Outcomes

	(1) Recipient Surveyed Endline	(2) Reached for survey helpline consent	(3) Reached for survey used mental health services after 6 months	(4) Reached for survey used mental health services after 6 months
Panel A: Pooled IV Estimates				
Sender shared (to anyone)	-0.030	0.041	0.006	0.012
	(0.075)	(0.094)	(0.066)	(0.074)
Control Mean	0.41	0.68	0.34	0.55
N	2668	1160	2668	1487
Sample Defined By:	All recipienthouseholds	Female recipient households	All recipienthouseholds	Male recipient households
Survey Respondents	Original recipient (male and female)	Only original female recipients	Females only: Original female recipient or any female in male recipient households	Women in household of male recipient
Panel B: ITT Estimates by Message Framing Arm				
Disclosed Compensation framing, *non-targeted*	-0.021	0.011	0.012	-0.008
	(0.028)	(0.037)	(0.027)	(0.032)
Disclosed Compensation framing, *targeted*	-0.043	0.030	-0.051**	-0.034
	(0.030)	(0.037)	(0.025)	(0.032)
Non-Disclosed Compensation framing, *non-targeted*	-0.029	-0.023	-0.042	-0.042
	(0.029)	(0.041)	(0.027)	(0.030)
Control Mean	0.41	0.68	0.34	0.55
N	2668	1160	2668	1487
Sample Defined By:	All recipient households	Female recipient households	All recipient households	Male recipient households
Survey Respondents	Original recipient (male and female)	Only original female recipients	Females only: Original female recipient or any female in male recipient households	Female in household of male recipient

This table tests for differential attrition across the sets of outcomes collected on recipients. Column 3 shows that there was differential attrition in the "disclosed compensation, targeted" message framing arm, relative to control, on the outcome of whether the respondent or anyone in their household had ever used mental health services, measured roughly 6 months after the main experiment. Note that only original female recipients (from the main experiment) were asked the helpline consent question used in the analysis (column 2) due to helpline programmatic priorities. For the outcome in column 3 *all* households of original recipients (male and female) were contacted for the 6 month follow-up questions, but the study interviewed only females in those households, again due to helpline programmatic priorities. Either the original female recipient or a (new) female respondent in the households of original male recipients was surveyed. Column 4 restricts the analysis to the sample of original male recipients and shows that there is no differential attrition for the outcome of using mental health services when restricting to female respondents in the households of original male recipients. Specifications cluster standard errors at the sender level and include no control covariates. * p < 0.10, ** p < 0.05, *** p < 0.01

Table D.16: Recipient Mental Health Take-Up by Restricted and Un-Restricted Samples

	(1) Ever used mental health services (after 6 months)	(2) Ever used mental health services (after 6 months)
Panel A: Pooled IV Estimates		
Sender shared (to anyone)	0.118	0.178
	(0.096)	(0.135)
Control Mean	0.230	0.190
Covariates	Lasso Double Selection	Lasso Double Selection
Sample	All respondents	Restricted, see notes
N	905	417
Panel B: ITT Estimates by Message Framing Arm		
Disclosed Compensation framing, *non-targeted*	0.072*	0.086
	(0.041)	(0.058)
Disclosed Compensation framing, *targeted*	0.062	0.100*
	(0.044)	(0.059)
Non-Disclosed Compensation framing, *non-targeted*	0.042	0.040
	(0.044)	(0.063)
Control Mean	0.23	0.19
Covariates	Lasso Double Selection	Lasso Double Selection
Sample	All Respondents	Restricted, see notes
N	905	417

This table shows that the marginally significant effect on mental health care take-up for recipients in the disclosed compensation arm remains qualitatively the same when considering the sample with or without attrition (noting that only the "disclosed compensation, targeted" arm exhibited differential attrition). Ever used mental health services is an indicator taking 1 if the respondent reports in the endline or 6 month follow-up that anyone in the household has ever used mental health services. Specification in column 1 includes all surveyed participants and column 2 is restricted to only female respondents in the household of an original male recipient – a subsample which did not display differential attrition (see Table D.15). Due to helpline programmatic priorities the survey measure was only collected with female respondents, who comprised both original female recipients, and, female respondents in the households of original male recipients. Covariates selected by lasso following Belloni et al. 2014. Standard errors are clustered at the sender level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.17: Types of Mental Health Care Taken-Up by Recipient

	(1) Called helpline (after 6 months)	(2) Used therapy in-person or remote (after 6 months)	(3) Used medication (after 6 months)	(4) Used other mental health care (after 6 months)
Panel A: Pooled IV Estimates				
Sender shared (to anyone)	-0.018	0.030	0.014	-0.012
	(0.044)	(0.075)	(0.069)	(0.070)
Control Mean	0.050	0.130	0.110	0.110
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	904	903	905	903
Panel B: ITT Estimates by Message Framing Arm				
Disclosed Compensation framing, *non-targeted*	-0.006	0.045	-0.010	0.026
	(0.018)	(0.034)	(0.027)	(0.030)
Disclosed Compensation framing, *targeted*	-0.006	0.013	0.006	-0.016
	(0.018)	(0.035)	(0.031)	(0.030)
Non-Disclosed Compensation framing, *non-targeted*	-0.008	0.028	0.029	-0.029
	(0.019)	(0.036)	(0.033)	(0.030)
Control Mean	0.05	0.13	0.11	0.11
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	904	903	905	903

This table show the types of mental health that the marginally significant effect on mental health care take-up for recipients in the disclosed compensation arm remains qualitatively the same when considering the sample with or without attrition (noting that only the "disclosed compensation, targeted" arm exhibited differential attrition). Ever used mental health services is an indicator taking 1 if the respondent reports in the endline or 6 month follow-up that anyone in the household has ever used mental health services. Specification in column 1 includes all surveyed participants and column 2 is restricted to only female respondents in the household of an original male recipient – a subsample which did not display differential attrition (see Table D.15). Due to helpline programmatic priorities the survey measure was only collected with female respondents, who comprised both original female recipients, and, female respondents in the households of original male recipients. Covariates selected by lasso following Belloni et al. 2014. Standard errors are clustered at the sender level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.18: Heterogeneity in Sender ability to target within friend group

	(1) Recipient depressed at baseline (0/1)	(2) Recipient depressed at baseline (0/1)	(3) Recipient depressed at baseline (0/1)	(4) Recipient depressed at baseline (0/1)	(5) Recipient depressed at baseline (0/1)	(6) Recipient depressed at baseline (0/1)
Highest need recipient in friend group	0.104***	0.106***	0.111***	0.106***	0.103***	0.106***
	(0.029)	(0.030)	(0.031)	(0.037)	(0.029)	(0.039)
Highest Need X Sender stigma 1st order	0.036					
	(0.029)					
Highest Need X Sender stigma 2nd order		-0.004				
		(0.029)				
Highest Need X Sender altruism			0.007			
			(0.031)			
Highest Need X Sender female				-0.002		
				(0.058)		
Highest Need X Sender social desirability					0.027	
					(0.029)	
Highest Need X Sender depressed						-0.001
						(0.058)
Control Mean	0.436	0.436	0.436	0.436	0.436	0.436
Covariates	No controls	No controls	No controls	No controls	No controls	No controls
N	1325	1308	1162	1330	1330	1329

This table shows that there is no significant heterogeneity by sender characteristics in senders' ability to identify which of their friends is more in need. Observations are at the recipient level. The sample is restricted to instances when the sender has more than 1 friend and includes only the recipients that were reached for the baseline survey. The independent variable is a binary variable of the sender having indicated that the recipient would benefit the most from mental health information. The dependent variable is an indicator for whether the recipient's PHQ-9 score at baseline indicates that the recipient likely has moderate to severe depression (10 or higher). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.19: Recipient Impacts: Outcome Family 1

	Called Helpline	Benefit from MH services Relative to Stigma (SD)	Willingness to Share Own Story (SD)	Willing to accept call from helpline (after 6 months)
Panel A: Pooled IV Estimates				
Sender shared (to anyone)	-0.017	-0.001	0.080	-0.045
	(0.021)	(0.188)	(0.217)	(0.126)
FDR-adjusted q-value	1	1	1	1
Control Mean	0.016	0.003	1.012	0.536
Double selection	Yes	Yes	Yes	Yes
N	1041	906	1042	812

	Called Helpline	Benefit from MH services Relative to Stigma (SD)	Willingness to Share Own Story (SD)	Willing to accept call from helpline (after 6 months)
Panel B: ITT Estimates by Message Framing Arm				
Disclosed Compensation framing, *non-targeted*	-0.009	-0.017	0.020	0.036
	(0.009)	(0.091)	(0.104)	(0.050)
Disclosed Compensation framing, *targeted*	-0.004	-0.060	0.150	-0.078
	(0.010)	(0.089)	(0.109)	(0.049)
Non-Disclosed Compensation framing, *non-targeted*	-0.008	-0.095	-0.012	0.002
	(0.009)	(0.098)	(0.108)	(0.051)
FDR-adjusted q-value				
Disclosed_non-targeted	1	1	1	1
Disclosed Compenation_targeted	1	1	1	1
Non-Disclosed_non-targeted	1	1	1	1
Control Mean	0.02	0.00	1.01	0.54
Double selection	Yes	Yes	Yes	Yes
N	1041	906	1042	812

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.20: Recipient Impacts: Outcome Family 2

	PHQ-9 Score	GAD-2 Score (SD)	Friend Support Index (SD)	Social Connectedness Index (SD)	Any conversations about mental health	Labor Assistance Freq. (SD)
Panel A: Pooled IV Estimates						
Sender shared (to anyone)	-0.096	0.015	0.135	0.372**	0.159**	0.448**
	(0.160)	(0.155)	(0.176)	(0.186)	(0.063)	(0.202)
FDR-adjusted q-value	.494	.856	.494	.083	.07	.071
Control Mean	-0.001	-0.004	0.000	0.000	0.111	-0.006
Double selection	Yes	Yes	Yes	Yes	Yes	Yes
N	1042	1042	1040	1042	1038	1042

	PHQ-9 Score	GAD-2 Score (SD)	Friend Support Index (SD)	Social Connectedness Index (SD)	Any conversations about mental health	Labor Assistance Freq. (SD)
Panel B: ITT Estimates by Message Framing Arm						
Disclosed Compensation framing, *non-targeted*	-0.019	-0.061	0.040	0.107	0.050*	0.092
	(0.077)	(0.073)	(0.085)	(0.091)	(0.030)	(0.095)
Disclosed Compensation framing, *targeted*	-0.038	0.077	0.110	0.205**	0.070**	0.201**
	(0.077)	(0.075)	(0.080)	(0.092)	(0.034)	(0.102)
Non-Disclosed Compensation framing, *non-targeted*	-0.027	0.094	0.030	0.075	0.041	0.085
	(0.081)	(0.085)	(0.093)	(0.092)	(0.031)	(0.098)
FDR-adjusted q-value						
Disclosed_non-targeted	1	.798	1	.798	.54	.798
Disclosed Compenation_targeted	1	.798	.793	.423	.423	.423
Non-Disclosed_non-targeted	1	.798	1	.798	.793	.798
Control Mean	-0.00	-0.00	0.00	0.00	0.11	-0.01
Double selection	Yes	Yes	Yes	Yes	Yes	Yes
N	1042	1042	1040	1042	1038	1042

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.21: Recipient Impacts: Outcome Family 3

	Expected Benefit from MH Care (SD)	Concern: Not considered reliable by friends (SD)	Concern: Confidentiality Breached (SD)	Stigma Beliefs 2nd Order (SD)	Own Stigma Index (SD)
Panel A: Pooled IV Estimates					
Sender shared (to anyone)	0.066	0.208	-0.051	0.052	0.066
	(0.168)	(0.176)	(0.176)	(0.198)	(0.176)
FDR-adjusted q-value	1	1	1	1	1
Control Mean	0.001	0.006	0.008	-0.005	0.005
Double selection	Yes	Yes	Yes	Yes	Yes
N	1035	1042	1041	909	1027

	Expected Benefit from MH Care (SD)	Concern: Not considered reliable by friends (SD)	Concern: Confidentiality Breached (SD)	Stigma Beliefs 2nd Order (SD)	Own Stigma Index (SD)
Panel B: ITT Estimates by Message Framing Arm					
Disclosed Compensation framing, *non-targeted*	0.015	0.062	-0.074	0.020	0.057
	(0.082)	(0.083)	(0.081)	(0.090)	(0.088)
Disclosed Compensation framing, *targeted*	-0.097	0.058	-0.013	-0.022	-0.005
	(0.083)	(0.086)	(0.087)	(0.092)	(0.090)
Non-Disclosed Compensation framing, *non-targeted*	-0.024	0.144*	0.082	0.112	0.149
	(0.091)	(0.085)	(0.091)	(0.093)	(0.102)
FDR-adjusted q-value					
Disclosed_non-targeted	1	1	1	1	1
Disclosed Compenation_targeted	1	1	1	1	1
Non-Disclosed_non-targeted	1	1	1	1	1
Control Mean	0.00	0.01	0.01	-0.01	0.01
Double selection	Yes	Yes	Yes	Yes	Yes
N	1035	1042	1041	909	1027

* p < 0.10, ** p < 0.05, *** p < 0.01

Table D.22: Recipient Impacts: Outcome Family 4

	Ever used mental health services (after 6 months)	Used any MH care 30 days	Knowledge Index (SD)	Knows of the Helpline	Exercise Past 7 Days	Shared MH Information With others (0/1)
Panel A: Pooled IV Estimates						
Sender shared (to anyone)	0.118	-0.008	0.143	0.031	-0.051	0.070
	(0.096)	(0.038)	(0.179)	(0.019)	(0.177)	(0.055)
FDR-adjusted q-value	.799	.799	.799	.799	.799	.799
Control Mean	0.234	0.073	0.000	0.012	-0.002	0.098
Double selection	Yes	Yes	Yes	Yes	Yes	Yes
N	905	1037	1042	1034	1042	1042
	Ever used mental health services (after 6 months)	Used any MH care 30 days	Knowledge Index (SD)	Knows of the Helpline	Exercise Past 7 Days	Shared MH Information With others (0/1)
Panel B: ITT Estimates by Message Framing Arm						
Disclosed Compensation framing, *non-targeted*	0.072*	-0.011	0.053	0.006	0.038	0.028
	(0.041)	(0.018)	(0.084)	(0.010)	(0.085)	(0.027)
Disclosed Compensation framing, *targeted*	0.062	0.018	0.110	0.006	-0.056	0.023
	(0.044)	(0.022)	(0.083)	(0.012)	(0.087)	(0.029)
Non-Disclosed Compensation framing, *non-targeted*	0.042	-0.000	0.073	0.001	-0.028	0.034
	(0.044)	(0.020)	(0.089)	(0.010)	(0.086)	(0.028)
FDR-adjusted q-value						
Disclosed_non-targeted	1	1	1	1	1	1
Disclosed Compenation_targeted	1	1	1	1	1	1
Non-Disclosed_non-targeted	1	1	1	1	1	1
Control Mean	0.23	0.07	0.00	0.01	-0.00	0.10
Double selection	Yes	Yes	Yes	Yes	Yes	Yes
N	905	1037	1042	1034	1042	1042

* p < 0.10, ** p < 0.05, *** p < 0.01

Table D.23: Recipient Impacts: Outcome Family 5

	Employed (0/1)	Earnings Monthly (JD)	Borrowed (0/1)	Lent (0/1)
Panel A: Pooled IV Estimates				
Sender shared (to anyone)	-0.060	-2.728	0.183*	-0.025
	(0.072)	(15.421)	(0.101)	(0.048)
FDR-adjusted q-value	1	1	.392	1
Control Mean	0.363	57.649	0.594	0.086
Double selection	Yes	Yes	Yes	Yes
N	1040	1040	1041	1041
	Employed (0/1)	Earnings Monthly (JD)	Borrowed (0/1)	Lent (0/1)
Panel B: ITT Estimates by Message Framing Arm				
Disclosed Compensation framing, *non-targeted*	-0.027	-1.486	0.033	-0.005
	(0.034)	(7.234)	(0.044)	(0.022)
Disclosed Compensation framing, *targeted*	-0.025	-4.639	0.106**	-0.014
	(0.033)	(6.934)	(0.045)	(0.025)
Non-Disclosed Compensation framing, *non-targeted*	-0.001	7.436	0.114**	-0.009
	(0.034)	(8.603)	(0.044)	(0.023)
FDR-adjusted q-value				
Disclosed_non-targeted	1	1	1	1
Disclosed Compenation_targeted	1	1	.138	1
Non-Disclosed_non-targeted	1	1	.138	1
Control Mean	0.36	57.65	0.59	0.09
Double selection	Yes	Yes	Yes	Yes
N	1040	1040	1041	1041

* p < 0.10, ** p < 0.05, *** p < 0.01

Figure D.3: Recipients' Baseline Self-Assessed Distress Levels: Current and If Hypothetically Started Using Care

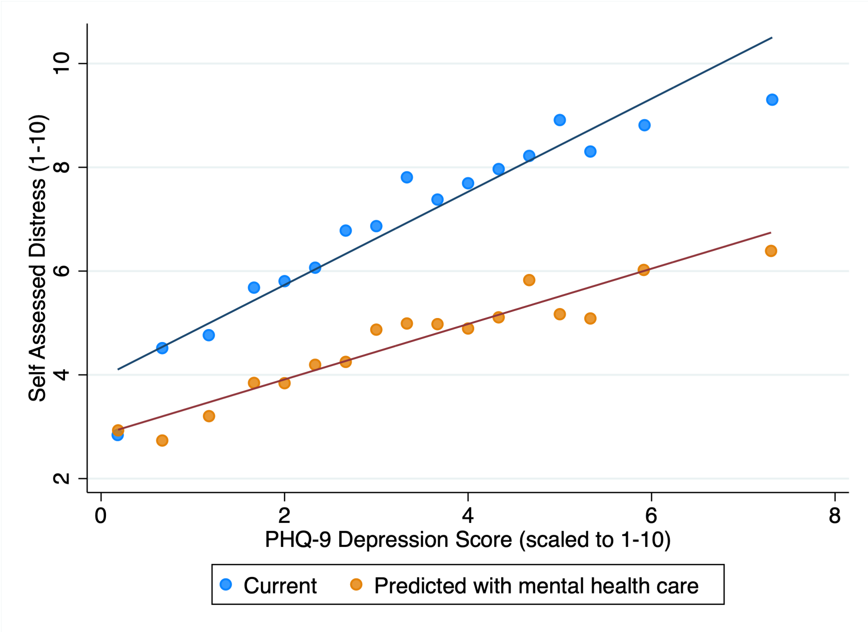


Figure D.4: Proportion of Representative Sample Agreeing: "If I were young and unmarried I would not marry someone who ever used mental health services."