Stigma and Social Cover: A Mental Health Care Experiment in Refugee Networks

Emma C. Smith*

Latest Version Here

This Version: March 14, 2025

Abstract

People might withhold useful information from others to avoid being associated with a stigmatized product or service. In a field experiment with 847 Syrian refugee friend groups, I investigate whether an external excuse can increase the exchange of information on a stigmatized topic – specifically, mental health services. First, I document simultaneously significant local knowledge about who may be depressed, positive beliefs about mental health treatment efficacy, and a reluctance to share information about services: only 22% of friends receive information. The study's main finding is that giving individuals social cover, by encouraging them to disclose that they are compensated to share information, raises sharing rates by 37%. These effects are strongest for senders who are prior mental health care users. In a reversal of the common prediction that financial incentives may crowd out prosocial behavior, I instead find that in this setting with stigma, increasing the observability of financial incentives crowds in prosocial behavior. In a follow-up experiment I show that senders can use the excuse of being paid without decreasing recipients' interest in the services.

*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, and seminar, workshop and conference participants at Georgetown University, Harvard University, NEUDC, UNHCR-WB Joint Data Center Research Conference on Forced Displacement, and Y-Rise. 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

Social services meant to help vulnerable people are often associated with stigma. This paper proposes that stigma does not just prevent people from using services, but can even prevent them from *learning* about such services[1]. Because people who use stigmatized services often do not want to be associated with them, users may be reluctant to share information with others. In fact, even non-users may worry that if they share information, others will assume they have used the stigmatized services.

I experimentally investigate whether image concerns constrain information sharing on stigmatized topics, in the context of mental health services for Syrian refugees in Jordan. 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 – one in three refugees have depression, anxiety or post-traumatic stress disorder (PTSD) (Blackmore et al., 2020)–, and in Jordan specifically the rate rises to more than half of Syrian refugees. In this setting I document high rates of mental health stigma, little care-seeking, and low knowledge of mental health services, together motivating the study hypothesis: fear of being labeled a mental health service user may prevent information from circulating.

Stigma and image concerns may not be relevant if people do not believe their friends need mental health services. But instead, I find that at baseline individuals have accurate knowledge of their friends' mental health need, above and beyond what can be explained by observable covariates. Prior to any intervention, the study asks one friend in the group to rank their friends by mental health need, and within 3 days those friends are surveyed on their mental health. Individuals that were identified as the most distressed in their friend group are 25% more likely to be experiencing depression. Baseline data also indicates that individuals recognize that they are psychologically distressed and self-report that they believe mental health services could lower their distress, yet only a third can name an organization providing mental health care. I therefore investigate whether image concerns sustain this knowledge gap.

The main experiment measures Syrian refugees' willingness to share information about a free counseling service with their friend group. 847 refugee friend groups participate in the experiment, consisting of one pre-selected "sender" and their elicited network of close friends. Senders are randomized either to control or to treatment, where the "sender" is asked to share

[1]For examples of stigma constraining service use, see Yang et al., 2023, Yu, 2023, Derksen et al., 2022, Celhay et al., 2022 Anders and Rafkin, 2022, Schanzenbach, 2009, Lasky-Fink and Linos, 2022, Bhargava and Manoli, 2015, Gesesew et al., 2017, Gesesew et al., 2017, Currie et al., 2001, Henderson and Thornicroft, 2009, Derksen et al., 2022, Suomi et al., 2022

a week-long mental health awareness campaign over WhatsApp with their network. The campaign repeatedly advertises a free counseling helpline, which "recipients" (the senders' friends) can take-up. I measure sharing rates with a combination of recipient self-reports, incentivized screenshots from senders and, importantly, unique trackable links, which provide a difficult-to-manipulate measure that validates the main recipient-level outcome.

Sharing this information might signal something negative about the sender's or their friend's social image.[2] For one, senders may worry that sharing the information will signal that they used the service themselves. Second, the sender may also worry that sharing information will insult her friend by insinuating that the friend needs mental health services. To identify these concerns I randomly vary two dimensions of the senders' messages. For the first source of variation, the study encourages some senders to reveal in their introduction that they are being paid to share the campaign. 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. The 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 say they are trying to share 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) since everyone can likely benefit even if just by having information to share with others, so again the sender's requested participation level is held constant. But the different messages to recipients creates variation in whether recipients think they were targeted. (I refer to this framing as "targeted" throughout, not because it varies the selection of who should receive information but because it leads recipients to think they were targeted.) The only differences in sender messages across treatment conditions comes from the one-sentence introduction the sender is instructed to use.

Disclosing that the sender is paid provides "social cover" because it dampens the signal of why the sender chose to share the information. Supposing that before someone who shared was regarded to very likely be a mental health care user, now an observer might think "Maybe she used these services, or maybe she is just sharing to get the payment." If senders share more when encouraged to disclose, then this indicates that the signal cost from sharing is binding.

While the payment disclosure dampens the signal of sharing, suggesting that the recipient is in need instead sharpens the signal. Recipients exposed to this "targeted" introduction are more likely to think that their friend thinks they are in need. If senders share less when encouraged to use the targeted framing, then it indicates that they internalize the image cost to their friend.

[2]Social image is defined following the intuition laid out in Bursztyn and Jensen, 2017: "an individual exhibits social image concerns when her utility depends on the posterior expectations of her type held by others, conditional on observing her behavior".

On the other hand if senders share more, it could indicate that senders think there is value to informing friends that they are a good fit for the service.

> 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.[3]
>
> Non-Disclosed Compensation + Non-targeted: I want to try to share this with all of my close friend. [4]

The first stage is consistent with low willingness in general to share information about mental health. Despite often knowing their friends' need, sender share the campaign with only 24 percent of the treatment group recipients. Moreover this is in spite of the fact that senders were screened at baseline on stating willingness to participate in the campaign.

But, encouraging senders to disclose that they are paid led to significantly higher sending rates. Consistent with the hypothesis that disclosing payment creates a stigma-alleviating excuse for senders, senders are 6.2 percentage points (37%) more likely to send the two framings that disclose the excuse of compensation, relative to the framing in which senders say they "want to share" but are still compensated privately (p-value 0.038).[5] The increase indicates that sharing mental health information carries a costly signal, and revealing that the sender is paid dampens this signal. On the other hand I detect no effect of encouraging the sender to say that they are targeting the recipient. 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 result in first-stage differences in which framing the sender used.

Heterogeneity analysis provides further evidence of image concerns. I use machine learning heterogeneity following Chernozhukov et al., 2018 after determining that no pre-specified dimensions of heterogeneity were associated with significant differences in responsiveness to the disclosure treatment. I find there is strong evidence of heterogeneous treatment effects of the

[3]In all conditions senders are incentivized to share with a pre-specified list of all their close friends. In the private instructions to the sender I remind the sender that, due to the high mental health burden in their community, all people may directly or indirectly benefit from learning about mental health services.

[4]Senders in all treatment arms were offered equal compensation, and were reminded of the compensation each time throughout the week when they were asked to send new content.

[5]Treatment and control are balanced but imbalance arose between the disclosed+targeted and non-disclosed+non-targeted framings (F-stat = 2.32, see appendix). Additionally 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 appendix shows that the results remain qualitatively very similar, with some loss of statistical significance, when forcing the inclusion of imbalanced covariates and dropping the affected week.

disclosure framing, and that one characteristic explains the heterogeneity better than any other – whether the sender was a prior user of mental health services herself. Prior users are more than twice as likely to send the "disclosed compensation" than the "non-disclosed compensation" framing (p-value 0.002), while non-users share both framings at similar rates (p-value 0.855). While not a direct prediction of the model, the fact that people who have used stigmatized services are most sensitive to the fear of "outing" themselves lends credence to the underlying mechanism.

Additional heterogeneity also reveals that senders prioritize sharing messages with friends who are most in need. When senders have social cover from the "disclosed compensation" framing, they send at equally high rates to recipients who do and do not likely have depression or anxiety (0.01 percentage point difference, p-value 0.78). But when not provided with social cover in the "non-disclosed compensation" framing, senders share less and target the most in-need recipients. Specifically recipients with likely depression or anxiety are 9.1 percentage points (59%, p-value 0.06) more likely than those without depression or anxiety to receive the "non-disclosed compensation" framing.[6] This points to senders making informed trade-offs between the marginal benefit their friends may get from information and the marginal social image cost.

Considering next the impacts on recipients, I find no average effect on take-up of formal mental health services. I measure take-up in several ways, using an endline survey roughly 1 month after senders shared content, match the sample list to the helpline's administrative data on new callers, and six months later, contact a subset of recipients to measure demand for the service. The most significant effects of the campaign were on activation of informal social support. Treated recipients experience a 0.32 standard deviation increase in social connectedness driven by a 0.41 standard deviation increase in the number of times the recipient spent time helping or being helped by a network member (p-values 0.037 and 0.015 respectively). Treated recipients also engage in 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 0.010). And, recipients also report borrowing 21 percentage points more often, with no effect on lending, in a setting in which the vast majority of borrowing is informal and often with little expectation of repayment (p-values 0.022 and 0.957 respectively).

I use a follow-up experiment to show that senders can use the excuse of being paid without recipients de-valuing services. Ex-ante, one may worry that disclosing that the sender is being paid may cause the recipient to discount the information being shared. The targeted framing could also impact recipients' demand, by helping her learn she is a good fit for the program or

[6]Depression and anxiety are measured at baseline and recipients with a PHQ-9 score of 10 or above are considered depressed while those with a GAD-2 score of 3 or higher are considered to have anxiety.

provoking backlash. This effect cannot be measured in the main experiment due to endogenous selection in who the sender messages. Therefore in the short follow-up experiment, with new senders, I shut down selection by having the study, rather than the sender, share the information on the sender's behalf, using her name. [7]

The average effect of disclosing that the sender is paid is insignificant, as is the average effect of "targeted" phrasing. This promising finding suggests that a policy of disclosing that senders are paid may increase sending without negatively affecting how recipients respond to the information. On a more cautionary note, I find that disclosing that the sender is paid interacts negatively with "targeted" phrasing, indicating that overall a "targeted" phrasing has a null or negative effect on sharing and service take-up.

Together these results show, first, that individuals withhold useful information about stigmatized services, despite knowing their friends' need. Generalizing the results suggest that, from food banks and unemployment assistance to HIV and substance use disorder treatment, users of stigmatized services may withhold referrals and recommendations to avoid signaling their own use. But dampening the signal of why the sender is sharing can significantly increase the flow of information, without necessarily decreasing recipients' demand for services. This highlights a new application of the externalities of financial incentives: in contrast to a prior literature that has largely focused on how financial incentives may crowd out prosocial actions, in this setting financial incentives crowd in prosocial actions (Bénabou and Tirole, 2006, Gneezy et al., 2011). This reversal is consistent with the presence of stigma, and may be applicable in broader contexts. The findings point to policy implications on how to raise knowledge of stigmatized products and services. Often only a small subset of people have this information, and so those in need could face large social costs if searching at random. If informed community members know who to target, as this and other work has shown (see Hussam, Rigol, and Roth, 2022, Goldberg et al., 2018 and Alatas et al., 2016), then focusing on the sending side of information sharing may be a promising strategy.

This paper contributes to three literatures. First, the study contributes to the literature on social learning, and specifically the role of image/reputational concerns. While the prior literature clearly documents reputational concerns in networks in general (Breza and Chandrasekhar, 2019 and Karing, 2018), and in information seeking specifically (Chandrasekhar et al., 2018 and Banerjee et al., 2018), this is one of the first studies to consider the *sharing* party's side, showing that individuals may withhold information due to image concerns. Notably other recent work investigates the related but separate question of how rationales influence willingness to share

[7] Due to the counseling service's programmatic priorities the follow-up experiment was conducted exclusively with women.

information publicly, and provides complementary findings to this paper (Bursztyn et al., 2023). Constraints on willingness to share information is under-explored but has important implications for existing models of social learning (Cefalà et al., 2024).

Second, the paper contributes to the literature on social image in general, by investigating a novel approach to alleviate social image concerns: directly dampening signals that lead to negative inferences. This new direction sheds light on how to incentivize behavior in the presence of real (not misperceived) social stigma, and in doing so builds significantly on a prior literature that largely focused on either correcting incorrect perceptions of stigma (Bursztyn et al., 2020) or avoiding social stigma concerns entirely by providing privacy (Bursztyn et al., 2019, Bursztyn and Jensen, 2017). This paper is part of a strand of recently, largely contemporaneous, work investigating this new direction (Raisaro, 2023, Buchmann et al., 2021 and Bursztyn et al., 2023).

Third, this work contributes to the literature on take-up of mental health services, in general and among refugees in particular. With mental illness the leading cause of disability globally (Bloom et al., 2012), and concentrated among vulnerable populations such as those in low-income countries and displaced people (Banerjee et al., 2023, Stillman et al., 2022), there is an urgent need to increase both supply and demand of mental health services. This paper contributes to understanding demand-side constraints, and is motivated by prior work that has established the negative relationship between mental health with economic well-being (Ridley et al., 2020), the efficacy of mental health treatment in low-resource settings including with forcibly displaced populations (Bhat et al., 2022, Harker Roa et al., 2023, Islam et al., 2021), and additional costs of poor mental health such as workplace discrimination (Ridley, 2022). In focusing on stigm and information as barriers to accessing services, this paper also contributes to a rich literature on stigma in other areas of public health, perhaps most notably regarding HIV (Yang et al., 2023). Finally, as a field experiment with refugees this study contributes to a small but growing body of randomized controlled trials in humanitarian settings (Alan et al., 2021, Baseler et al., 2023, Hussam, Kelley, et al., 2022, Tamim et al., 2025).

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 results and Section 6 concludes.

2 Context

I conduct the study with a sample of Syrian refugees living in Jordan, and their social network which may include non-refugees. Most Syrian refugees in Jordan live outside of camps among

the host population, and a majority have been in Jordan for 9 years, having been predominantly displaced in 2013 by the Syrian Civil War, which began in 2011. In this setting there is a large mental health burden, with roughly 50% of the adult population having symptoms aligned with clinical depression (Stillman et al., 2022). However use of mental health services remains low and cannot be explained by self-reported efficacy beliefs. Prior to beginning the experiments I conduct a nationally representative survey of 1516 Syrian refugees in Jordan. I again find that roughly 50% of the adult population has symptoms of clinical depression, and that 85% indicate the belief that mental health services can be effective. Yet less than 7% of households had someone seek mental health services in a 2 week recall period, and 45% said that they would not marry someone who once sought professional mental health services. These statistics motivate the study's focus on the role of stigma in social learning about mental health services.

In an effort to increase knowledge of available services and decrease stigma toward care-seekers, I collaborated with the International Rescue Committee to spread mental health awareness content designed by the INGO. I then partnered with the Jordan River Foundation, a local Jordanian organization which offers a free counseling helpline, to advertise the organization's services and measure take-up.

2.1 Motivating Descriptive Evidence

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 to construct the intended "recipient" sample. I then implement a main information sharing experiment in which I "seed" information about a mental health phone counseling service and randomize the framing of the information to study stigma barriers to sharing. 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

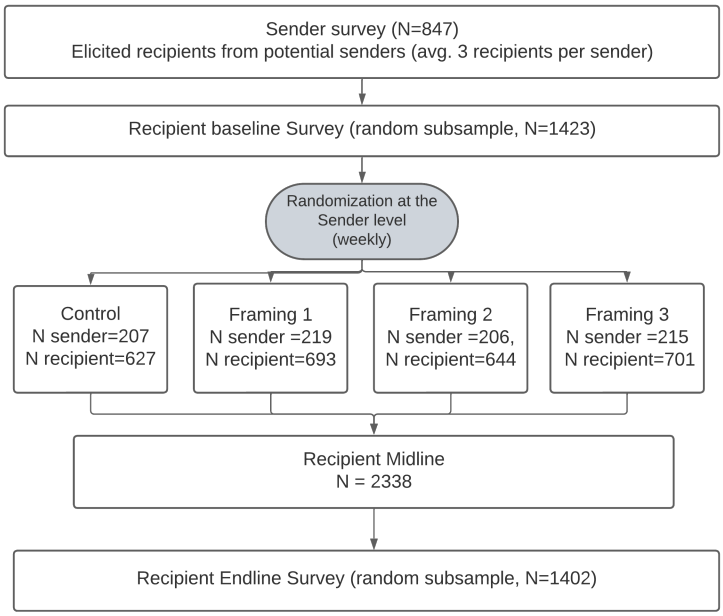
The study sample comprised a "sender" group who was asked to share awareness content with their friends (when assigned to treatment), and a "recipient" group who were the senders' friends. The enrolled "sender" sample comprised 847 individuals who were nominated by peers in an otherwise-unrelated representative survey of Syrian refugees across Jordan in late 2021

through early 2022.[8] Potential senders could be nominated if the nominator felt the person fell into any of the following three categories: being "well-regarded or well-know", or "community-minded", or "good at spreading news", with the final category informed by the literature on identifying individuals with high diffusion centrality (Banerjee et al., 2013).

The sender sample was surveyed by phone from January 3, 2023 to February 22, 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. The final sender sample consisted of 847 senders who provided the phone numbers for their friends and expressed willingness to participate in the WhatsApp intervention (described below). The final recipient sample comprised those senders' friends and consisted of 2665 individuals.[9]

The study design is summarized in Figure 1 below.

Figure 1: Experimental Design



[8]The representative sample was drawn from the UNHCR universe of registered Syrian refugee households. Out of 1516 surveys of the representative sample, 726 respondents agreed to nominate individuals and the average number of nominations was 2.5. Respondents were presented with the three nomination types in random order. Respondents were not made aware of the mental health focus on the intended study.

[9]5% of the recipient sample appeared in more than 1 friend group.

3.2 Targeting

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. The respondent was asked[10]:

> "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."

3.3 Eligibility for the Experiment

Next, still within the same survey, the senders were asked to share the phone numbers for their friends. Conditional on sharing any phone numbers for their friends, 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 randomized either to treatment or control with 207 senders assigned to control and 640 assigned to one of three treatments described below. Randomization was at the sender level and stratified on gender and the sender's original nominator if the nominator identified multiple senders. Control group senders were not contacted again for the study and did not receive the awareness content during the experiment period.

3.4 WhatsApp Messaging Intervention

Treatment group senders received the mental health awareness campaign designed by the International Rescue Committee in Jordan. The content was developed with extensive input from Syrian refugee community members and the professional and cultural expertise of a Jordanian psychologist. The campaign content 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. Each batch of content advertised the Jordan River Foundation's free phone counseling helpline. The content was sent

[10]Less than five percent of the sender sample declined to do this ranking.

in 3 batches over 8 days. The campaign was administered on a rolling basis in weekly batches, such that all 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.[11] To remind the sender who to message, the recipient friends' names were listed in the instructions that the senders got from the study on each of the three days that they got content throughout the campaign.[12]

3.5 Recipient Baseline

Each week, immediately following each sender survey and before the next wave of the WhatsApp intervention was sent, baseline phone surveys were attempted with the new recipients. Of the 2665 recipients, 1423 were reached for a baseline 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. At baseline recipients completed 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.

3.6 Main Experiment

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 with his or her friends began with "Here is some mental health information I received from [an NGO]." 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. [13]

[11] See appendix for examples of the content and instructions layout. Instructions were re-sent with each of the three batches of content, and three reminders were sent, each one day after a batch of content was sent.

[12] 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, depending on the respondent's preference.

[13] Senders in all treatment arms were offered equal compensation, and were reminded of the compensation each time throughout the week when they were asked to send new content.

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. In all three conditions the sender was told to share the content with all of the friends who she mentioned in the baseline survey, and specifically those friends' 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 framings enable me to test key predictions of the model. 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. If disclosing that there is compensation leads to different rates of sending then it is evidence that there is indeed a social image signal from sending the content, and an excuse in the form of a financial incentive dampens that signal. If dampening the signal specifically increases sending, then it shows that the social image signal of sending is negative, consistent with the presence of stigma.

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 for 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 knows the recipient may be in need. From a policy perspective this might 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. Formally this can come from the recipient having belief-based utility, here simply meaning that she gets negative utility from her belief that others think she is vulnerable. 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.

3.7 Follow-Up Data Collection for Recipients

For each wave of recipients, the week after the campaign was implemented all recipients were contacted to completed a two-question midline phone survey that asked if they had received the campaign and if they had used the advertised helpline. 2,550 recipients were reached for the midline check-in, representing 95% of the total recipient sample.

An endline phone survey was conducted with a random sample of the recipient sample three weeks after the recipient's last message was scheduled to be received. 1,051 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.

Six months after the intervention, female recipients were contacted and asked if they would like the helpline to contact them directly to receive free phone counseling. 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.

3.8 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 and coded as 1 (relative to 0) if any of these data sources indicate that the content was ever shared. For the primary measure of sender follow-through I consider whether the sender sent to anyone in her friend group.

I additionally construct the outcome from the recipient's perspective, 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. Measured rates of sending and receiving are not equivalent for two reasons. First, a sender might not share with everyone in her friend group. Second, in many sender screenshots it was impossible to conclusively determine the recipient of the WhatsApp message pictured.[14] In these cases the sender was coded as having shared, but no recipient was coded as "1" in return.

I use personalized trackable links as 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 unique devices. By assigning each sender unique links, I am able to measure which senders' content was engaged with more regardless of whether the sender

[14]In Middle Eastern culture it is common for people to call their friends by nicknames such as "Mother of Mohammad". Often participants' WhatsApp contacts were saved with nicknames, while the study sample lists contained only full names.

or recipient self-reported sending or receiving the content.

3.9 Follow-Up Experiment

Six months after the main experiment I conduct a follow-up experiment to measure the exogenous impact of the framings on recipient demand for phone counseling. The main experiment provides evidence on the exogenous impact of different message framings on sending rates, but does not identify whether the framing that senders use affects recipients' demand for services. This is because in the main experiment the sender decides whether the recipient gets a message and therefore the estimated effects of messages on recipients are endogenous. Recipients' demand for phone counseling could be affected by the framings in multiple ways. While the "disclosed compensation" framing was effective at increasing sending rates, one might be concerned that this framing will lead recipients to de-value the helpline after inferring that the sender does not think the helpline is useful. Or, recipients may infer from the "disclosed compensation" framing that stigma is high (hence the sender's desire to use an excuse) and be deterred from using the helpline. The "targeted" framing might encourage or discourage recipient take-up. The "targeted" framing may increase take-up if recipients learn from the framing that they are a good fit for the service. But if the "targeted" framing causes the recipient to feel singled out and exposed there could be a backlash effect. I use the second experiment to identify whether these effects are at play.

Enumerators re-contacted recipients from the original study and surveyed 443 women by phone.[15] I refer to these participants as focus respondents in the follow-up experiment. After eliciting the focus respondent's social network the enumerator asked the respondent for permission for the study to contact her friends and inform them about the helpline, while mentioning the focus respondent's name. The enumerator elicited the focus respondent's permission to be allowed to use any of 4 different framings of the content when introducing the helpline to the focus respondent's friends and respondents who agreed received a financial incentive.[16] 85% of respondents agreed that their friends could be contacted by the study with any of the messages.[17]

The four framing conditions reflect the same framings used in the main experiment with the addition of the framing "non-disclosure + targeted".

[15] Due to programmatic priorities of the phone counseling service I included only women.

[16] Respondents received a small financial compensation that increased with the number of friends who they listed and agreed could be messaged and the average compensation was $2.20.

[17] The design allows me to test for differences in focus respondents' willingness to let the study contact their friends depending on different framings. However there were no significant differences in willingness to have friends contacted with different framings. The focus respondents' rates of agreement were not significantly different from senders' average rate of agreeing verbally to participate in the main experiment. There 92% of senders said initially that they would share the campaign, but then only a fraction followed through. These results suggest that the constraints measured in this study become binding when senders have to take proactive actions.

Disclosed Compensation + Non-targeted: Our project offered to pay your friend [focus respondent name] to help us check if her friends are interested in this helpline, and she mentioned that you are one of the people she knows.

Disclosed Compensation + Targeted: Our project offered to pay your friend [focus respondent name] to help us check if her friends are interested in this helpline, and she indicated that you or people you know might benefit from the information.

Non-Disclosed Compensation + Non-targeted: Your friend [focus respondent name] wanted us to check if her friends are interested in this helpline, and she mentioned that you are one of the people she knows.

Non-Disclosed Compensation + Targeted: Your friend [focus respondent name] wanted us to check if her friends are interested in this helpline, and she indicated that you or people you know might benefit from the information.

If the focus respondent agreed to all four framings then immediately after the survey the enumerator attempted to contact the friends by phone, and read a randomly assigned framing. Framings were randomized at the final recipient level and stratified on the focus respondent's treatment status in the main experiment and the first framing the focus respondent was exposed to in the follow-up experiment. Immediately after introducing the helpline the enumerators asks if the friend is interested in using the helpline, which is recorded as a binary outcome of helpline demand.

4 Conceptual Framework

The conceptual framework closely follows the signaling model proposed by Chandrasekhar et al., 2018, adapted for sending rather than seeking information and focusing exclusively on the social image costs of sending information.[18] Senders decide whether to send information given a health benefit to the recipient and social image costs to both the sender and recipient. The sender's image cost arises from being associated with mental health services or, more simply, people thinking that she is a mental health care user. When being a prior user is stigmatized, as it is in this context, there is a social image cost to sending information. Likewise recipients face a social image cost of receiving information, which senders may internalize. The social image cost to recipients comes from the fact that people who are more in need will benefit from information more, and so those who receive information are more likely to be in need.

[18] Unlike Chandrasekhar et al., 2018 I do not consider an "interaction" cost on top of the social image cost.

In the sender population a fraction π are prior users of mental health services (type A) and the remainder are non-users (type B). There are two types of recipients, the vulnerable type V and the unvulnerable type UV, with ω representing the proportion of recipients who are the vulnerable type. 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.

$$U(S) = \underbrace{\xi(h\mathbf{1}_{s=1})}_{\substack{\text{Health benefit} \\ \text{to recipient} \\ \text{(sender's belief)}}} - \underbrace{\varphi(\pi_{post})}_{\substack{\text{social image cost} \\ \text{to sender}}} - \underbrace{\phi(\omega_{post})}_{\substack{\text{social image cost} \\ \text{to recipient}}} \tag{1}$$

where h is the recipient's health benefit from the information and $\xi(h)$ are the sender's beliefs about the benefit for the recipient where $\xi(.)$ is a monotonic function of h and h is distributed normally. $\pi_{post} = P(Type_s = A|S)$ is the posterior belief that the sender is a prior user given the observed sending decision. Likewise $\omega_{post} = P(Type_r = V|S)$ is the probability that the recipient is vulnerable given the sending decision. $\varphi(.)$ and $\phi(.)$ are continuous bounded functions $[0, 1] \to \mathbf{R}$.

Assumption 1. Prior users have higher efficacy beliefs, such that $\xi_A(h) = \xi_B(h) + q$ for all h. Let $\alpha \equiv \xi_B(h)$.

Assumption 2. Vulnerable types have higher health benefit than non-vulnerable types, such that $F_V(h) < F_{NV}(h)$ for all h, where $F(h)$ is the cumulative distribution of h.

There will be an equilibrium cutoff level of recipient need α_A^* for which prior user senders share with recipients with this α or greater and non-user senders share when $\alpha > \alpha_A^* + q$. The equilibrium is further characterized in the appendix and follows the results of Chandrasekhar et al., 2018.

In equilibrium prior users will be more likely to send compared to non-prior users, and likewise vulnerable types will be more likely to receive than non-vulnerable types. Therefore sending information signals likely "prior user" and/or "vulnerable". This signal is a social cost incurred from sending or receiving. Introducing a monetary incentive M for sending leads to a revised threshold level of need, α_A^{*M}. When $\alpha_A^{*M} < \alpha_A^*$, ie. cash leads to more sending, and this is observable, then there can be a secondary effect on the social signal of sending. By inducing users and non-users to share, the incentive can reduce the proportion of prior users among the people who share. It then becomes less "telling" that someone who shares is a prior user. The same mechanisms lead those who receive when there is a known financial incentive to be less vulnerable in expectation than those who receive when there is no incentive. These effects decreases the cost of sending, specifically via a social image channel. (See appendix for

further details.)

4.1 Identifying Social Image Effects

The social image constraint on information sharing can be identified by varying the observability of a constant financial incentive. The direct effect of a 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 financial incentive has a secondary effect on the social signal, and it varies with its observability.

Let Let $R(S) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS})$ be the social image cost of sending. Define $R(S|M = 0)$ as the social image cost associated with sending in the equilibrium without financial incentives, and $R(S|M > 0)$ as the social image cost associated with sending in the equilibrium with an observable financial incentive M.

When sending is private, recipients cannot observe equilibrium sending behavior. Therefore recipients will form beliefs about the cutoff α_A^*, and sender and recipient types π_{post} and ω_{post}, based on the information they know about the sender's incentives. When the sender discloses that she is paid, the recipient will know the sender's full objective function and form beliefs about type that are consistent with the equilibrium in which there is a financial incentive, and the equilibrium cutoff accordingly reflects those beliefs:

$$\alpha_A^*(Disclose = 1) = R(S|M > 0) - q - M \tag{2}$$

When the sender does not disclose that she is paid, the recipient will form beliefs about type that are consistent with the equilibrium in which the sender is *not* paid. Therefore the equilibrium cutoff when the sender does not disclose payment reflects a social image cost $R(S|M = 0)$ as if there is no financial incentive, despite the existence of the private financial incentive M:

$$\alpha_A^*(Disclose = 0) = R(S|M = 0) - q - M \tag{3}$$

Comparing the cutoffs shows that when the financial incentive is constant across information conditions, the cutoff with disclosure will be different from the cutoff without disclosure if and only if social image concerns receive non-zero weight in the sender's objective function and the existence of a known financial incentive affects inferences about type (captured by $R(.)$).

$$\alpha_A^*(Disclose = 1) - \alpha_A^*(Disclose = 0) = R(S|M > 0) - R(S|M = 0) \tag{4}$$

The model prediction maps to the design to generate the following testable hypothesis:

H1: If image concerns are binding then senders nudged to disclose their financial compensation will be more likely to share.

A second test for whether social image concerns are binding is to vary how informative, ie. revealing, the information that is shared is. For example telling people that information is being targeted based on need will increase ω^S. Let $\widetilde{\omega}^S > \omega^S$ be the information bundled with targeting messaging and accordingly define the corresponding cutoffs for sending:

$$\alpha_A^*(targeted) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\widetilde{\omega}^S) - \phi(\omega^{NS}) - q \tag{5}$$

$$\alpha_A^*(non-targeted) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q \tag{6}$$

$$\tag{7}$$

Then targeted and non-targeted cutoffs $\alpha_A^*(targeted) - \alpha_A^*(non-targeted)$ will be different if and only if the recipient's reputation receives non-zero weight in the sender's objective function.

This model predictions maps to the "targeted" framing in the design and forms the testable hypothesis:

H2: If senders assign positive weight to their friend's image then they will be less likely to share messages with a "targeted" framing that emphasizes the recipient's need.

Finally, the sender's utility function is characterized assuming the sender weights health benefit to the recipient against image costs to the sender and recipient. This forms the testable prediction that sharing rates will be increasing in recipient need and decreasing in social image cost. More specifically:

H3: If senders trade-off health benefit to recipients with image costs then sharing rates will be increasing in recipient need and decreasing when senders do not disclose compensation or do "target" the recipient.

5 Main Results

In the main analysis I first establish individuals' private knowledge of their friends' mental health status. Second I test how varying the perceived social image concerns of sharing stigmatized information lead senders to be more or less willing to share information with their friends, and explore mechanisms using heterogeneity analysis. Third I investigate 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 the helpline services and other related outcomes.

5.1 Sender Knowledge

A key rationale for involving community members in outreach efforts for various programs is that these individuals have private knowledge of who will benefit most from programs. This is particularly valuable when the factors determining an individual's fit for a program are not easily observable, as in the case of mental health.

I first asked senders to rank their friends in order of who would most benefit from learning about mental health resources.[19] I next surveyed the sender's friends (the recipients) within the same week and collected the PHQ-9 depression questionnaire and the GAD-2 screening for anxiety. Enumerators were trained to not reveal to the recipients that they had been referred by a friend nor that the survey was the baseline for a mental health awareness campaign. Column 1 of Table 1 regresses a recipient-level indicator for being the person who would benefit the most on an indicator for whether the recipient has depression or anxiety at baseline.[20] This predictive analysis reveals that the friend identified to be more in need is 11 percentage points, or 25% more likely to have depression. In column 2 I add demographic controls selected using lasso double selection. The results suggest that senders have information above and beyond observable demographic characteristics. Including the demographic controls reduces the coefficient on the sender's ranking by less than 2 percentage points. In columns 3 and 4 I replicate the analysis but instead study whether the ranking predicts anxiety. I find the sender's selection of who is most in need is only marginally significantly associated with anxiety, and not significantly so when controlling for recipient demographics although similar in magnitude.[21]

5.2 Experimental Effect of Message Framing on Sending Rates

The effect of each framing on sending rates is estimated using first a full flexible specification and then a specification in which the disclosure arms are pooled:

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

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

[19] The list of friends was established before the ranking question.

[20] "Has depression" or "has anxiety" is used as a shorthand to mean that the person likely has depression or anxiety based on the standard scoring cut-offs for the PHQ-9 and GAD-2.

[21] 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.

p_r is a binary indicator of whether the recipient received a message. F_{1s} is an indicator for the "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 "compensation, targeted" framing (Framing 2 :"An NGO is compensating me to share this information with my friends who I feel will benefit from the information.") F_{3s} is an indicator for the "non-compensation, 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. The three framings were assigned mutually exclusively over the sender sample and together comprise the complete treatment group.[22] X'_s and X'_r are covariate vectors for the sender and recipient respectively and Γ 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 pre-specified.[23] 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 that the sender level (ε_s).

Despite senders agreeing at baseline to share the campaign and knowing their friends are in need, most recipients never receive the campaign. As shown in Table 2 only 22% of recipients in the treatment group ever receive the campaign. In column 2 we see that 16% of senders' links were clicked on. Effort, attention and financial cost are unlikely to fully explain this. Recipients were by definition people with whom the sender communicates frequently; senders were reminded repeatedly to share the campaign, and the financial incentive for participation exceeded the phone credit costs of sharing the campaign. I turn to the experimental variation in message framing to test whether image concerns can explain the low sending rates.

Table 3 presents the main result of the effect on sharing rates of framings that increase or decrease social image concerns. Column 1 shows the rates of sending in each of the three framing arms. We see that, consistent with the hypothesis that disclosing compensation will increase willingness to share, the two "disclosed compensation" arms (non-targeted and targeted), are each shared 5.6 percentage points (34%) more than than the effect of "non-disclosed compensation, non-targeted" framing (which had a 16.1 percentage point effect on sharing) though this disaggregated result is statistically insignificant at traditional significance levels (p-values 0.113 and 0.109). Contrary to the hypothesis that senders would withhold messages that strongly signalled that the recipient is in need, the difference between the "targeted" and

[22] 5% of the recipient sample individuals appeared in more than 1 friend group, and so could have receive multiple treatments which were assigned randomly and independently. The results are robust to dropping these recipients with degree greater than 1.

[23] In the pre-specified analysis the treatment indicators as defined did not align with the final implemented design, and therefore the first specification to be tested here is the fully flexible specification.

"non-targeted" framings is less than 0.01 percentage points.[24] Because I measure no meaningful difference between the targeted and non-targeted framings, I combine the two into one pooled "disclosed compensation" framing. In column 2 we see that disclosure led to a 6.2 percentage point (37%) increase in willingness to share compared to the non-disclosed compensation framing (p-value 0.038).[25] Columns 3 and 4 replicate the analysis at the sender level using link click data.[26] There the effect of disclosure is significant when fully disaggregating the framing arms. Comparing the "disclosed compensation, non-targeted" and "non-disclosed compensation, non-targeted" framings shows a 7 percentage point (57%) increase in clicks when compensation is disclosured (p-value 0.048). The result is similar and more significant when again pooling the disclosure framings (6.5 percentage point difference, p-value 0.029).[27]

5.3 Drivers of senders' responsiveness to disclosure

Heterogeneity analysis allows us to dig further whether the increase in sharing when nudged to disclose compensation is related to social image. The effect of disclosure does not vary with any of the 6 dimensions of pre-specified heterogeneity. I turn then to machine learning heterogeneity which reveals that whether the sender is prior user most differentiates responsiveness to the disclosure treatment. I follow Chernozhukov et al., 2018 to first test whether in fact the disclosure effect is heterogeneous, and if yes then what characteristics are associated with the variation in treatment effects. Figure 4 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.[28] One characteristic dominates any other, which is whether the sender is a prior user of mental health services herself. Table 4 shows that senders who are prior users of mental health services are 25 percentage points (133%) more likely to send the disclosed compensation framing than those who have not used mental health services before. In the non-disclosed compensation framing senders who are prior users are no more likely to

[24] The comparison is conducted holding constant "disclosed compensation".

[25] The non-disclosed compenation framing is non-targeted

[26] This measure may combine both sending rates and recipient engagement. The appendix shows that the patter of results is unchanged when restricting to instances in which clicks occurred on more than 1 unique device.

[27] Control group senders never received the campaign from the study and have no links that could be clicked. The control mean of 1.2% receipt of the campaign may be due to spillovers (which seems unlikely given the overall low rates of sharing) or misreporting by recipients in the midline or endline surveys.

[28] I restrict to characteristics that vary significantly between the first and fifth quantiles at the 99% confidence level, and then order characteristics by the magnitude of the difference.

send messages. Prior users' high sensitivity to the excuse provided by disclosing compensation offers strong suggestive evidence that disclosure operates through the social signaling mechanism described in the model.

5.4 Do senders target those most in need?

The interaction of recipient need with message framing helps further characterize senders' behavior. The model assumes that senders' behavior is the result of a tradeoff between social image costs of sending information and potential health benefits to recipients. Table 5 and Figure 6 provides evidence consistent with this model. Column 1 establishes that senders do not on average target high need recipients. But, columns 2 and 3 show that patterns of targeting depend on whether the sender has social cover. Senders nudged to use the excuse of being compensated send to high and low need recipients at similar rates. But when senders are not nudged to use this excuse they more strictly target the few messages they are willing to send to high-need recipients. Focusing first on recipients who do not have depression or anxiety at baseline we see that, as in the main results, senders in the non-disclosed compensation framing send significantly less (7.4 percentage point difference, p-value 0.087). But senders in the non-disclosed compensation arm significantly increase their willingness to send to recipients with depression or anxiety at baseline (9.1 percentage point difference, p-value 0.059).

5.5 Recipient Impacts

I first investigate the effect of the receiving the campaign on recipient outcomes, regardless of the framing used. To analyze the recipient impacts I estimate the effect of treatment on the treated using a two stage least squares regression model.[29]

$$T_r = \gamma_0 + \gamma_1 A_s^T + \psi_s A_s^T X_s + X_r' \lambda_1 + X_s' \lambda_2 + \Gamma + \nu_r \tag{10}$$

$$y_r = \pi_0 + \pi_1 \widehat{T_r} + + X_r' \phi 1 + X_s' \phi_2 + \Gamma + \eta_r \tag{11}$$

where T_r is an indicator taking 1 if the recipient's sender shared any campaign messages and A_s^T is an indicator taking 1 if the recipient's sender was assigned to treatment. $A_s^T X_s$ is a vector of interactions between the sender's treatment assignment and sender's baseline characteristics from which predictors are chosen by lasso to predict exogenous recipient of the campaign. As

[29]Estimating recipient effects with the IV approach was pre-specified. But to increase the power of the instrument I deviate from the pre-specified instrument of only the treatment indicator and instead also include interactions of sender covariates with the treatment indicator. I use lasso to select which of these interactions to include as instruments.

before X_s', X_r', and Γ are a vector of sender covariates, a vector of recipient covariates, and survey week fixed effects, respectively. The specific covariates and fixed effects included in each estimation are selected using the lasso double-selection procedure.

I additionally consider whether the message framing arms caused differential demand for the helpline. Because rates of sender follow-through are significantly different across arms, I focus on the intention to treat specification:

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

Estimated impacts on recipients are in general noisy because treatment only induced a 22 percentage point increase in the probability that recipients received the campaign. However Table 6 shows that even small impacts on pro-active take-up of the advertised helpline can be rejected. The point estimate on calling the helpline is -0.012 (se. 0.019). In addition to the self-reported helpline take-up measure, I check whether the phone numbers of study participants appear in the administrative in-take survey completed by the helpline with new callers during the study period. I find no instances of matching phone numbers. When female participants are asked directly 6 months later if they want the helpline to contact them, there is no significant effect of pooled treatment. The point estimate is negative, at -0.111 (se. 0.108).

Table 7 tests for impacts on recipients' mental health attitudes and shows that in little evidence of impacts on attitudes. Columns 1-3 show that there are no significant impacts on expected benefit from mental health services, own stigma perceptions, nor second order stigma beliefs, though moderate effects on these outcomes cannot be rejected due to low power. Column 4 shows that whether participants had held any conversations about mental health in the past 6 months increased by 15.3 percentage points (p-value = 0.010).[30] This constituted a 138% increase in the probability of having these conversations, highlighting how rare opportunities are for knowledge exchange in this setting, and the potential for the peer-led communication campagin to spark additional conversations.

Table 8 shows that the campaign lead to substantial activation of social and economic support. I find that treated recipients experience a 0.32 standard deviation increase in their social connectedness (p-value = 0.037). Columns 2 and 3 show that treated recipients engage in mutual labor assistance 0.41 standard deviations more (p-value = 0.015) and borrowed 19 percentage points more (p-value = 0.022) while there was no effect on lending.[31]

It may be that different framings of information affect recipients' demand for the helpline. For

[30] This included only phone or face to face conversations, not messages such as the campaign itself.

[31] Co-authors and I document in Stillman et al., 2022 that the vast marjority of loans in this setting are informal.

example the disclosed compensation framing could lead the recipient to devalue the information, or infer that stigma is high, which could both decrease demand for the service. The targeted framing could help recipients learn about their fit for the program, or could lead to backlash if the recipient feels singled out. I investigate these impacts using the follow-up experiment, which, unlike the main experiment, enables me to measure recipients' reactions to message framings that are assigned randomly rather than with selection. Table 9 presents the effects of exogenously implemented introductions to the helpline on recipients' demand. Recall that in the follow-up experiment focus respondents gave the study permission to contact their friends and introduce the helpline using any of four introductions, using the focus respondent's name. Because there is no pure control group I compare rates of accepting the helpline across framing arms. I analyze the results using two pre-specified specification: a fully flexible specification in which each of the four framings are entered as separate treatments, and a specification that enters the framings as two cross-randomized treatments:

$$\begin{aligned} p_r =& \alpha_0 + \alpha_1 DiscloseXNontarget_r + \alpha_2 DiscloseXTarget_r \\ &+ \alpha_3 NonDiscloseXTarget_r + X_s'\phi_1 + X_r'\phi_2 + \Gamma + \varepsilon_r \end{aligned} \tag{13}$$

$$p_r = \beta_0 + \beta_1 Disclose_r + \beta_2 Target_r + \delta DiscloseXTarget_r + X_s'\phi_1 + X_r'\phi_2 + \Gamma + \varepsilon_r \tag{14}$$

In Table 9, Column 1 reports the specification in Equation (13) and Column 2 reports the specification in Equation (14). The second specification can be obtained by calculating a difference in difference from the results of the first specification. Column 1 shows that the only significant pairwise comparison of treatment framings is the comparison of disclosed and targeted versus disclosed and non-targeted, which has a marginally significant point estimate of 4.9 percentage points (p-value 0.090). Column 2 reports the specification described in Equation 14, which uses pooled versions of the disclosed and targeted treatments, and includes an interaction of the two treatments. Column 2 shows that while there is no significant effect of neither disclosure nor "targeted" phrasing, there is a negative interaction of the two which results in a 14.1 percentage point decrease in willingness to accept the helpline relative to when the respondent does not know the sender is paid and was not read "targeted" phrasing. These results first promisingly indicate that compensation can be disclosed without recipients disregarding the information shared. But second the results show that disclosing compensation interacts negatively with targeted phrasing, and therefore these approaches should likely not be used together.

6 Conclusion

In this paper I consider whether friends withhold potentially useful but stigmatizing information due to social image concerns for themselves and the recipient. I first show that peers are well positioned to facilitate social learning, as they have private knowledge of who is most in need. I then use information sharing experiments within friend groups to test whether stigma in fact leads friends to withhold potentially helpful information about mental health services. I find that friends do withhold information and these concerns are driven by concern for the sender's own social image rather than concern for revealing the recipient's vulnerability. Nudging senders to disclose that they are being financially compensated increases the chance that intended recipients get the information by 37%. This responsiveness to having an excuse for sharing information is consistent with senders demanding social cover when sharing stigmatized information. That prior users of mental health services are most responsive to the suggestion to disclose is consistent with senders worrying that others will infer information about them if they share information about stigmatized services. In a follow up experiment I investigate whether senders' disclosure may lead recipients to de-value the information, and promisingly find that the excuse of being paid can be used without undermining recipients' interest in the advertised services. Taking these results together suggests that a policy of encouraging senders to say that they are paid may help to increase sharing in settings with high stigma, without reducing message recipients' responsiveness to the information.

References

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*.

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., 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.

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

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

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.

Buchmann, N., Field, E. M., Glennerster, R., Nazneen, S., & Wang, X. Y. (2021). *A signal to end child marriage: Theory and experimental evidence from bangladesh* (tech. rep.). National Bureau of Economic Research.

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

Bursztyn, L., Egorov, G., & Jensen, R. (2019). Cool to be smart or smart to be cool? understanding peer pressure in education. *The Review of Economic Studies*, *86*(4), 1487–1526.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, *110*(10), 2997–3029.

Bursztyn, L., & Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, *9*, 131–153.

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

Celhay, P. A., Meyer, B. D., & Mittag, N. (2022). *Stigma in welfare programs* (tech. rep.). National Bureau of Economic Research.

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.

Currie, J., Grogger, J., Burtless, G., & Schoeni, R. F. (2001). Explaining recent declines in food stamp program participation [with comments]. *Brookings-Wharton papers on urban affairs*, 203–244.

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.

Gesesew, H. A., Tesfay Gebremedhin, A., Demissie, T. D., Kerie, M. W., Sudhakar, M., & Mwanri, L. (2017). Significant association between perceived hiv related stigma and late presentation for hiv/aids care in low and middle-income countries: A systematic review and meta-analysis. *PLoS One*, *12*(3), e0173928.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of economic perspectives*, *25*(4), 191–210.

Goldberg, J., Macis, M., & Chintagunta, P. (2018). *Incentivized peer referrals for tuberculosis screening: Evidence from india* (tech. rep.). National Bureau of Economic Research.

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.

Henderson, C., & Thornicroft, G. (2009). Stigma and discrimination in mental illness: Time to change. *The Lancet*, *373*(9679), 1928–1930.

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.

Hussam, R., Rigol, N., & Roth, B. N. (2022). Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review*, *112*(3), 861–898.

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*.

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

Lasky-Fink, J., & Linos, E. (2022). It's not your fault: Reducing stigma increases take-up of government programs.

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.

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].

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, *370*(6522), eaay0214.

Schanzenbach, D. (2009). *Experimental estimates of the barriers to food stamp enrollment*. Institute for Research on Poverty, University of Wisconsin-Madison.

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.

Suomi, A., Schofield, T., Haslam, N., & Butterworth, P. (2022). Is unemployment benefit stigma related to poverty, payment receipt, or lack of employment? a vignette experiment about australian views. *Analyses of Social Issues and Public Policy*, *22*(2), 694–711.

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.

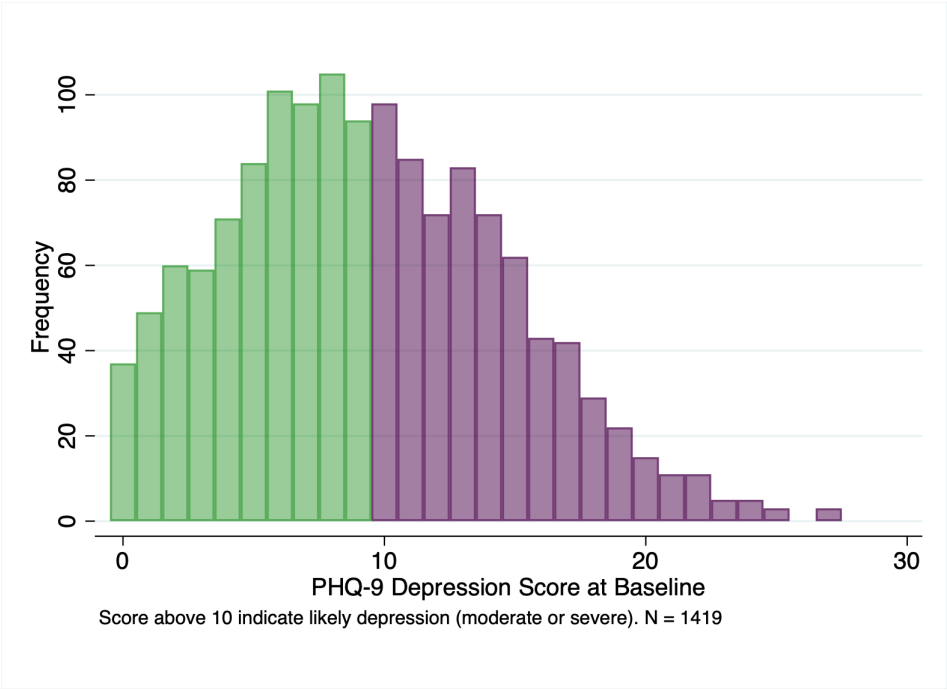
WHO, W. M. H. S. C., et al. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *Jama*, *291*(21), 2581–2590.

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.

7 Tables and Figures

Figure 2: Prevalence of Depression among Recipients at Baseline



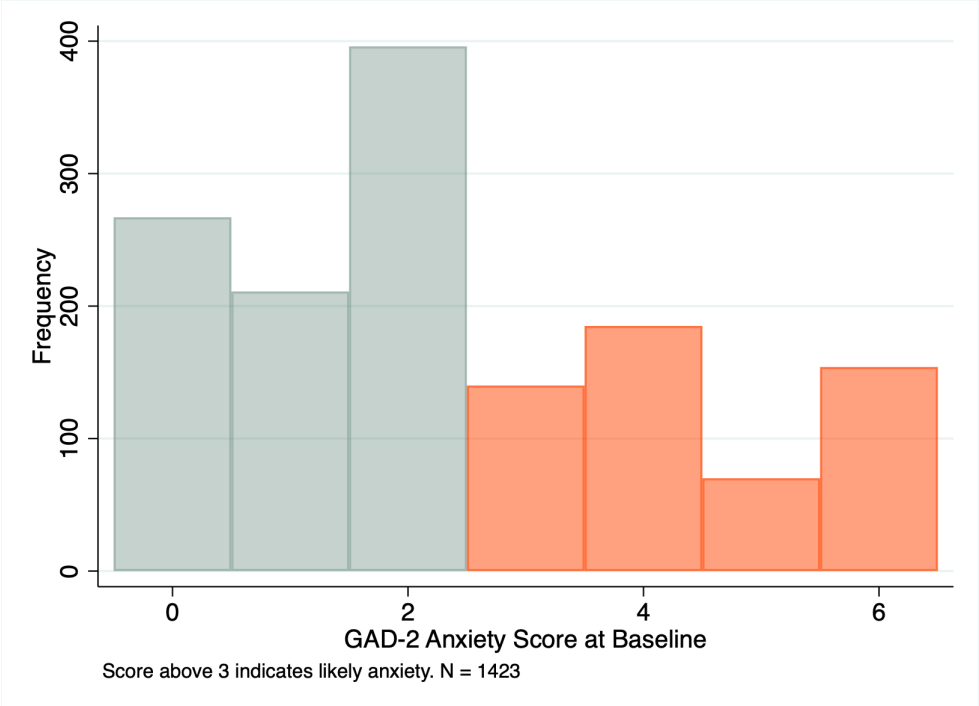


Figure 3: Prevalence of Likely Anxiety Among Recipients at Baseline

Table 1: Sender ability to target

	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	0.110***	0.096***	0.052*	0.043
	(0.029)	(0.028)	(0.028)	(0.028)
Control Mean	0.435	0.435	0.374	0.374
Covariates	No controls	Demographics Selected by lasso	No controls	Demographics Selected by lasso
N	1326	1326	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 is an imprecise 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 1B. Heterogeneity in Sender ability to target

	(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	0.111***	0.109***	0.109***	0.108***	0.107***	0.112***
	(0.030)	(0.029)	(0.030)	(0.038)	(0.029)	(0.039)
Highest Need X Sender stigma 1st order	0.038					
	(0.029)					
Highest Need X Sender stigma 2nd order		-0.004				
		(0.029)				
Highest Need X Sender altruism			0.008			
			(0.031)			
Highest Need X Sender female				0.000		
				(0.059)		
Highest Need X Sender social desirability					0.028	
					(0.029)	
Highest Need X Sender depressed						-0.007
						(0.058)
Control Mean	0.435	0.435	0.435	0.435	0.435	0.435
Covariates	No controls	No controls	No controls	No controls	No controls	No controls
N	1326	1326	1326	1326	1326	1325

This table shows that there is no significant heterogeneity by sender characteristics in senders' ability to identify which of their friends is 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 2: Sender Compliance

	(1) Received campaign (recipient-level)	(2) Any clicks (0/1) (sender-level)
Treatment (sender asked to share)	0.210***	0.162***
	(0.015)	(0.015)
Control Mean	0.012	–
Covariates	Lasso Double Selection	Lasso Double Selection
N	2665	847

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, *non-targeted*	0.217***		0.191***	
	(0.025)		(0.028)	
Disclosed Compensation , *targeted*	0.217***		0.173***	
	(0.026)		(0.028)	
Non-Disclosed Compensation framing, *non-targeted*	0.161***	0.165***	0.121***	0.123***
	(0.024)	(0.024)	(0.025)	(0.025)
Disclosed Compensation framing, *pooled*		0.227***		0.188***
		(0.019)		(0.020)
p-values				
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.113]		[.048]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.109]		[.133]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.997]		[.634]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.038]		[.029]
Control Mean	0.012	0.012	–	–
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2665	2665	847	847

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

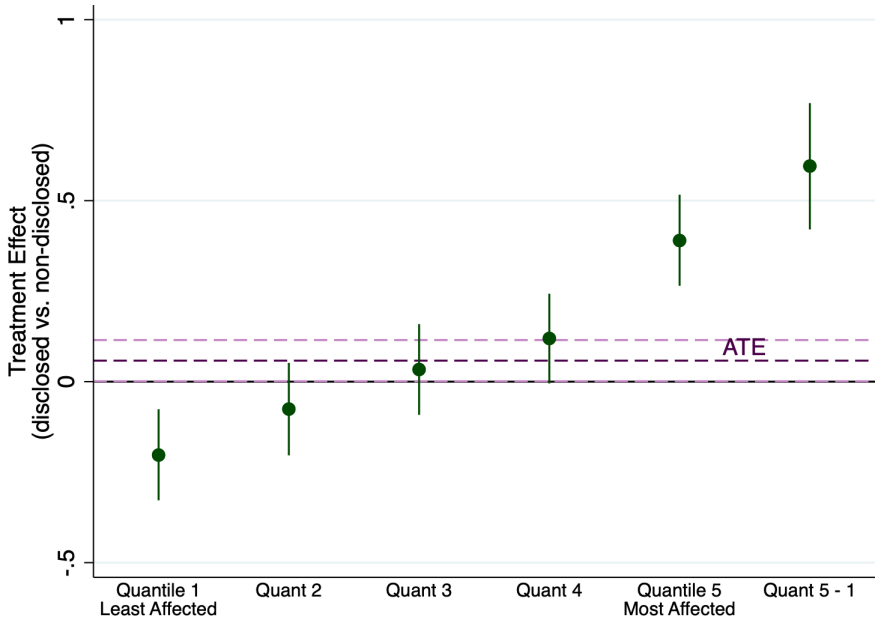


Figure 4: Sorted Group Average Treatment Effects (Effect of Disclosure)

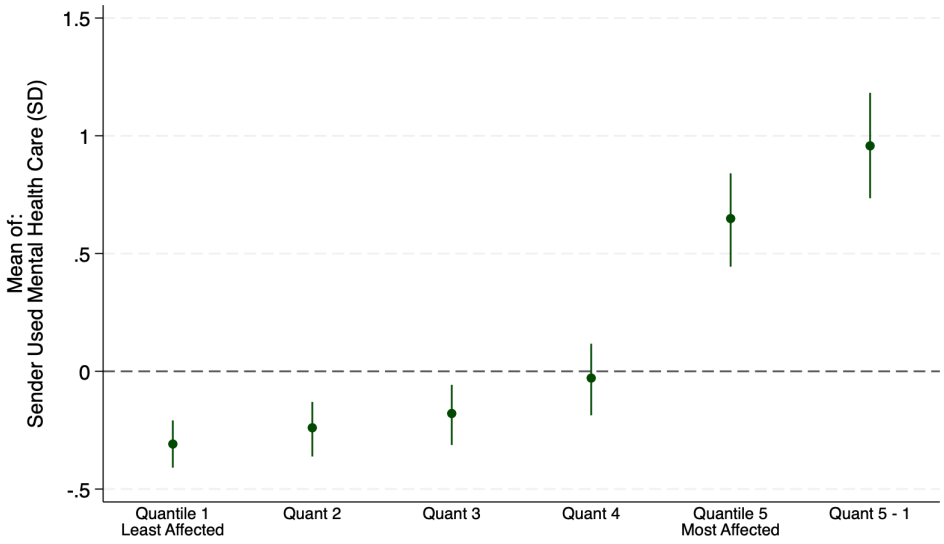


Figure 5: Mean of "Sender Used Mental Health Care" by Quantile of Impact of Disclosure

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 X			
Sender has used mental health services	0.227***		
	(0.045)		
Treatment			
(sender asked to share)	0.168***		
	(0.015)		
Sender has used mental health services			
X Disclosed Compensation, *non-targeted*		0.359***	
		(0.071)	
X Disclosed Compensation , *targeted*		0.111*	
		(0.065)	
X Non-Disclosed Compensation framing, *non-targeted*		0.002	0.018
		(0.079)	(0.074)
Sender used mental health services X			
Compensation framing, *pooled*			0.251***
			(0.052)
Disclosed Compensation, *non-targeted*		0.166***	
		(0.022)	
Disclosed Compensation , *targeted*		0.199***	
		(0.029)	
Non-Disclosed Compensation framing, *non-targeted*		0.179***	0.182***
		(0.030)	(0.029)
Disclosed Compensation framing, *pooled*			0.188***
			(0.019)
Sender has used mental health services	-0.019	0.024	0.009
	(0.015)	(0.029)	(0.022)
p-values for differences of means for non-users			
$\text{Disclosed}_{non-targeted}$ − $\text{Non-Disclosed}_{non-targeted}$		[.705]	
$\text{Disclosed}_{targeted}$ − $\text{Non-Disclosed}_{non-targeted}$		[.618]	
$\text{Disclosed}_{targeted}$ − $\text{Disclosed}_{non-targeted}$		[.345]	
$\text{Disclosed}_{pooled}$ − $\text{Non-Disclosed}_{non-targeted}$			[.855]
Control Mean			
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2665	2665	2665

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. 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: Interaction of Recipient Need and Message Framing

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)	(3) Received campaign (recipient-level)
Treatment X			
Recipient. Depression/Anxiety (0/1)	0.007		
	(0.029)		
Treatment			
(sender asked to share)	0.223***		
	(0.023)		
Recipient Depressed/Anxious			
X Disclosed Compensation, *non-targeted*		0.016	
		(0.047)	
X Disclosed Compensation , *targeted*		0.021	
		(0.048)	
X Non-Disclosed Compensation framing, *non-targeted*		0.093*	0.091*
		(0.048)	(0.048)
X Disclosed Compensation framing, *pooled*			0.010
			(0.036)
Disclosed Compensation, *non-targeted*		0.215***	
		(0.037)	
Disclosed Compensation , *targeted*		0.226***	
		(0.038)	
Non-Disclosed Compensation framing, *non-targeted*		0.158***	0.155***
		(0.037)	(0.037)
Disclosed Compensation framing, *pooled*			0.229***
			(0.028)
Baseline Depression or Anxiety (0/1)	-0.002	-0.035	-0.024
	(0.017)	(0.021)	(0.020)
p-values for differences of means for non-depressed recipients			
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$		[.248]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$		[.178]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$		[.841]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$			[.087]
Control Mean			
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	1423	1422	1422

This table shows how the interaction of recipient mental health need with assignment to treatment and message framing impacts the sending decision. The dependent variable in columns 1-3 is an indicator for whether the given recipient received any campaign message. 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. Column 1 shows that on average senders did not target on recipient need. Column 2 tests for targeting by recipient across the three framing arms. We see that the "non-compensation, non-targeted" framing was targeted to more depressed or anxious recipients, and this targeting closes the gap in sending between the compensation and non-compensation groups. Column 3 repeats this analysis when pooling the two compensation framings. The second panel shows the differences in rates of sharing between framing arms for the base group, which is those without depression or anxiety. P-values for the difference in means are reported in brackets in the bottom panel. Robust standard errors clustered at the sender level. The sample includes only recipients in the subsample reached to be surveyed at baseline. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 6: Heterogeneous Framing Effects by Recipient Mental Health Need

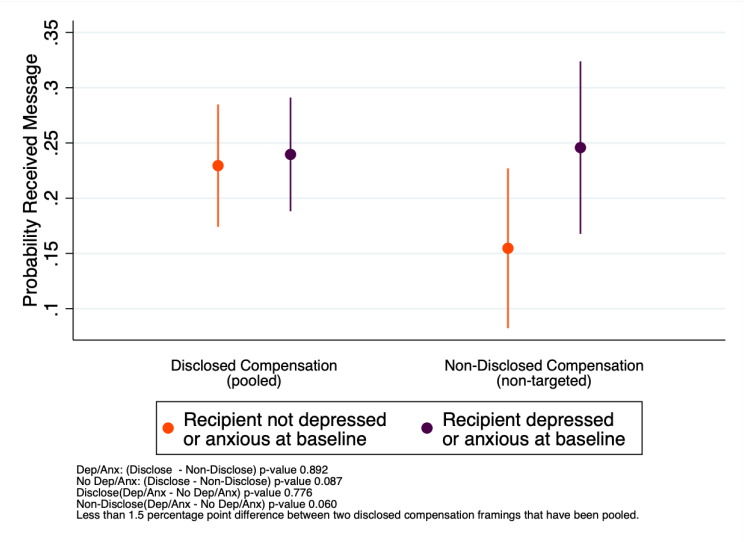


Table 6: IV Estimates of Impact on Recipient Demand for Mental Health Services

	(1) Called Helpline	(2) Willing to accept call from helpline
IV		
Sender Participated (sent to anyone)	-0.011	-0.017
	(0.019)	(0.101)
FDR-adjusted q-value	1	1
Control Mean	0.016	0.541
Double selection	Yes	Yes
N	1021	794

Expected benefit from mental health care is a standardized index of the respondent's expectation on a scale of 1 to 10 of what her psychological distress level would be if she began using mental health services regularly, minus her assessment on the same scale of her current level of psychological distress. Own stigma is a standardized index comprising the respondent's discriminatory attitudes toward mental health careseekers in the marriage market and the labor market. The second order stigma beliefs outcome is a standardized index comprising the respondent's beliefs about others' discriminatory attitudes toward mental health care-seekers in the marriage market and labor market. The dependent variable in column 5 is an indicator of whether the respondent had conversations about mental health with anyone outside her household in the past 6 months. 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: IV Estimates of Impacts on Recipient Mental Health Attitudes

	(1) Expected Benefit from MH Care (SD)	(2) Own Stigma Index (SD)	(3) Stigma Beliefs 2nd Order (SD)	(4) Any conversations about mental health
IV				
Sender participated (sent to anyone)	0.182	-0.013	-0.091	0.153***
	(0.148)	(0.146)	(0.150)	(0.059)
FDR-adjusted q-value	.973	1	1	.046
Control Mean	0.009	0.001	-0.020	0.111
Double selection	Yes	Yes	Yes	Yes
N	1015	1022	1022	1019

Expected benefit from mental health care is a standardized index of the respondent's expectation on a scale of 1 to 10 of what her psychological distress level would be if she began using mental health services regularly, minus her assessment on the same scale of her current level of psychological distress. Own stigma is a standardized index comprising the respondent's discriminatory attitudes toward mental health careseekers in the marriage market and the labor market. The second order stigma beliefs outcome is a standardized index comprising the respondent's beliefs about others' discriminatory attitudes toward mental health care-seekers in the marriage market and labor market. The dependent variable in column 5 is an indicator of whether the respondent had conversations about mental health with anyone outside her household in the past 6 months. 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 8: IV Estimates of Impacts on Recipient Social and Economic Support

	(1) Social Connectedness Index (SD)	(2) Labor Assistance Freq. (SD)	(3) Borrowed (0/1)	(4) Lent (0/1)
IV				
Sender participated (sent to anyone)	0.317**	0.413**	0.193**	0.002
	(0.152)	(0.169)	(0.084)	(0.040)
FDR-adjusted q-value	.052	.046	.124	1
Control Mean	-0.005	-0.005	0.594	0.086
Double selection	Yes	Yes	Yes	Yes
N	1022	1022	1021	1021

The dependent variable in column 1 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 2 is a standardized index of the number of times 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 dependent variable in column 3 is an indicator variable for whether the repondent has borrowed from anyone househod the household in the past 30 days. The dependend variable in column 4 is an indicator for whether the respondent has lent to anyone outside the household in the past 30 days. 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 9: Experiment 2: Impact of Exogenous Messaging on Demand for Phone Counseling

	(1) Willing to accept call from helpline (0/1)	(2) Willing to accept call from helpline (0/1)
Disclosed Compensation, *non-targeted*	0.049	
	(0.052)	
Disclosed Compensation , *targeted*	-0.033	
	(0.055)	
Non-Disclosed Compensation framing, *targeted*	0.053	
	(0.049)	
Disclosed Compensation framing, *pooled*		0.048
		(0.051)
Targeted framing, *pooled*		0.053
		(0.048)
Compensation X Targeted framing		-0.141**
		(0.071)
Reference category mean:		
Non-Disclosed compensation, *non-targeted*	0.685	0.685
p-value		
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	0.120	
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{targeted}$	0.940	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{targeted}$	0.090	
Double selection	Yes	Yes
N	652	652

This table shows that when message framings are implemented exogenously in experiment 2, the "compensation + targeted" framing causes a significant decrease in demand for the helpline which parallels the impacts on recipients in experiment 1. The compensation framing without the targeted framing does not decrease take-up. The dependent variable is the willingness of the new recipients to be contacted by the helpline to receive phone counseling. Robust standard errors clustered at the original recipient level. Covariates are selected using lasso double-selection from a list of sender and recipient covariates following Belloni et al. 2014. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Appendix

Appendix A. Attrition

Table 10: 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)	-0.022			-0.023		
	(0.027)			(0.027)		
Disclosed Compensation, *non-targeted*		-0.031			-0.021	
		(0.029)			(0.028)	
Disclosed Compensation , *targeted*		-0.036			-0.044	
		(0.030)			(0.030)	
Non-Disclosed Compensation framing, *non-targeted*		-0.036	-0.033		-0.030	-0.029
		(0.030)	(0.031)		(0.029)	(0.030)
Disclosed Compensation framing, *pooled*			-0.030			-0.031
			(0.027)			(0.026)
F-Statistic	.71	.706	.789	.728	.792	.76
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	2665	2665	2665	2665	2665	2665

This table shows that there was no difference in the probability of being treated for recipients reached for the baseline and endline surveys. Standard errors are clustered at the sender level. * p < 0.10, ** p < 0.05, *** p < 0.01

Table 11: Baseline Recipient Attrition by Sender's Ranking of Recipient Need

	(1) Recipient Surveyed Baseline	(2) Recipient Surveyed Baseline
Ranked Recipient Need	-0.004	
	(0.007)	
Highest need recipient in friend group		0.012
		(0.021)
Control Mean	1	1
Covariates	Network Size	Network Size
N	2548	2548

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

Appendix B. Intervention Content

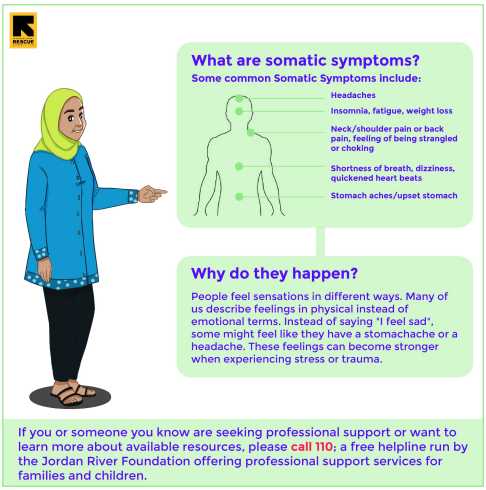


Figure 7: Example of Campaign Content

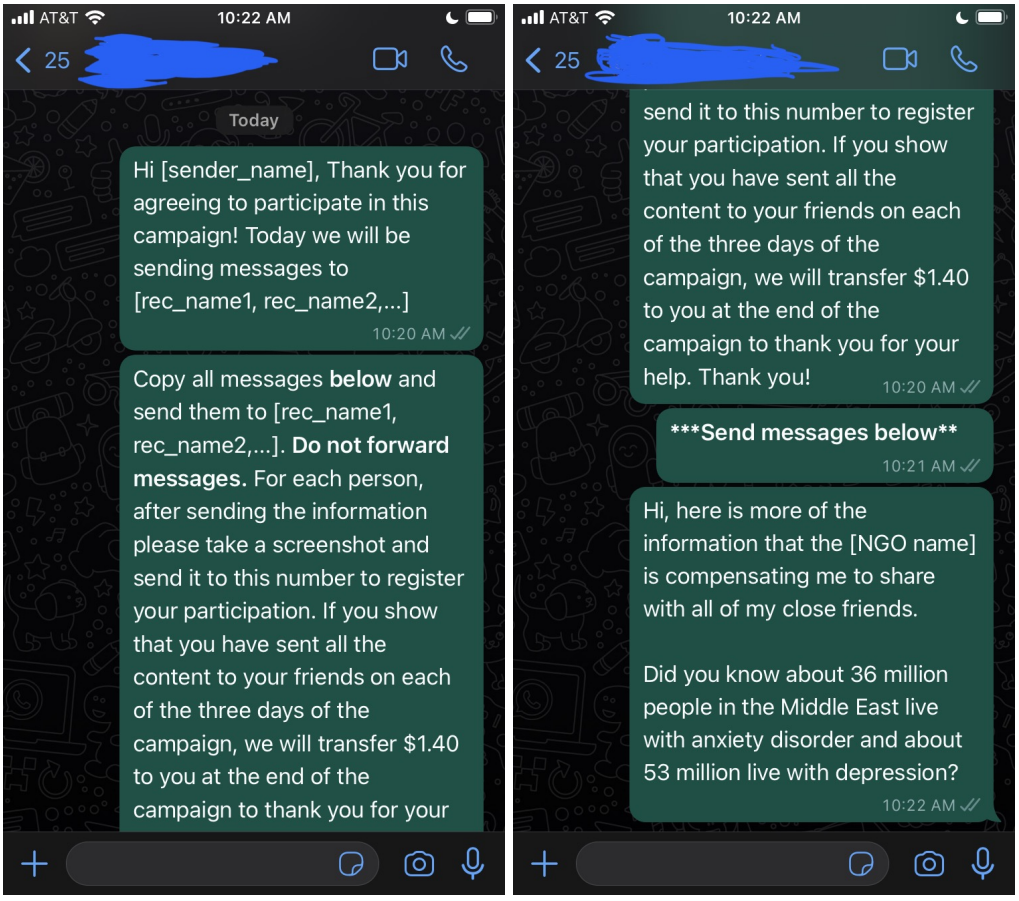


Figure 8: Example of Campaign Instructions

Appendix C.

In the sender population a fraction π are prior users of mental health services (type A) and the remainder are non-users (type B). There are two types of recipients, the vulnerable type V and the unvulnerable type UV, with ω representing the proportion of recipients who are the vulnerable type. 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.

$$U(S) = \underbrace{\xi(h\mathbf{1}_{s=1})}_{\substack{\text{Health benefit} \\ \text{to recipient} \\ \text{(sender's belief)}}} - \underbrace{\varphi(\pi_{post})}_{\substack{\text{social image cost} \\ \text{to sender}}} - \underbrace{\phi(\omega_{post})}_{\substack{\text{social image cost} \\ \text{to recipient}}} \tag{15}$$

where h is the recipient's health benefit from the information and $\xi(h)$ are the sender's beliefs about the benefit for the recipient where $\xi(.)$ is a monotonic function of h and h is distributed normally. $\pi_{post} = P(Type_s = A|S)$ is the posterior belief that the sender is a prior user given the observed sending decision. Likewise $\omega_{post} = P(Type_r = V|S)$ is the probability that the recipient is vulnerable given the sending decision. $\varphi(.)$ and $\phi(.)$ are continuous bounded functions $[0, 1] \to \mathbf{R}$.

There will be an equilibrium cutoff level of recipient need α_A^* for which the sender shares with recipients with this α or greater.

Assumption 1. Prior users have higher efficacy beliefs, such that $\xi_A(h) = \xi_B(h) + q$ for all h. Let $\alpha \equiv \xi_B(h)$.

Assumption 2. Vulnerable types have higher health benefit than non-vulnerable types, such that $F_V(h) < F_{NV}(h)$ for all h, where $F(h)$ is the cumulative distribution of h.

The equilibrium sharing rate is determined by a cutoff α^* at which the value of information to the recipient is at least as great as the social image losses incurred by sharing the information. The sender shares with all recipients for whom $\alpha \geq \alpha^*$, with α^* given by

$$\alpha_A^* = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q \tag{16}$$

where π^S is the posterior belief that the sender is a user if she sends, and π^{NS} is the posterior belief that she is user if she does not send.

Under the stated assumptions in equilibrium prior users will be more likely to send compared to non-prior users, and likewise vulnerable types will be more likely to receive than non-vulnerable types. Therefore sending information signals prior user and/or vulnerable. Holding constant the recipient's vulnerability, $\pi^S > \pi^{NS}$. And holding constant the sender's probability

of being a user, $\omega^S > \omega^{NS}$.

Introducing a monetary incentive M for sending leads to a revised cutoff characterized as

$$\alpha_A^* = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q - M \tag{17}$$

Let $R(S) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS})$ be the social image concern. Differentiating the cutoff α_A^* with respect to the monetary incentive M shows that if the change in the social image concern terms for a change in M ($\frac{\partial R(S)}{\partial M}$) is less than 1 than the the cutoff decreases in M, resulting in more sharing.

Letting there only be on type of recipient and focusing on the case of two types of senders, the posteriors are defined as:

$$\pi_{post}^S = \frac{(1 - F(\alpha_A^*))\pi}{(1 - F(\alpha_A^*))\pi + (1 - F(\alpha_A^* + q))(1 - \pi)} \tag{18}$$

$$\pi_{post}^{NS} = \frac{F(\alpha_A^*))\pi}{(F(\alpha_A^*))\pi + (F(\alpha_A^* + q))(1 - \pi)} \tag{19}$$

The equilibrium condition is

$$\alpha_A^* = \varphi(\pi_{post}^S) - \varphi(\pi_{post}^{NS}) - q - M \tag{20}$$

Noting that $\frac{\partial \pi_{post}^S}{\partial M} = \frac{\partial \pi_{post}^S}{\partial \alpha_A^*} \frac{\partial \alpha_A^*}{\partial M}$ and similarly for $\frac{\partial \pi_{post}^{NS}}{\partial M}$,

Taking the partial with respect to M gives

$$\frac{\partial \alpha_A^*}{\partial M} = -\left(1 + \frac{\partial \varphi(\pi_{post}^{NS})}{\partial \alpha_A^*} - \frac{\partial \varphi(\pi_{post}^S)}{\partial \alpha_A^*}\right)^{-1} \tag{21}$$

This shows that the cutoff is decreasing in the cash incentive if $\frac{\partial \varphi(\pi_{post}^S)}{\partial \alpha_A^*} - \frac{\partial \varphi(\pi_{post}^{NS})}{\partial \alpha_A^*} < 1$.

If, in the simplest case, φ is a constant, then we can focus just on $\frac{\partial \pi_{post}^S}{\partial \alpha_A^*}$ and $\frac{\partial \pi_{post}^{NS}}{\partial \alpha_A^*}$. Solving for the partial derivatives of the posteriors with respect to α_A^* yields

$$\frac{\partial \pi_{post}^S}{\partial \alpha_A^*} = \frac{-\pi f(\alpha)\left[((1 - F(\alpha))\pi + (1 - F(\alpha + q))(1 - \pi)\right] + (1 - F(\alpha))\pi\left[\pi f(\alpha) + f(\alpha + q)(1 - \pi)\right]}{\left[\pi(1 - F(\alpha)) + (1 - \pi)(1 - F(\alpha + q))\right]^2} \tag{22}$$

$$\frac{\partial \pi_{post}^{NS}}{\partial \alpha_A^*} = \frac{\pi f(\alpha)\left[F(\alpha)\pi + F(\alpha + q)(1 - \pi)\right] - F(\alpha)\pi\left[(\alpha) + f(\alpha + q)(1 - \pi)\right]}{\left[(\alpha) + (1 - \pi)F(\alpha + q)\right]^2} \tag{23}$$

Taking the difference $\frac{\partial \pi_{post}^S}{\partial \alpha_A^*} - \frac{\partial \pi_{post}^{NS}}{\partial \alpha_A^*}$ yields

$$\frac{\partial \pi^S_{post}}{\partial \alpha^*_T} - \frac{\partial \pi^{NS}_{post}}{\partial \alpha^*_T} \tag{24}$$

$$= (\pi - 1)\pi \left[\frac{(F(\alpha+q)-1)f(\alpha) - f(\alpha+q)F(\alpha) + f(\alpha+q)}{[(F(\alpha+q)-1) + \pi(1-F(\alpha+q)+F(\alpha)-1]^2} + \frac{f(\alpha)(-F(\alpha+q)) + f(\alpha+q)F(\alpha)}{(F(\alpha+q)-1+\pi(-F(\alpha+q)+F(\alpha))+1)^2} \right] \tag{25}$$

If the cutoff is decreasing in the monetary incentive, then a sufficient condition for the stigma concern to π^S_{post} to also be decreasing in the incentive is if probability distribution function $f(\alpha)$ is increasing (ie. $f(\alpha^*_A) < f(\alpha^*_A)$).

8.1 Identifying Social Image Effects

The social image constraint on information sharing can be identified by varying the observability of a constant financial incentive.

Define $R(S|M = 0)$ as the social image cost associated with sending in the equilibrium without financial incentives, where π^S and π^{NS} are beliefs about whether the sender is a user under "send" and "not send" respectively, and ω^S and ω^{NS} are beliefs about whether the recipient is the vulnerable type under "send" and "not send" respectively:

$$\begin{aligned} \alpha^*_A &= \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q \\ &= R(S|M=0) - q \end{aligned}$$

Define $R(S|M > 0)$ as the social image cost associated with sending in the equilibrium with an observable financial incentive M:

$$\begin{aligned} \alpha^*_A &= \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q - M \\ &= R(S|M>0) - q - M \end{aligned}$$

When sending is private recipients cannot observe equilibrium sending behavior. Therefore recipients will form beliefs about the cutoff α^*_A, and sender and recipient types π and ω, based on the information they know about the sender's objective function. When the sender discloses that she is paid, the recipient will know the sender's full objective function and form beliefs about type that are consistent with the equilibrium in which there is a financial incentive, and the equilibrium cutoff accordingly reflects those beliefs:

$$\alpha^*_A(Disclose = 1) = R(S|M>0) - q - M \tag{26}$$

When the sender does not disclose that she is paid, the recipient will form beliefs about type that are consistent with the equilibrium in which the sender is *not* paid. Therefore the equilibrium cutoff when the sender does not disclose payment reflects a social image cost with inaccurate beliefs, despite the existence of the private financial incentive M:

$$\alpha_A^*(Disclose = 0) = R(S|M = 0) - q - M \tag{27}$$

Comparing the cutoffs shows that when the private financial incentive is constant, the cutoff with disclosure will be different from the cutoff without disclosure if and only if social image concerns receive non-zero weight in the sender's objective function and the existence of a known financial incentive affects inferences about type.

$$\alpha_A^*(Disclose = 1) - \alpha_A^*(Disclose = 0) = R(S|M > 0) - R(S|M = 0) \tag{28}$$

A second test for whether social image concerns are binding is to vary how informative, ie. revealing, the information that is shared is. For example telling people that information is being targeted based on need will increase ω^S. Let $\widetilde{\omega}^S > \omega^S$ be the information bundled with targeting messaging and accordingly define the corresponding cutoffs for sending:

$$\alpha_A^*(targeted) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\widetilde{\omega}^S) - \phi(\omega^{NS}) - q \tag{29}$$

$$\alpha_A^*(non - targeted) = \varphi(\pi^S) - \varphi(\pi^{NS}) + \phi(\omega^S) - \phi(\omega^{NS}) - q \tag{30}$$

$$\tag{31}$$

Then targeted and non-targeted cutoffs $\alpha_A^*(targeted) - \alpha_A^*(non - targeted)$ will be different if and only if the recipient's reputation receives non-zero weight in the sender's objective function.

9 Appendix D. 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

10 Appendix E. Robustness

Table 12: 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, *non-targeted*	0.217***		0.191***		0.087***	
	(0.025)		(0.028)		(0.020)	
Disclosed Compensation , *targeted*	0.217***		0.173***		0.063***	
	(0.026)		(0.028)		(0.018)	
Non-Disclosed Compensation framing, *non-targeted*	0.161***	0.165***	0.121***	0.123***	0.040**	0.048***
	(0.024)	(0.024)	(0.025)	(0.025)	(0.016)	(0.016)
Disclosed Compensation framing, *pooled*		0.227***		0.188***		0.084***
		(0.019)		(0.020)		(0.014)
p-values						
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.113]		[.048]		[.058]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.109]		[.133]		[.323]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.997]		[.634]		[.389]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.038]		[.029]		[.074]
Control Mean	0.012	0.012	–	–	–	–
Covariates	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection	Lasso Double Selection
N	2665	2665	847	847	847	847

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 13: Robustness: Sender follow-through excluding recipients with duplicate treatments

	(1) Received campaign (recipient-level)	(2) Received campaign (recipient-level)
Disclosed Compensation, *non-targeted*	0.237***	
	(0.027)	
Disclosed Compensation , *targeted*	0.228***	
	(0.027)	
Disclosed Compensation framing, *pooled*		0.233***
		(0.019)
Non-Disclosed Compensation framing, *non-targeted*	0.162***	0.162***
	(0.024)	(0.024)
p-values		
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	[.037]	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	[.06]	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	[.812]	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		[.018]
Control Mean	0.012	0.010
Covariates	Lasso Double Selection	Lasso Double Selection
N	2546	2546

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 14: Robustness: Sender follow-through accounting for implementation imbalance and glitch

	(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, *non-targeted*	0.225***		0.211***		0.208***	
	(0.024)		(0.030)		(0.027)	
Disclosed Compensation , *targeted*	0.225***		0.184***		0.191***	
	(0.026)		(0.031)		(0.031)	
Disclosed Compensation framing, *pooled*		0.233***		0.202***		0.203***
		(0.018)		(0.022)		(0.021)
Non-Disclosed Compensation framing, *non-targeted*	0.172***	0.175***	0.142***	0.144***	0.150***	0.152***
	(0.026)	(0.025)	(0.025)	(0.025)	(0.028)	(0.027)
p-values						
$\text{Disclosed}_{non-targeted} - \text{Non-Disclosed}_{non-targeted}$	0.997		0.523		0.660	
$\text{Disclosed}_{targeted} - \text{Non-Disclosed}_{non-targeted}$	0.119		0.079		0.112	
$\text{Disclosed}_{targeted} - \text{Disclosed}_{non-targeted}$	0.127		0.280		0.302	
$\text{Disclosed}_{pooled} - \text{Non-Disclosed}_{non-targeted}$		0.051		0.076		0.122
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	2665	2665	1772	1772	1772	1772

This table tests alternate ways to address imbalance and an implementation glitch that caused a random subset of senders in Framing 3 to not receive one of the three batches of campaign content. In column 1 the specification controls for whether the sender experienced this glitch and baseline covariates that were by chance imbalanced but that are not selected using lasso. In column 2 the entire affected week is dropped, which leads to a large loss in power. In column 3 the affected week is dropped and baseline covariates that were by chance imbalanced but that are not selected by lasso for inclusion are included. Robust standard errors clustered at the sender level. The dependent variable in all columns of this table is an indicator for whether the recipient received the campaign. A recipient is recorded to have received the campaign if the sender shared a screenshot with the study documenting having shared any message, 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 15: Balance Checks: Sender Treatment vs. Control

	Control Mean/SD	Treatment Mean/SD	Dif. Means SE
Female	0.435	0.416	-0.019
	[0.035]	[0.019]	(0.040)
Age	39.396	40.520	1.124
	[0.753]	[0.450]	(0.877)
In camp	0.130	0.159	0.029
	[0.023]	[0.014]	(0.028)
Employed	0.353	0.355	0.002
	[0.033]	[0.019]	(0.038)
Jordanian	0.101	0.113	0.011
	[0.021]	[0.013]	(0.024)
Stigma (1st order)	-0.181	0.025	0.206***
	[0.062]	[0.040]	(0.074)
Stigma (2nd order)	0.019	-0.033	-0.052
	[0.069]	[0.039]	(0.079)
Dictator allocation	4.011	3.917	-0.094
	[0.223]	[0.124]	(0.255)
Social desirability	-0.018	-0.015	0.004
	[0.063]	[0.040]	(0.075)
PHQ-2	0.064	0.046	-0.018
	[0.067]	[0.040]	(0.078)
GAD-2	0.059	0.024	-0.035
	[0.068]	[0.040]	(0.079)
Depression/Anxiety likely	0.541	0.546	0.005
	[0.035]	[0.020]	(0.040)
Used MH care before	0.198	0.166	-0.032
	[0.028]	[0.015]	(0.031)
Social connectedness	-3.445	-3.443	0.003
	[0.045]	[0.028]	(0.053)
MH Efficacy beliefs	0.069	0.021	-0.048
	[0.065]	[0.038]	(0.075)
Network size	3.208	3.328	0.120
	[0.118]	[0.068]	(0.136)
F-Stat	0.0	0.0	0.0
N Observations	207	640	847

This table tests for balance on sender baseline covariates comparing the pooled treatment to control. Columns 1 and 2 show the means with standard deviations in brackets below. The third column shows the difference in the means, with the standard error in parentheses below. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Balance Checks: Recipient Treatment vs. Control

	Control Mean/SD	Treatment Mean/SD	Dif. Means SE
Female	0.464	0.447	-0.016
	[0.027]	[0.015]	(0.031)
Age	37.512	38.546	1.034
	[0.578]	[0.355]	(0.679)
In Camp	0.128	0.174	0.046 *
	[0.021]	[0.014]	(0.025)
Employed	0.364	0.358	-0.005
	[0.027]	[0.015]	(0.030)
Refugee	0.947	0.895	-0.051 **
	[0.014]	[0.011]	(0.018)
Stigma (1st order)	-0.097	0.023	0.120 *
	[0.055]	[0.031]	(0.063)
Stigma (2nd order)	0.041	-0.008	-0.049
	[0.053]	[0.031]	(0.061)
Social desirability	-0.080	0.024	0.104
	[0.053]	[0.031]	(0.061)
PHQ-9	9.562	9.342	-0.220
	[0.288]	[0.163]	(0.331)
GAD-2	2.485	2.399	-0.086
	[0.103]	[0.057]	(0.118)
Depression/Anxiety likely	0.567	0.536	-0.030
	[0.027]	[0.015]	(0.031)
Used MH care before	0.067	0.047	-0.020
	[0.014]	[0.006]	(0.015)
F-Stat	1.76	1.76	1.76
N Observations	341	1082	1423

This table tests for balance on recipient baseline covariates comparing the pooled treatment to control. Columns 1 and 2 show the means with standard deviations in brackets below. The third column shows the difference in the means, with the standard error in parentheses below. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Balance Checks: Sender Framings vs. Control

	Control Mean/SD	Framing 1 Mean/SD	Framing 2 Mean/SD	Framing 3 Mean/SD	$(\bar{x}_0 - \bar{x}_1)$ SE	$(\bar{x}_0 - \bar{x}_2)$ SE	$(\bar{x}_0 - \bar{x}_3)$ SE
Female	0.435	0.397	0.408	0.442	-0.038	-0.027	0.007
	[0.035]	[0.033]	[0.034]	[0.034]	(0.048)	(0.049)	(0.048)
Age	39.396	41.068	40.374	40.102	1.672	0.978	0.706
	[0.753]	[0.768]	[0.804]	[0.771]	(1.076)	(1.101)	(1.078)
In camp	0.130	0.155	0.136	0.186	0.025	0.005	0.056
	[0.023]	[0.025]	[0.024]	[0.027]	(0.034)	(0.034)	(0.035)
Employed	0.353	0.329	0.393	0.344	-0.024	0.041	-0.008
	[0.033]	[0.032]	[0.034]	[0.032]	(0.046)	(0.048)	(0.047)
Jordanian	0.101	0.068	0.102	0.167	-0.033	0.000	0.066
	[0.021]	[0.017]	[0.021]	[0.026]	(0.027)	(0.030)	(0.033)
Stigma (1st order)	-0.181	-0.042	-0.001	0.118	0.140	0.180	0.299
	[0.062]	[0.066]	[0.071]	[0.072]	(0.091)	(0.094)	(0.095)
Stigma (2nd order)	0.019	-0.068	-0.122	0.087	-0.087	-0.141	0.068
	[0.069]	[0.067]	[0.066]	[0.069]	(0.097)	(0.096)	(0.097)
Dictator allocation	4.011	4.062	3.722	3.962	0.051	-0.289	-0.049
	[0.223]	[0.203]	[0.226]	[0.213]	(0.302)	(0.318)	(0.309)
Social desirability	-0.018	-0.036	0.009	-0.015	-0.018	0.027	0.003
	[0.063]	[0.068]	[0.071]	[0.068]	(0.093)	(0.095)	(0.093)
PHQ-2	0.064	0.047	0.060	0.033	-0.017	-0.004	-0.031
	[0.067]	[0.067]	[0.071]	[0.069]	(0.095)	(0.098)	(0.097)
GAD-2	0.059	0.051	0.053	-0.030	-0.009	-0.006	-0.089
	[0.068]	[0.070]	[0.072]	[0.065]	(0.098)	(0.099)	(0.094)
Depression/Anxiety likely	0.541	0.555	0.563	0.521	0.014	0.022	-0.020
	[0.035]	[0.034]	[0.035]	[0.034]	(0.048)	(0.049)	(0.049)
Used MH care before	0.198	0.151	0.199	0.149	-0.047	0.001	-0.049
	[0.028]	[0.024]	[0.028]	[0.024]	(0.037)	(0.039)	(0.037)
Social connectedness	-3.445	-3.458	-3.426	-3.443	-0.013	0.019	0.002
	[0.045]	[0.048]	[0.051]	[0.049]	(0.066)	(0.068)	(0.066)
MH Efficacy beliefs	0.069	0.082	-0.086	0.061	0.014	-0.154	-0.008
	[0.065]	[0.058]	[0.074]	[0.063]	(0.087)	(0.098)	(0.090)
Network size	3.208	3.320	3.296	3.367	0.112	0.088	0.160
	[0.118]	[0.112]	[0.128]	[0.114]	(0.162)	(0.174)	(0.164)
F-stat relative to control	.75	.75	1.17	1.7	.75	1.17	1.7
F-stat relative to M2	.8	.8			.8		
F-stat relative to M3	1.58	1.58	2.15		1.58	2.15	
N Observations	207	219	206	215	426	413	422

This table tests for balance on sender baseline covariates comparing across framing arms. Columns 1-2 show the means with standard deviations in brackets below. Columns 5-7 show the difference in the means of each treatment arm with control, with the standard error in parentheses below. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$