American Psychological Association

Journal of Consulting and Clinical Psychology

2019, Vol. 87, No. 4, 370–379
http://dx.doi.org/10.1037/ccp0000389

0022-006X/19/$12.00

Randomized Controlled Trial of an Online Machine Learning-Driven Risk Assessment and Intervention Platform for Increasing the Use of Crisis Services

Adam C. Jaroszewski
Harvard University

Robert R. Morris
Koko, New York, New York

Matthew K. Nock
Harvard University

Objective: Mental illness is a leading cause of disease burden; however, many barriers prevent people from seeking mental health services. Technological innovations may improve our ability to reach underserved populations by overcoming many existing barriers. We evaluated a brief, automated risk assessment and intervention platform designed to increase the use of crisis resources provided to those online and in crisis. ***Method:*** Participants, users of the digital mental health app Koko, were randomly assigned to treatment or control conditions upon accessing the app and were included in the study after their posts were identified by machine learning classifiers as signaling a current mental health crisis. Participants in the treatment condition received a brief Barrier Reduction Intervention (BRI) designed to increase the use of crisis service referrals provided on the app. Participants were followed up several hours later to assess the use of crisis services. ***Results:*** Only about one quarter of participants in a crisis (21.8%) reported being "very likely" to use clinical referrals provided to them, with the most commonly endorsed barriers being they "just want to chat" or their "thoughts are too intense." Among participants providing follow-up data (41.3%), receipt of the BRI was associated with a 23% increase in the use of crisis services. ***Conclusion:*** These findings suggest that a brief, automated BRI can be efficacious on digital platforms, even among individuals experiencing acute psychological distress. The potential to increase help seeking and service utilization with such procedures holds promise for those in need of psychiatric services. ***Trial Registration:*** clinicaltrials.gov identifier: NCT03633825.

What is the public health significance of this article?
This study provides evidence that a brief, automated barrier reduction procedure increased the rate of service utilization among individuals in acute distress on digital platforms. These findings suggest that automated procedures may have the potential to increase help-seeking behavior among those in need of mental health services on a large scale.

Keywords: suicide, risk assessment, treatment barrier, help seeking, digital platform

Supplemental materials: http://dx.doi.org/10.1037/ccp0000389.supp

Mental illness is highly prevalent and a leading cause of disease burden (Kessler et al., 2007) accounting for nearly 25% of all years lost to disability (Whiteford et al., 2013). Unfortunately, only about one third (35.4%) of those with a mental disorder seek help during the year their disorder begins (Wang et al., 2007). People who report experiencing suicidal thoughts and behaviors (STBs) initiate contact with providers and use services at comparably low rates. Cross-national estimates indicated that approximately only 40% of people experiencing STBs in the past year received treatment for their mental health in the year prior (Bruffaerts et al., 2011).

Across many different types of mental health problems, perceived barriers prevent people from seeking and/or using mental health services. These barriers can be either attitudinal/evaluative in nature (e.g., stigma, low perceived efficacy of treatments) and/or structural (e.g., lack of financial means, transportation is-

Adam C. Jaroszewski, Department of Psychology, Harvard University; Robert R. Morris, Koko, New York, New York; Matthew K. Nock, Department of Psychology, Harvard University.

The research was supported by the Chet and Will Griswold Suicide Prevention Fund and the For the Love of Travis Fund. Robert R. Morris is a co-founder of Koko, a for-profit enterprise that provides mental health and safety services to large social networks.

Correspondence concerning this article should be addressed to Adam C. Jaroszewski, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: jaroszewski@fas.harvard.edu

sues, availability of care; Gulliver, Griffiths, & Christensen, 2010). For instance, among people who perceive a need for care but did not seek help, nearly 75% report that they wanted to handle their mental health issues on their own (Mojtabai et al., 2011). Some barriers appear to be unique to or particularly salient among people reporting STBs, such as the purposeful refusal or avoidance of help (known as "help-negation;" e.g., Clark & Fawcett, 1992; Deane, Wilson, & Ciarrochi, 2001; Rudd, Joiner, & Rajab, 1995). Studies also suggest that fear of hospitalization is a relatively prevalent (40%) barrier reported by callers to crisis lines experiencing suicidal thoughts (Gould, Munfakh, Kleinman, & Lake, 2012).

Technological innovations on digital platforms, including web-based services (e.g., Google, Facebook) and digital mental health apps, represent an important opportunity to improve our ability to reach underserved populations by overcoming many of these existing barriers (Kazdin & Blase, 2011). Such innovations afford greater access to resources (e.g., psychoeducation), assessment, and feedback (to both client and service provider), and time-sensitive risk identification, triage, and resource allocation (e.g., in crises involving STBs and/or other concerns). Unfortunately, few studies have experimentally tested the clinical efficacy of services deployed on digital platforms (e.g., <%1 of digital mental health apps; Donker et al., 2013). Moreover, whereas web services of many kinds (e.g., Facebook, Amazon's Alexa) have taken steps to identify high-risk users and refer them to crisis resources, little empirical work has investigated whether people who receive crisis referrals on digital platforms actually use them. Thus, despite the fact that innovations on digital platforms may address structural barriers to help seeking (e.g., geography), their ability to address attitudinal barriers (e.g., low perceived need for help) remains largely untested.

The purpose of this study was to develop and evaluate a brief, automated risk assessment and intervention platform designed to increase the use of crisis resources among individuals routed to a digital mental health app who were identified as likely experiencing a mental health crisis. Drawing on user feedback and research identifying common barriers to seeking help for mental health-related issues (e.g., Gulliver et al., 2010; Hom, Stanley, & Joiner, 2015; Mojtabai et al., 2011) and ways to overcome them via brief interventions (Nock & Kazdin, 2005), we developed and evaluated a brief psychoeducation intervention designed to reduce perceived barriers to using crisis resources (e.g., fear of police intervention/hospitalization) by clearing up misconceptions on which these barriers may be based. The primary aim was to evaluate whether this intervention increased the rate of using crisis resources. It was hypothesized that individuals assigned to the intervention condition would report using resources at higher rates than individuals in the control condition.

Method

Sample and Recruitment

Participants were recruited from 39,450 Koko (http://itskoko.com) users who signed up for the service between August 10, 2017 and September 20, 2017. Koko (née Panoply) provides safety services for large online social networks. Among its offerings is an online peer-to-peer crowdsourcing platform that teaches users cognitive reappraisal strategies that they then use to help other users manage negative emotions. The Koko network is strictly peer-to-peer, with no clinician or counselor oversight. Interactions between users are supervised by a combination of machine learning algorithms and on-demand human moderators. Koko uses a text-based user interface and is available on messaging services (e.g., Facebook Messenger, Kik), as well as desktop and mobile browsers. Because Koko is anonymous, no personally identifiable information is required or recorded. As such, no age, gender, location, or other sociodemographic information were collected. In a separate survey of Koko users from 2017 ($N = 496$), 65.0% of users were female and the majority were young adults ($M = 18.24$, $SD = 5.80$; Morris, 2018). Before using the Koko app, users accepted a terms and conditions agreement, which grants the collection of user content (e.g., "posts") and sharing of this content with third parties for research purposes. As part of its ongoing quality improvement efforts, Koko randomly assigns users to different versions of the app and examines differences in user engagement with the app, as was done in the current study.

Procedure

Randomization. Users ($N = 39{,}450$) were randomly assigned to either the treatment ($n = 19{,}612$) or control ($n = 19{,}838$) condition as soon as they accessed the platform (the CONSORT flow diagram; Figure 1). Randomization was conducted on a remote server at the moment each user account was created. Specifically, MongoDB was used to generate a random string of numbers and letters for each user, creating an anonymous ID. These ID values were summed and then passed to a modulo-2 operation, generating a remainder of zero (control condition) or one (treatment condition). Users were randomized upon accessing the platform, as opposed to later on when identified as being in crisis, because we used the platform's extant randomization tool, which has a primary purpose of AB-testing user-experience features (e.g., the platform's color palette) initiated at the user's point of entry onto the platform, not subsequent to classification. No AB tests or any other manipulations were conducted on the platform during the present study period.

Assessment and identification of those in crisis. Once on Koko, users can compose short descriptions (i.e., posts) of stressful situations and share them with the network. Among 39,450 Koko users, 74.3% ($n = 29{,}304$) made a post on the network (i.e., viewable by other users). Koko uses a hybrid human-machine computation system that evaluates the semantic content of posts in real time. As soon as a post is created on Koko, it is evaluated by a suite of machine learning classifiers, consisting of recurrent neural networks with word embeddings (Kshirsagar, Morris, & Bowman, 2017). In addition to other codes, each post is assigned a binary classification of "crisis" or "not crisis." The crisis classification is defined as possibly at risk of serious, imminent physical harm, either through self-inflicted actions or through abuse from a third party. Each post classified as crisis has a confidence threshold from zero to one. If the confidence is above .95, the system rules automatically. Otherwise, the user's post is independently reviewed by three moderators all of whom have been trained to identify potentially at-risk individuals. On average, the classifiers return results within 44 ms, while the moderators return judgments within 2 min ($M = 77.4$ s). The classifiers demonstrated excellent performance (area under the curve = .93, sensitivity =

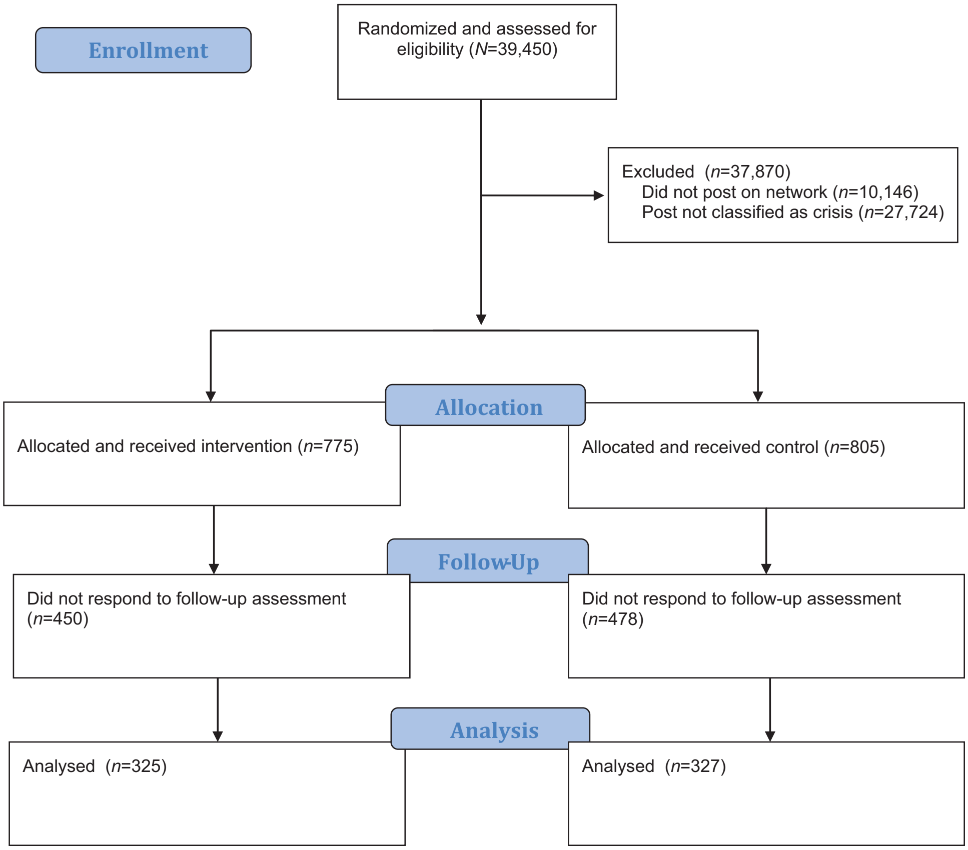
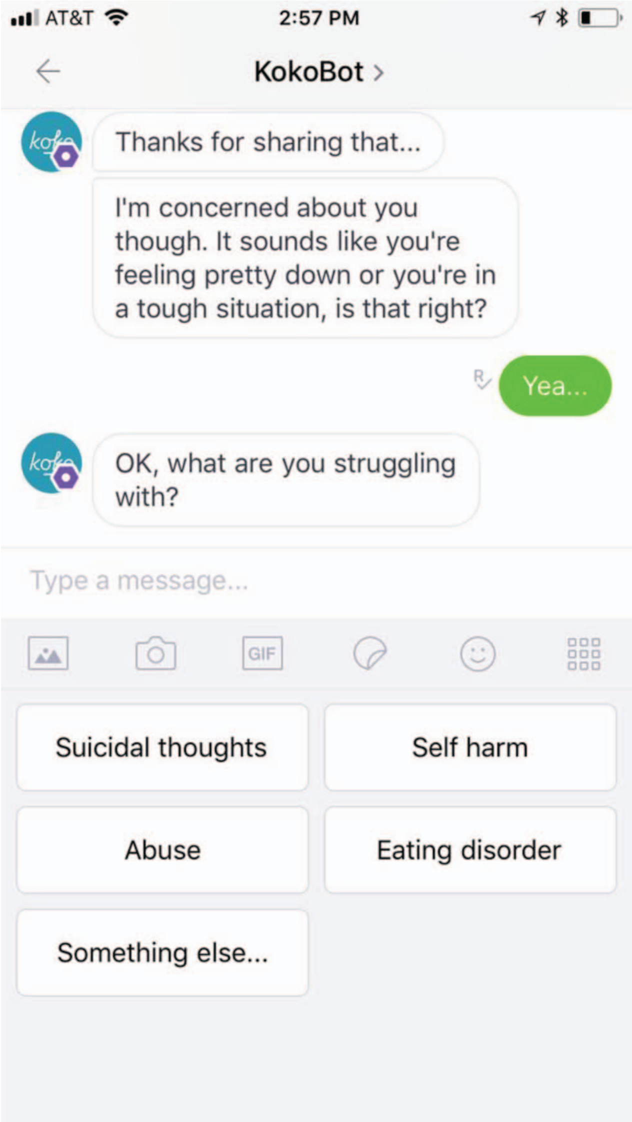

Enrollment
Randomized and assessed for eligibility (N=39,450)
Excluded (n=37,870)
Did not post on network (n=10,146)
Post not classified as crisis (n=27,724)
Allocation
Allocated and received intervention (n=775)
Allocated and received control (n=805)
Follow-Up
Did not respond to follow-up assessment (n=450)
Did not respond to follow-up assessment (n=478)
Analysis
Analysed (n=325)
Analysed (n=327)

Figure 1. CONSORT flow diagram. See the online article for the color version of this figure.

.64, specificity = .98, positive predictive value = .90, negative predictive value = .93, accuracy = .93; see Table 1 in the online supplemental materials for classifier confusion matrix) for identifying posts containing language representing a possible crisis as described above. Classifier performance metrics were derived from a large ground-truth sample consisting of >10,000 posts and coding labels. Ground-truth labels were established by consensus between human coders who were extensively trained and regularly audited for coding accuracy. Of the 29,304 users in this sample who made a post, 5.4% (*n* = 1,580) made a post that was classified as representing a crisis.

Users classified as being in crisis were immediately messaged by the Koko system and given a series of assessments. Users were first asked to identify what particular issue was troubling them most (e.g., "suicidal ideation," "self-injury," "eating disorder"). They also were given an opportunity to say they had been misclassified and weren't actually experiencing a crisis. Only users who reported serious concerns (e.g., "suicide," "self-injury") were given additional questions concerning suicidal ideation, past history of suicide attempts, and any current plans to attempt suicide. Questions were adapted from protocols used by the National Suicide Prevention Lifeline (Joiner et al., 2007). Users were assigned a "high-risk" status if they presented any past history of an attempt, current plan, or significant ideation. As soon as a user met the threshold for high risk the assessment stopped and crisis resources were presented immediately. All assessments were made within the Koko messenger platform, using a text- and button-based user interface (Figure 2). Koko uses simple language and emojis that are common to many messenger-based bots and applications. It also makes occasional use of a first-person persona ("KokoBot") to help users navigate the platform.

Among those classified as being in crisis, 65.5% (*n* = 1,036) acknowledged being in crisis, among whom the most commonly reported concerns were suicidal ideation (52.6%) and self-injury (21.1%), followed by eating disorders (7.4%), physical abuse (1.6%), unspecified abuse (1.1%), emotional abuse (0.5%), and otherwise unspecified (15.7%). Among the 34.5% of participants not acknowledging being in crisis, 51.6% (*n* = 281) reported being misclassified and not in crisis ("false crisis"), and 48.4% (*n* = 264) described posting about someone other than themselves ("someone else's crisis"). We decided a priori to include in the study and analyses all participants identified by the classifier as possibly being in crisis, including participants reporting they were misclassified, for two reasons: (a) the accuracy of self-reported crisis status may be biased by some of the very same barriers the intervention was intended to reduce (e.g., fear of police, preference for self- vs. professional management), and (b) based on our classifier's false discovery rate, we expected about 9.5% of participants to be falsely classified; however, it was impossible to know for sure which participants were falsely classified given the previously mentioned potential for biased reporting. Thus, including all participants identified by the classifier in the study and



Thanks for sharing that...

I'm concerned about you though. It sounds like you're feeling pretty down or you're in a tough situation, is that right?

Yea...

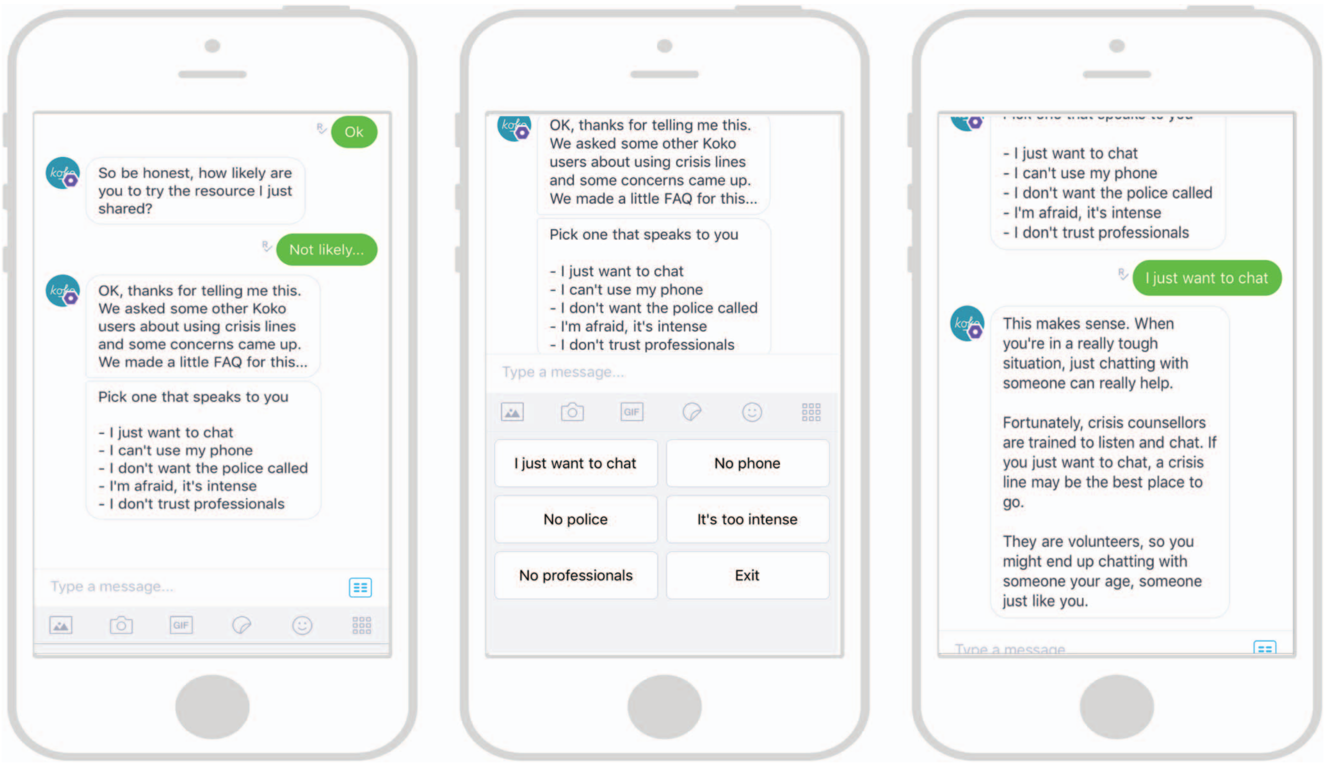
OK, what are you struggling with?

Figure 2. A screenshot of the Koko interface. A bot (KokoBot) helps users compose posts and responses and connects users to other members of the network. Users can respond to questions by typing in the text field or by selecting one of the buttons. Only one option can be selected for any particular question. See the online article for the color version of this figure.

analyses was the only way to ensure that all participants meeting study inclusion criteria (i.e., likely experiencing a crisis) were, in fact, included. Additionally, among participants providing follow-up data, a substantial portion reporting false crisis or someone else's crisis at baseline indicated they used crisis services at follow up (40.7% and 47.8%, respectively), and a chi-square test indicated that the type of crisis category was not associated with different rates of crisis resource service use ($\chi^2(5) = 2.71$, $p = .74$). This suggests some reports of false crisis and someone else's crisis may have been inaccurate. Therefore, the 1,580 participants classified as being in crisis represent the sample that was the focus of our randomized controlled trial (RCT). They all received the same intervention regardless of crisis type because these crises often co-occur (e.g., physical abuse and emotional abuse) and because the intervention was designed to send people to crisis services regardless of the specific nature of the crisis.

Experimental conditions. All participants classified as being in crisis are automatically shown a list of crisis resources (e.g., the National Suicide Prevention Lifeline). This list varies depending on a user's country of origin and the type of issue reported (e.g., the National Eating Disorders Association is provided for United States-based individuals who struggle with eating disorders). After seeing the crisis service referrals, the participant experience differs for those in the control versus treatment conditions. Participants randomly assigned to the control condition continue using the Koko platform as usual. Participants randomly assigned to the treatment condition, by contrast, receive additional intervention. They are first asked if they would use the resources presented on screen (i.e., "Be honest, how likely are you to try the resources I just shared?"). Individuals who answered "very likely" continued using the Koko service as usual. Individuals who answered "not likely . . . " were presented with an interactive Barrier Reduction Intervention (BRI).

The BRI was designed to overcome common concerns and misconceptions (i.e., barriers) related to using crisis services, thereby increasing utilization of these services. It works by first asking the user about what potential barriers may keep them from using the crisis service referrals, and then, based on the user's response, by providing information intended to help the user overcome the potential barrier(s) they selected. In order to create a list of potential barriers for the BRI, we asked thousands of Koko users to tell us why they had avoided the crisis service referrals that were presented to them on Koko. A random set of 245 responses were analyzed and coded by trained raters to identify common reasons. Among the most common barriers reported was related to users believing they weren't actually in danger and could handle the situation themselves (13.11%). This reason for avoiding crisis referrals was not easily addressed by the BRI, because the BRI targeted barriers that are based on misconceptions (e.g., fear of police intervention) capable of being clarified with factual information in 2–3 sentences. It was impossible to know for certain whether participants were mistaken in their belief about not being in danger. Therefore, we did not want to suggest that participants were mistaken in this belief lest they perceive the intervention as irrelevant and/or invalidating, likely decreasing help-seeking behavior. The decision to not target this barrier was made a priori. However, the tendency to minimize symptom severity and the preference for self-management are common barriers among individuals at elevated risk for suicide and, therefore, better identifying and intervening upon them will be the subject of future exploration. Five other common barriers were presented to users in the form of an in-app menu that was ordered by most to least common (Figure 3). By exploring the menu of barriers, users could read brief messages designed to dispel common misconceptions or concerns related to each barrier. For example, a common concern among Koko users was that calls to lifelines invariably result in visits by the police or other emergency services. Users who feared this possibility could tap on the associated button and learn that active rescues such as these are extremely rare and occur in less than one percent of all cases. Whenever possible, we used language throughout the intervention to help validate the experiences



Ok

So be honest, how likely are you to try the resource I just shared?

Not likely...

OK, thanks for telling me this. We asked some other Koko users about using crisis lines and some concerns came up. We made a little FAQ for this...

Pick one that speaks to you

- I just want to chat
- I can't use my phone
- I don't want the police called
- I'm afraid, it's intense
- I don't trust professionals

OK, thanks for telling me this. We asked some other Koko users about using crisis lines and some concerns came up. We made a little FAQ for this...

Pick one that speaks to you

- I just want to chat
- I can't use my phone
- I don't want the police called
- I'm afraid, it's intense
- I don't trust professionals

- I just want to chat
- I can't use my phone
- I don't want the police called
- I'm afraid, it's intense
- I don't trust professionals

I just want to chat

This makes sense. When you're in a really tough situation, just chatting with someone can really help.

Fortunately, crisis counsellors are trained to listen and chat. If you just want to chat, a crisis line may be the best place to go.

They are volunteers, so you might end up chatting with someone your age, someone just like you.

Figure 3. Screenshots of the barrier reduction intervention on Koko. Users were presented a list of common concerns that they could explore by tapping one of the buttons. See the online article for the color version of this figure.

of the users (see Figures 2 and 3). We referenced other Koko users who may have felt similarly and acknowledged that each concern had merit and could not be rejected outright. We did some pilot testing of the barriers used in the BRI; however, no data were systematically collected or analyzed during the pilot testing period. Therefore, we implemented this RCT as a formal test of the BRI.

Outcome assessment. Five hours after completing the screening and receiving crisis resources, users were sent a notification request to answer some follow-up questions. Notifications could be delivered through an in-app notification, a text message, or an e-mail, depending on the messenger/online platform being used and the user's notification preferences. The median time between being shown crisis resources and initiating the follow-up survey was 11 hr and 12 min. The follow-up assessment included questions regarding resource utilization ("Did you try any of the resources I sent you?"). Follow-up data were obtained for 652 (41.3%) participants. There were no significant differences between those who participated in the follow-up assessment and those who did not on all study variables collected during baseline assessment. However, participants who did not complete the follow-up assessment were marginally more likely to be categorized as high risk (50.6% vs. 45.8%, $\chi^2(1) = 3.51$, $p = .06$, $RR = 1.11$, 95% CI [0.99, 1.26]).

Data Analytic Plan

Outcomes. We used chi-square tests to determine whether the treatment and control groups were equivalent on baseline factors (e.g., suicidal thoughts/behaviors). We conducted a chi-square test to examine the hypothesis that individuals assigned to the intervention condition would report using crisis resources at higher rates than individuals in the control condition. Follow-up analyses consisted of a chi-square test to examine whether groups differed in terms of which crisis resource they used and a series of logistic regressions to examine potential moderators (e.g., history of suicidal thoughts/behavior, risk level) of the effect of intervention on the rate of using crisis resources. Analysis of these preexisting, anonymous data was approved by the Harvard University Institutional Review Board.

Missing data. We used chi-square tests to determine whether follow-up data were missing completely at random (MCAR) or whether missingness depended on observed data collected at baseline (i.e., risk level, crisis category, allocation). These tests were nonsignificant, suggesting that data were MCAR and, therefore, that effect estimates based on complete-case analysis were possibly accurate and unbiased. However, the relationship between baseline risk level and missingness at follow up was marginally significant ($p = .06$). Therefore, we analyzed imputed data and conducted complete-case analysis with covariate adjustment (cf., Groenwold, Donders, Roes, Harrell, & Moons, 2012) to provide additional assurance that estimates from nonmissing data were possibly accurate. We performed multivariate imputation by chained equations using the "mice" package (van Buuren & Groothuis-Oudshoorn, 2011) in R (R Core Team, 2008), generating 1,000 imputed data sets with 50 iterations. Results from analyses run on each dataset were pooled according to Rubin's (1987) rules. Both imputation results and results from complete case analysis with covariate adjustment are presented in Table 2 in the online supplemental materials. Results for nonimputed, complete cases are reported in the present paper (n = 652); however, we also report proportions of participants using crisis services and related statistics (e.g., number needed to treat; NNT) based on the full sample providing baseline data (n = 1,580; i.e., intent-to-treat [ITT] analyses). Notably, nonimputed, imputed, and covariate adjusted complete-case analyses produced nearly identical results.

Results

Group Equivalence at Baseline

To rule out the alternative hypothesis that any observed between-group differences in the use of crisis services might be due to differences in baseline factors like history of suicidal thoughts/behaviors or level of risk, we compared participants assigned to the treatment and control conditions on such factors. No between-group differences were observed, suggesting that randomization resulted in group equivalence at baseline (Table 1).

Barriers to Using Crisis Services

Participants randomized to the treatment condition were initially asked how likely they were to use the crisis resources presented to them on the Koko app. Only 21.8% of participants reported being very likely to use such resources. These participants were routed directly to the resources (as opposed to administering the BRI before presenting the resources) so as to not impede their access to them. "Very likely" participants were about 60% more likely (RR = 1.59) to actually use crisis resources (as reported during follow up) than "not likely" participants (69.0% vs. 43.3%, $\chi^2(1) = 14.86$, $p < .001$, 95% CI [1.27, 1.95]). Participant report of likelihood of using crisis services was unrelated to the baseline presence of suicidal thoughts/behaviors or risk level (presented in Table 3 in the online supplemental materials).

Table 1
Descriptive Statistics and Between-Group Comparisons

Variable	Control (n = 327), %	Treatment (n = 325), %	Between-group comparison		
			χ^2	df	p
Suicide ideation			.91	2	.63
Yes	47.4	42.7			
No	7.0	5.8			
Unknown	45.6	45.2			
Suicide plan			1.03	2	.59
Yes	13.8	16.6			
No	6.7	6.5			
Unknown	79.5	76.9			
Suicide attempt			2.14	2	.34
Yes	26.9	22.2			
No	20.5	23.0			
Unknown	52.6	54.8			
Crisis category			7.01	5	.53
Suicide	34.5	31.5			
Self-harm	12.9	13.5			
Someone else	17.3	18.1			
False crisis	17.3	16.7			
Eating disorder	5.2	5.6			
Other	12.8	14.6			
Risk level			.10	1	.74
Low risk	53.5	54.8			
High risk	46.5	45.2			

Participants who reported being not likely to use crisis services were presented with potential barriers to their use of such services. Most participants (78.2%) endorsed one of the reasons queried. The most commonly endorsed reasons/barriers were: "I just want to chat" (53.1%), "Don't want police called" (42.5%), "Thoughts too intense" (30.7%), "Don't trust professionals" (19.7%), and "Don't have a phone" (13.7%; see Figure 3).

Effect of the Intervention

After participants in the treatment condition were presented with potential barriers, they received the BRI. Among participants providing follow-up data, participants assigned to the BRI condition were 23% (RR = 1.23) more likely to later use crisis services than participants in the control condition (48.9% vs. 39.8%, $\chi^2(1) = 5.55$, $p = .02$, 95% CI [1.03, 1.46], NNT = 10.98). (Note: the corresponding statistics for ITT analyses were 20.51% vs. 16.14%, RR = 1.20, $p = .02$, 95% CI [1.04, 1.42], NNT = 22.93). There was no between-group difference regarding which crisis resources participants used (Table 2). There also were no moderators of the observed treatment effect. More specifically, a series of logistic regressions revealed that neither suicidal ideation (RR = 0.98, $p = .95$), presence of suicide plan (RR = 0.82, $p = .59$), past suicide attempt (RR = 1.15, $p = .57$), risk level (RR = 0.99, $p = .96$), nor the type of possible crisis (e.g., suicide RR = 0.95, $p = .87$) moderated the effect of the intervention on the rate of using crisis resources (presented in Table 4 in the online supplemental materials). Exploratory analyses revealed that neither the type of barrier explored (e.g., "just want to chat"), nor the total or unique number of barriers explored was associated with increased probability of crisis resource use among participants receiving the BRI (presented in Table 5 in the online supplemental materials). At follow up, participants were asked whether the KokoBot "did a good job helping [them]," and rated their experience as either "good" or "bad," with nearly 75% (n = 484) rating their experience as good. Ratings did not differ between treatment and control participants, $\chi^2(1) = 1.69$, $p > .05$; however, across both groups, participants reporting actually using crisis resources were nearly one third (RR = 1.32) more likely to rate Koko as good at helping them than participants who reported not using resources (79.2% vs. 70.2%, $\chi^2(1) = 6.79$, $p < .01$, 95% CI [1.07, 1.67].

Discussion

The purpose of this study was to develop and evaluate a brief BRI designed to increase the use of crisis resources among individuals identified as being at risk for a mental health crisis. There were three main findings in this study. First, among participants assigned to the BRI who provided follow-up data, only about one quarter (21.8%) presented with crisis service referrals report being likely to use them, and only about two thirds of that group (69.0%) reported actually doing so. Second, participants' most-endorsed barriers to using crisis services included preferring to chat with

Table 2
Descriptive Statistics and Between-Group Comparisons for Follow-Up Assessment Data

			Between-group comparison		
Variable	Control (n = 327), %	Treatment (n = 325), %	χ^2	*df*	*p*
Used resources			5.55	1	**.02**
Yes	39.8	48.9			
No	60.2	51.1			
Which resources used			6.26	5	.28
Crisis text line	14.4	18.5			
Suicide hotline	4.6	5.2			
International	1.8	2.5			
Recovery road	.9	1.8			
Unknown	17.4	20.3			
Not available	60.9	51.7			

Note. Bold type denotes $p < .05$.

other people on their phone/computer instead of seeking help through the provided crisis service referrals, fearing that the police may be called, and perceiving that their thoughts were too intense to share with a professional at one of the crisis referrals. Third, and perhaps most importantly, a brief BRI significantly increased participants' likelihood of using crisis services. Each of these findings warrants further comment.

The low rate of help seeking observed in this study, even among those reporting they were likely to use crisis referrals provided to them, is consistent with data from previous studies among individuals with severe mental health issues and STBs (Bruffaerts et al., 2011; Hom et al., 2015). Whereas digital platforms hold great promise by overcoming structural barriers to seeking help (e.g., geographical), attitudinal barriers persist (e.g., preference for informal help), underscoring the need to develop and evaluate interventions that target barriers to help seeking, in general, and on digital platforms, in particular. Among participants in the treatment condition providing follow-up data who reported they were not likely to use the crisis service referrals provided to them, 53.1% reported just wanting to chat and 30.7% reported perceiving that their thoughts were too intense. The preference for informal help and the perception that suicidal thoughts are too intense or dangerous to share with a service provider are consistent with findings from studies on help negation (e.g., Wilson & Deane, 2010). Help negation is the tendency to purposefully refuse or avoid help, and is commonly observed among individuals experiencing STBs. Researchers have proposed that suicidal thoughts and concomitant affective states may themselves contribute to help negation by exacerbating existing vulnerabilities (e.g., behavioral avoidance, cognitive distortion, inhibited recall), leading to ineffective problem solving (e.g., Deane et al., 2001; Rudd et al., 1995; Weishaar, 1996; Wilson & Deane, 2010). Therefore, people experiencing suicidal thoughts may avoid speaking with a professional because suicidal thoughts and the accompanying distress may contribute to thinking errors that interfere with the ability to select more effective solutions than avoidance, such as seeking professional help.

The main finding of this study demonstrates that a brief, automated barrier reduction strategy can be efficacious, even among individuals experiencing acute psychological distress. Moreover, a vast majority of participants (78.2%) providing follow-up data who reported they were not likely to actually use the crisis resources available on Koko did, in fact, explore at least one psychoeducation section, suggesting that this brief and automated intervention was capable of engaging even a reluctant and acute subgroup of participants. While we did hypothesize that the psychoeducation intervention would increase crisis resource utilization, no variables we tested moderated this effect; however, statistical power was limited to sufficiently assess several potential moderating variables (e.g., suicide ideation, suicide plan, and suicide attempt) due to small and unbalanced subgroup sample sizes.

There has been a dramatic increase in recent years of the number of digital platforms focused on mental health; however, very few studies investigating such platforms have used experimental designs and collected follow-up data from participants (<1%; Donker et al., 2013). This makes it difficult to judge whether digital platforms related to mental health are actually efficacious. Digital platforms that help those with mental health problems are desperately needed to curb the enormous burden associated with mental illness by providing services to those with unmet needs. Platforms most capable of bridging this gap must be acceptable to users, scalable (e.g., using automated procedures), and sensitive to the concerns (i.e., barriers) of people reluctant or unwilling to seek or utilize traditional mental health services. This study represents an important step toward the goal of developing and evaluating a brief, automated intervention designed to reduce barriers in order to increase the utilization of crisis resources among individuals in acute need.

These findings have important implications for psychological and public health researchers as well as creators of digital apps focused on mental health. This intervention can be implemented and studied by psychological and public health researchers in a range of settings (e.g., web- and mobile-based digital mental health services/apps) and across a variety of populations with relative ease. Digital platform designers can integrate brief, automated psychoeducation interventions on their platforms to reduce barriers to treatment and increase service utilization and user engagement. While further research is needed to test and improve this intervention, individual and public health may benefit from its immediate use. Psychoeducation interventions may help people overcome barriers to treatment, increasing service use among vulnerable populations, such as those experiencing mental health crises. If true, such interventions could have large downstream effects, such as helping reduce the disease burden associated with mental health problems as well as the burden associated with physical and medical health issues by reducing stress on and access blocks to health care safety nets like emergency departments.

Careful discussion must continue among researchers, institutional review boards, medical ethicists, app developers, and consumers of health apps about how to best conduct clinical research on digital platforms. Important ethical questions remain unanswered, including to what extent users should be informed that their experience on apps and digital platforms might be curated and/or experimentally manipulated to test and improve health- and other-related outcomes? How to accurately calculate and balance the potential benefits and costs of manipulating users' experience, particularly in health-related apps? And should research conducted in health-related apps involving experimental manipulation provide a higher benefit-to-cost ratio compared with other types of apps (e.g., financial) given the higher likelihood that users may be part of a vulnerable population?

Limitations

Despite these positive implications, several important limitations of this study deserve comment. First, the generality of the findings may be restricted. This study was completed among individuals identified as likely experiencing a mental health crisis while they were using a digital app. Thus, these results may not generalize to individuals with less acute mental health problems or to those not technologically savvy or otherwise reluctant to use digital platforms. Second, the classifier provides a probability estimate of the possibility an individual is at risk of serious, imminent harm, not an objective metric of crisis status, such as the observation of suicidal behavior. Therefore, it is possible that some individuals classified as being in crisis were misclassified. Despite this, a majority of participants indicated they were currently struggling with one of the common mental health issues provided prior to the BRI, and a significant portion of participants providing follow-up data went on to use crisis services, including participants who initially reported they were not in crisis. However, future studies may benefit by attempting to corroborate crisis classification status by, for example, assessing users with expert clinicians following automated crisis classification. The strength of this study lies in demonstrating that individuals in possible crisis, who are often disinclined to seek help or utilize services, are amenable to a brief and automated intervention. That is to say, in this study we prioritized maximizing internal validity rather than the external validity of the intervention. Additionally, because classification of a potential mental health crisis was based on text data from users' posts, it is possible that users who did not post or who posted very little text were incorrectly classified as not being in crisis. Sparsity of information is a limitation affecting all text-based classifiers as well as human clinical judgment alike. Future studies may benefit by combining different types of data (e.g., passive data, user-platform engagement data) to improve accurate classification. Additionally, a majority of participants who completed baseline assessment (58.7%) did not provide follow-up data. This dropout rate is consistent with other similar studies conducted on digital platforms (Eysenbach, 2005; Fleming et al., 2018); however, these results may not generalize to all individuals experiencing a crisis while using a digital platform. Additionally, low response rates may represent potential bias among individuals who responded (e.g., higher conscientiousness), leading to overestimating the effect of intervention. Results from imputed data and covariate-adjusted complete case analysis were nearly identical to nonimputed data, suggesting that results from nonimputed data are possibly accurate and unbiased; however, it is possible that unobserved variables were associated with missingness at follow up. Future studies should increase efforts aimed at collecting both follow-up data from participants who do not initially respond and additional baseline variables potentially associated with the probability of responding at follow up. Doing so might provide information on potential biases or other factors (e.g., possible engagement in self-injurious behaviors) among responders, allowing for more accurate estimates of effects as well as potential limitations of this intervention. Notably, there were no significant differences between those who provided follow-up data and those who did not on variables collected at baseline; however, participants classified as high risk were marginally less likely to provide follow-up data than low risk participants, suggesting that individuals with a history of severe STBs (e.g., past suicide attempt) may be less amenable to using crisis service referrals, which is consistent with research on help negation (Hom et al., 2015).

Several aspects of the assessment and intervention procedures limit the conclusions that can be drawn from this study. First, at follow up, participants reported whether they actually used crisis resources. Although the measure of treatment utilization is consistent with the methods used in other studies, it may have been subject to various biases (e.g., social desirability). Future studies of treatment utilization should aim to not rely solely on self-report, but instead use objective measures whenever possible, such as tracking technology on digital platforms (e.g., cookies). Second, although this study examined presence of STBs as well as the type of mental health crisis participants were experiencing, many other factors that may influence utilization of crisis resources were not evaluated. A broader range of participant characteristics that may influence crisis resource use should be included in future studies, such as sociodemographics, psychiatric symptoms, treatment history, and evaluative/attitudinal barriers to help seeking. Third, the intervention consisted of two parts: (a) querying participants about the likelihood they would actually use crisis resources available on Koko, and (b) the psychoeducation component (i.e., the BRI) consisting of sections of textual information, each addressing one common barrier to treatment (e.g., fear of hospitalization). Importantly, only participants reporting that they were "not likely" to use crisis resources received the psychoeducation component of the intervention, whereas participants reporting they were "very likely" to use crisis resources did not receive the psychoeducation component. Because participants in the control group did not report on their likelihood of using available crisis resources, we cannot directly compare the portion of treatment and control participants who were not likely to utilize resources. Thus, our ability to evaluate the effect of the intervention among this subgroup of participants is hindered. In addition, it is possible that asking participants how likely they were to use crisis resources played a causal role in whether they subsequently used crisis resources. Importantly, we asked participants to report on their likelihood of utilizing resources for ethical reasons. All participants in the study were identified as experiencing a mental health crisis. Thus, if a participant in crisis is likely to use crisis resources on their own without any intervention, it would be unethical to impede their access to resources by first administering the intervention to them. Future studies should ask all participants to report on their likelihood of using crisis resources. This will allow for the comparison of "not likely" resource respondents in both groups and help isolate the effect of the psychoeducation component of the intervention on this subgroup. Finally, exploratory analyses revealed that neither the type of barrier explored, nor the total or unique number of barriers explored increased the probability that participants receiving the BRI would use crisis resources. Due to small and unbalanced subgroup sample sizes, statistical power was limited to sufficiently assess the causal effect of these "molecular" parts of the intervention; however, these findings suggest the causal effect of the intervention may be driven by either unidentified variable(s) and/or the "molar" treatment package. Future studies would benefit from a larger sample size to overcome this limitation.

Conclusion

The intervention developed and evaluated in this study was delivered in a brief and automated format. This minimal design yielded promising results. Modifications to certain aspects of this intervention should be evaluated in future studies. For example, the intervention could be modified with ease to control for the effect of querying participants' likelihood of utilizing treatment resources, thereby isolating the effect of the psychoeducation component of the intervention. Also, asking for more information from participants at baseline (e.g., sociodemographics, psychiatric symptoms, evaluative/attitudinal barriers) may provide potential moderators and mediators to test. Moreover, future studies that expand the psychoeducation component to include additional potential barriers to utilizing mental health services (e.g., preference for self-management, help negation) as well as barriers that are predicted to be particularly relevant to participants based on their sociodemographic and psychiatric variables, may elicit stronger intervention effects and are likely to provide more information regarding the potential dose-response relationship between treatment and outcome. These proposed modifications represent only a small range of potential ways to research barrier reduction strategies on digital platforms. Future work in this space is necessary to understand whether new or a subset of known barriers to help seeking and/or treatment utilization emerge or prove particularly potent on digital platforms such as web-based services and digital mental health apps. The potential to increase help seeking and service utilization among underserved populations with brief, automated procedures holds great promise for researchers, clinicians, public health, and particularly for those in need of mental health services.

References

Barnard, J., & Rubin, D. B. (1999). Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika, 86,* 948–955. http://dx.doi.org/10.1093/biomet/86.4.948

Bruffaerts, R., Demyttenaere, K., Hwang, I., Chiu, W. T., Sampson, N., Kessler, R. C., . . . Nock, M. K. (2011). Treatment of suicidal people around the world. *The British Journal of Psychiatry, 199,* 64–70. http://dx.doi.org/10.1192/bjp.bp.110.084129

Clark, D. C., & Fawcett, J. (1992). Review of empirical risk factors for evaluation of the suicidal patient. In B. Bongar (Ed.), *Suicide: Guidelines for assessment, management and treatment* (pp. 16–48). New York. NY: Oxford University Press.

Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2001). Suicidal ideation and help-negation: Not just hopelessness or prior help. *Journal of Clinical Psychology, 57,* 901–914. http://dx.doi.org/10.1002/jclp.1058

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: A systematic review. *Journal of Medical Internet Research, 15,* e247. http://dx.doi.org/10.2196/jmir.2791

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7,* e11. http://dx.doi.org/10.2196/jmir.7.1.e11

Fleming, T., Bavin, L., Lucassen, M., Stasiak, K., Hopkins, S., & Merry, S. (2018). Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *Journal of Medical Internet Research, 20,* e199. http://dx.doi.org/10.2196/jmir.9275

Gould, M. S., Munfakh, J. L. H., Kleinman, M., & Lake, A. M. (2012). National suicide prevention lifeline: Enhancing mental health care for suicidal individuals and other people in crisis. *Suicide and Life-Threatening Behavior, 42,* 22–35. http://dx.doi.org/10.1111/j.1943-278X.2011.00068.x

Groenwold, R. H. H., Donders, A. R. T., Roes, K. C. B., Harrell, F. E., Jr., & Moons, K. G. M. (2012). Dealing with missing outcome data in randomized trials and observational studies. *American Journal of Epidemiology, 175,* 210–217. http://dx.doi.org/10.1093/aje/kwr302

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry, 10,* 113. http://dx.doi.org/10.1186/1471-244X-10-113

Hom, M. A., Stanley, I. H., & Joiner, T. E., Jr. (2015). Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: A review of the literature. *Clinical Psychology Review, 40,* 28–39. http://dx.doi.org/10.1016/j.cpr.2015.05.006

Joiner, T., Kalafat, J., Draper, J., Stokes, H., Knudson, M., Berman, A. L., & McKeon, R. (2007). Establishing standards for the assessment of suicide risk among callers to the National Suicide Prevention Lifeline. *Suicide and Life-Threatening Behavior, 37,* 353–365.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science, 6,* 21–37. http://dx.doi.org/10.1177/1745691610393527

Kessler, R. C., Angermeyer, M., Anthony, J. C.,Graaf, R. D., Demyttenaere, K., Gasquet, I., . . . Ustün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry, 6,* 168–176.

Kshirsagar, R., Morris, R., & Bowman, S. (2017). *Detecting and explaining crisis*. Retrieved from http://arxiv.org/abs/1705.09585

Mojtabai, R., Olfson, M., Sampson, N. A., Jin, R., Druss, B., Wang, P. S., . . . Kessler, R. C. (2011). Barriers to mental health treatment: Results from the National Comorbidity Survey Replication. *Psychological Medicine, 41,* 1751–1761. http://dx.doi.org/10.1017/S0033291710002291

Morris, R. (2018). *Demographics of Koko Users*. Internal Koko report: unpublished.

Nock, M. K., & Kazdin, A. E. (2005). Randomized controlled trial of a brief intervention for increasing participation in parent management training. *Journal of Consulting and Clinical Psychology, 73,* 872–879. http://dx.doi.org/10.1037/0022-006X.73.5.872

R Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley. http://dx.doi.org/10.1002/9780470316696

Rudd, D. M., Joiner, T. E., & Rajab, H. (1995). Help negation after acute suicidal crisis. *Journal of Consulting and Clinical Psychology, 63,* 499–503. http://doi.org/10.1037/0022-006X.63.3.499

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45,* 1–67. http://dx.doi.org/10.18637/jss.v045.i03

Wang, P. S., Angermeyer, M., Borges, G., Bruffaerts, R., Tat Chiu, W., Girolamo, D. E. G., . . . Ustün, T. B. (2007). Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry: Official Journal of the World Psychiatric Association, 6,* 177–185. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18188443

Weishaar, M. E. (1996). Cognitive risk factors in suicide. In P. Salkorskis (Ed.), *Frontiers of cognitive therapy* (pp. 226–249). New York, NY: Guilford Press.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., . . . Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet, 382,* 1575–1586. http://dx.doi.org/10.1016/S0140-6736(13)61611-6

Wilson, C. J., & Deane, F. P. (2010). Help-negation and suicidal ideation: The role of depression, anxiety and hopelessness. *Journal of Youth and Adolescence, 39,* 291–305. http://dx.doi.org/10.1007/s10964-009-9487-8

Received March 27, 2018
Revision received January 2, 2019
Accepted January 10, 2019 ■