

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: Evidence from internet search behavior in a large U.S. cohort

Nicholas C. Jacobson^{a,b,c,d,*}, Elad Yom-Tov^{e,h}, Damien Lekkas^{a,d}, Michael Heinz^{a,f}, Lili Liu^d, Paul J. Barr^{a,g}

^a Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway, EverGreen Center, Suite 315, Lebanon, NH, 03756, United States

^b Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Williamson Building, 3rd Floor, 1 Medical Center Drive, Lebanon, NH, 03756, United States

^c Department of Psychiatry, Geisel School of Medicine, Dartmouth College, 1 Medical Center Drive, Lebanon, NH, 03756, United States

^d Quantitative Biomedical Sciences Program, Dartmouth College, NH, United States

^e Microsoft Research, 13 Shenkar Street, Herzeliya, 4672513, Israel

^f Dartmouth-Hitchcock Medical Center, 1 Medical Center Drive, Lebanon, NH, 03756, United States

^g The Dartmouth Institute, Geisel School of Medicine, Dartmouth College, Williamson Building, 5th Floor, 1 Medical Center Drive, Lebanon, NH, 03756, United States

^h Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa, 3200003, Israel

ARTICLE INFO

Keywords:

Online screening tool
Machine learning
Internet search behavior
Suicidal ideation
Suicidal intent

ABSTRACT

Introduction: Most people with psychiatric illnesses do not receive treatment for almost a decade after disorder onset. Online mental health screens reflect one mechanism designed to shorten this lag in help-seeking, yet there has been limited research on the effectiveness of screening tools in naturalistic settings.

Material and methods: We examined a cohort of persons directed to a mental health screening tool via the Bing search engine ($n = 126,060$). We evaluated the impact of tool content on later searches for mental health self-references, self-diagnosis, care seeking, psychoactive medications, suicidal ideation, and suicidal intent. Website characteristics were evaluated by pairs of independent raters to ascertain screen type and content. These included the presence/absence of a suggestive diagnosis, a message on interpretability, as well as referrals to digital treatments, in-person treatments, and crisis services.

Results: Using machine learning models, the results suggested that screen content predicted later searches with mental health self-references (AUC = 0.73), mental health self-diagnosis (AUC = 0.69), mental health care seeking (AUC = 0.61), psychoactive medications (AUC = 0.55), suicidal ideation (AUC = 0.58), and suicidal intent (AUC = 0.60). Cox-proportional hazards models suggested individuals utilizing tools with in-person care referral were significantly more likely to subsequently search for methods to actively end their life ($HR = 1.727$, $p = 0.007$).

Discussion: Online screens may influence help-seeking behavior, suicidal ideation, and suicidal intent. Websites with referrals to in-person treatments could put persons at greater risk of active suicidal intent. Further evaluation using large-scale randomized controlled trials is needed.

Psychiatric illness is highly prevalent and burdensome for both individuals and society, impacting nearly one in five adults (18.9% of the population) annually (Bose, 2017). Mental and substance use disorders account for nearly 7.4% of disease burden worldwide, and they are the

single largest contributor to years lived with disability (22.9%) (Whiteford et al., 2013). The cost of psychiatric illness in 2010 was estimated at 2.5 trillion U.S. dollars and expected to increase to 6 trillion U.S. dollars by the year 2030 (Bloom et al., 2011).

* Corresponding author. Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway, Suite 300, Office # 333S, Lebanon, NH, 03766, United States.

E-mail addresses: Nicholas.C.Jacobson@dartmouth.edu (N.C. Jacobson), eladyt@microsoft.com (E. Yom-Tov), Damien.Lekkas.GR@dartmouth.edu (D. Lekkas), Michael.V.Heinz@hitchcock.org (M. Heinz), Lili.Liu.GR@dartmouth.edu (L. Liu), Paul.J.Barr@dartmouth.edu (P.J. Barr).

<https://doi.org/10.1016/j.jpsychires.2020.11.010>

Received 23 July 2020; Received in revised form 30 October 2020; Accepted 3 November 2020

Available online 9 November 2020

0022-3956/© 2020 Elsevier Ltd. All rights reserved.

There is an ever-growing array of evidence-based pharmacotherapies and psychotherapies available for the treatment of psychiatric illnesses both in-person and online; however, numerous studies suggest a significant lag time between the onset of symptoms and seeking mental health care, showing a low probability of treatment contact during the first 1–2 years of the illness, across multiple illness domains (Carter et al., 2003; Maggaard et al., 2017; Shidhaye et al., 2015; Thompson et al., 2008). On average, there is a treatment delay of approximately eight years after initial symptom onset (Wang et al., 2005). Factors associated with longer delays in seeking treatment include younger age at onset of symptoms, slower problem recognition, a more severe level of disability, self-stigma, inability to access treatment, being male, and identifying as part of a minority population (Maggaard et al., 2017; Thompson et al., 2008).

There has been considerable thought from researchers about methods to close this treatment gap, with efforts factoring the quantity of the services provided, the cost of the services, and the tier of care (Shidhaye et al., 2015). Given their cost and scalability, self-care models have been a focus of a great deal of research (Gulliver et al., 2010), with the most recent attention on the promise of digital treatments to meet this need (Wilhelm et al., 2020). Being aware of their symptoms, many persons seek information online as a first step towards care receipt, they often come to mental health screens which can in turn connect them to these potential treatments. Fitting this need, online screening tools are designed to quickly help determine if testers are experiencing symptoms of a mental health problem. Depending on the results of screening tools, testers could be directed to a variety of care mechanisms, including self-help care, digital treatments, and in-person mental health care (Webb et al., 2017).

Research on the evaluation of these digital screens can evaluate the potential benefits and costs across ‘efficacy’ (i.e. under controlled circumstances) and ‘effectiveness’ (i.e. real-world conditions) trials (Singal et al., 2014). Although there have been reviews (Hassem and Laher, 2019; Iragorri and Spackman, 2018) and a number of research endeavors looking at the (i) accuracy/efficacy (Cronly et al., 2018; Fitzsimmons-Craft et al., 2019; McDonald et al., 2019), (ii) user demographic characteristics (Choi et al., 2018; Rowlands et al., 2015), (iii) user levels of engagement (King et al., 2015), and/or (iv) user opinions and preferences of mental health digital screening tools as they apply to more demographically and clinically homogeneous populations (Batterham and Cleave, 2017; Drake et al., 2014; Pretorius et al., 2019), there is no known research to date which has evaluated the impact of these screens within large and more generalizable populations. Many mental health advocacy sites provide mental health screens to the public based on these efficacy studies. However, broader-scale effectiveness studies are needed to further assess the naturalistic impact of these digital screens on participants’ help-seeking behaviors and mental health. It is these large-scale effectiveness studies on whole populations that will afford researchers the ability to study both information seeking and help seeking behaviors within the context of day-to-day life.

This study aimed to address the current lack of research in the effectiveness of digital screens in large, generalizable populations. Specifically, the goal of the current work was to examine the impact and qualities of widely used, freely available online mental health screening on potential benefits, including (1) further mental health information seeking (i.e. searching for information related to one’s mental health conditions or problems), (2) mental health treatment seeking (i.e. later searching for psychiatrists), and (3) potential indicators of receipt of care (e.g. searches for specific psychoactive medications, may be an indicator of received mental treatment (Pretorius et al., 2019) and has been validated in prior research examining the impact of adverse drug reactions (Yom-Tov and Gabrilovich, 2013), as well as potential risks, including (4) suicidal ideation (i.e. searching about thoughts related to suicide and/or death, e.g. “I have suicidal thoughts”) (Jacobson et al., 2020b), and (5) active suicidal intent (i.e. information-seeking about suicide methods, e.g., “how to commit suicide”). Given the goals of the

current work, we employed methods to both predict these future mental health outcomes using online screen characteristics within a machine learning framework, as well as models to examine the differential impact of a given screen characteristic on the likelihood of a later mental health outcome (i.e. Cox Proportional Hazards Models). Thus, our research questions were as follows:

- (1) Are the characteristics of digital mental health screens predictive of future mental health information seeking, mental health treatment seeking, proxies of receipt of care, suicidal ideation, or active suicidal intent?
- (2) What characteristics of digital mental health screens are associated with the differential likelihood of future mental health information seeking, mental health treatment seeking, proxies of receipt of care, suicidal ideation, or active suicidal intent?

1. Methods

1.1. Participants and search data

We extracted data from all users who made queries related to mental health screening tools within the United States to the Microsoft Bing search engine (see Fig. 1). Gender and age information was provided by users at the time of registration to Bing. Users were identified using a cookie. Data for the study consisted of queries by users who clicked a link to one of the identified mental health screening tools during their searches and maintained only queries by those users, between December 1st, 2018 and January 31st, 2020. Bing users are a representative sample of internet users in the United States (Rosenblum and Yom-Tov, 2017).

The analysis described herein was approved by the Behavioral Sciences Research Ethics Committee of the Technion, approval number 2018–032. For each query through January 31st, 2020, we extracted an anonymous user identifier, its time and date, the query text, the list of pages displayed to the user, and whether they were clicked. Each query was labeled into one or more of 62 topics (e.g., shopping, tourism, and health) using a proprietary classifier.

1.2. Mental health screening tool identification

Mental health screening tools were found by examining pages shown in response to queries containing the list of searches shown in Appendix A. Any web link shown 5 or more times during the data period was examined by the authors to verify if it was a screening tool. Screening tools were grouped according to their topic (i.e., depression, anxiety).

1.3. Codification of websites

Each website obtained through select search queries was manually reviewed. Presence of the feature quality was coded as a ‘1’, while a score of ‘0’ represented an absence of the feature quality. The features of the website/screen search result include the presence/absence of a/an (i) screen on that page, (ii) link to a screening tool on that page, (iii) provided suggestive diagnosis upon completion of the screen, (iv) accompanying message with details of the diagnosis, (v) a provided form of internal digital treatment, (vi) referral to facilitate in-person patient content, (vii) referral to digital treatment, (viii) digital treatment including human contact, and (ix) referral to crisis services. Raters completed each screen with the most severe clinical profiles possible (e.g. rating the degree of suicidality as “all the time”). To control for observer bias, all features of the website/screen search results were evaluated by two individuals separately and Cohen’s Kappa was calculated.

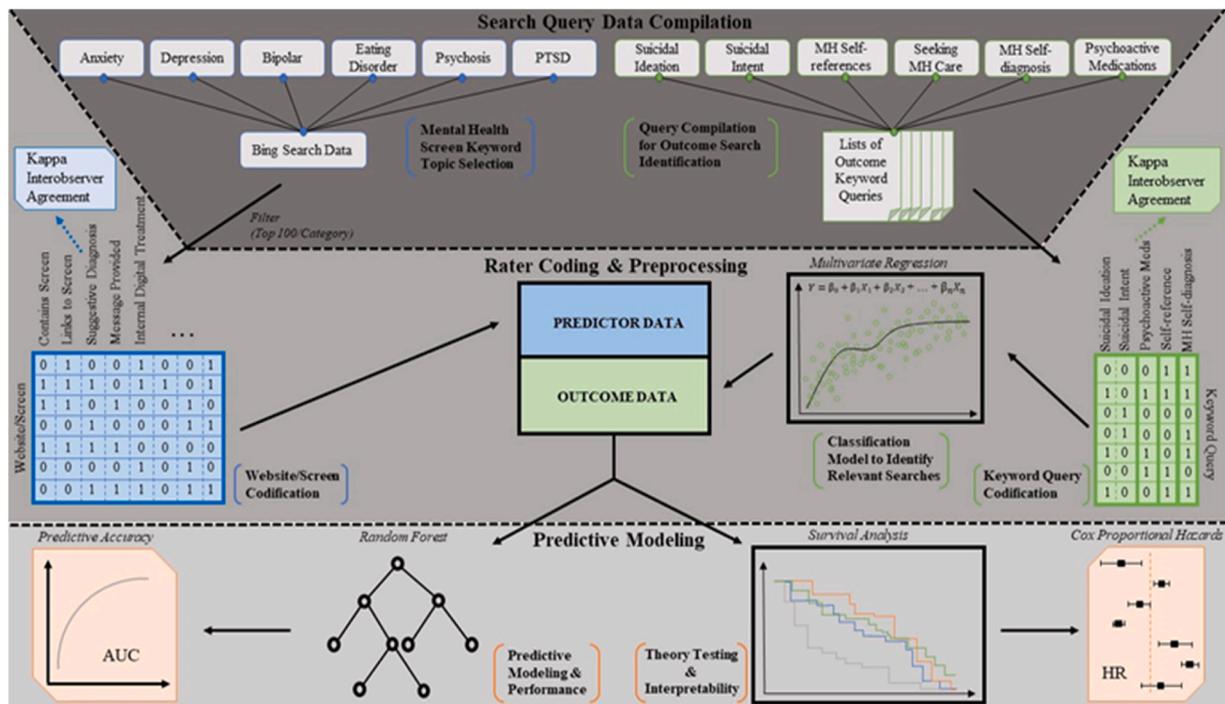


Fig. 1. Data collection and analysis pipeline.

1.4. Mental health outcomes: query classes and query labeling

Our process of identification and refinement of outcomes proceeded in three stages: (1) query compilation for outcome identification, (2) paired independent rater coding of the outcomes, and (3) using a classification model to form the respective outcomes.

Mental Health Query Compilation. Our primary outcomes consisted of search queries of several topics: (1) mental health self-diagnosis, (2) mental health self-references, (3) seeking mental health care, (4), suicidal intent, (5) suicidal ideation, and (6) psychoactive medications. To identify outcomes, seed keywords were first used for each given outcome (see Table 1). The goal was to be over-inclusive and then to further discriminate the outcomes using independent raters and predictive models (as described below).

Paired Independent Rating of Common Queries for each Respective Outcome. Two independent raters rated the most common queries from each of the respective outcomes with the goal of filtering only those that were the representative of the intended query outcome. Due to the large number of potential queries, and to preserve participants' privacy, less common search queries were then evaluated by a predictive model trained using data from the independent raters.

Query Labeling using Predictive Models. Utilizing a multivariate regression model to filter irrelevant queries from each outcome except for the mental health care seeking category, the attributes for the model were: (1) word and word pairs appearing in the query, (2) the span of time of queries and their appearance counts, (3) the number of users who used the query, and (4) whether the query included one of the following psychiatric conditions: depression, anxiety, BPD, paranoia, or schizophrenia.¹ These derived classes were then used as the model outcomes. Classifier accuracy was assessed with the Area Under Curve

¹ Note that the queries related to psychiatric conditions were utilized to capture lay-language of the following psychiatric conditions: (1) major depressive disorder as depression, (2) generalized anxiety disorder, social anxiety disorder, or panic disorder as anxiety, (3) borderline personality disorder as BPD, (4) paranoid personality disorder as paranoia, and (5) schizophrenia as schizophrenia.

(AUC) metric of the labeled queries using 10-fold cross-validation.

1.5. Analysis

Machine Learning Models. We predicted the future behavior of a user as a function of each of their visits to online web screens. The independent variables included: (1) screening tool topic (i.e. anxiety, bipolar, depression, eating disorder, psychosis, post-traumatic stress), (2) screening tool attributes, (3) hour of the day and day of the week at which the screening tool was clicked, (4) whether the screening tool was a Mental Health America (MHA, mhanational.org) screening tool or from another online web domain, and (5) number of previous searches which resulted in a click to a screening tool. Additionally, in secondary models, we added (6) past interests to predict future behavior as represented by the distribution of query topics prior to the first screening tool click by each user to ascertain whether the online screen information provided incremental information to their general search pattern types (e.g. health). Linear, linear with interactions, and random forest models with 100 trees were used to predict whether the user would make each of the target behaviors in future. These models used either variables 1–5, variable 6, or all independent variables. The models were evaluated using the AUC (Duda et al., 1973) and validated using 10-fold cross-validation.

Cox Proportional Hazards Models. Additionally, we utilized Cox proportional hazards models using independent variables 1–5 to test specific online screening attributes which were associated with the differential future search behavior of each mental health outcome (to test research question 2). The dependent variable was the time to each target behavior (if there were several instances of each, then this was based on the closest target behavior).

2. Results

2.1. Descriptive statistics, interrater agreement, and query labeling

A total of 130 screening tools were identified, and the total sample included 126,060 participants. Among users whose age and gender were

Table 1
Query labeling agreement, model accuracy, and sample queries of mental health outcomes.

Topic	Seed Keyword Query	Labeler agreement (Kappa)	Model accuracy (AUC)	Example queries
Mental health self-diagnosis	"I have" or "I've been diagnosed", but not the terms "do I have" or "if I have"	1.00	0.93	● I think I have depression
Mental health self-references	"I" or "my"	0.94	0.85	● I have anxiety
Seeking mental health care	"doctor", "clinic", "psychiatrist", "psychologist", "doctor", "hospital", "PCP", or "primary care"	0.85	0.86	● am i depressed quiz
Suicidal intent	"suicide" or "kill myself"	0.79	0.74	● do i have ptsd
Suicidal ideation	"suicide" or "kill myself"	0.85	0.83	● Psychiatrist near me
Psychoactive medications	index of drug names by trade name, use, and class provided in Stahl's Essential Pharmacology (Stahl et al., 2017); this includes both the generic or brand name of a psychiatric medication	0.95	0.99	● Psychologist near me
				● How to commit suicide
				● I want to kill myself
				● I have suicidal thoughts
				● I am suicidal
				● Fluoxetine
				● Zoloft

Note. This table contains the outcome topic categories, the seed keywords for each outcome, the interrater agreement, the model AUC for each outcome, and example queries from each outcome.

known, 72% were female. Average age was 38.5 years (range: 19–86, $SD = 14.0$). On average, 18,008 users ($SD = 22,040$) were found per screening tool topic (range: 2396–65,922, See Fig. 2). Regarding the interrater agreement of the website coding (see Table A1, Appendix B), Cohen's Kappa statistics suggested that there was high substantial agreement (0.71–0.78) between evaluators in the presence/absence of (i) a screen on that page, (ii) a diagnosis, and (iii) an accompanying message with details of the diagnosis. Although the agreement in referrals to digital treatment was fair (0.32), all other features had moderate agreement (0.49–0.56; median = 0.53). Query labeling agreement values, classification model accuracies, and examples of queries are illustrated in Table 1. As the table shows, high agreement was reached between labelers (>0.75 for all topics), and relevant queries could be identified using the filtering model with high accuracy (>0.7 for all topics).

2.2. Predicting future mental health outcomes based on online screening attributes

The result of predicting the outcomes using a Random Forest is shown in Fig. 3. We focus on Random Forest, since the linear models achieved lower accuracy across outcomes and attribute classes. Overall, prediction accuracies were high for mental health self-diagnosis and self-references, as well as seeking care. Other outcomes were more difficult to predict. Moreover, using both sets of attributes (screening tool parameters and user interests) attained the highest accuracy, but using only user interests reached almost the same accuracies for most outcomes. This suggested that past user interests are the most indicative

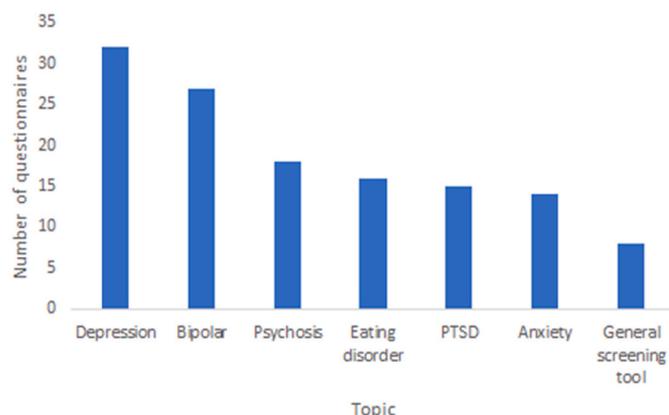


Fig. 2. Number of screening tools per topic.

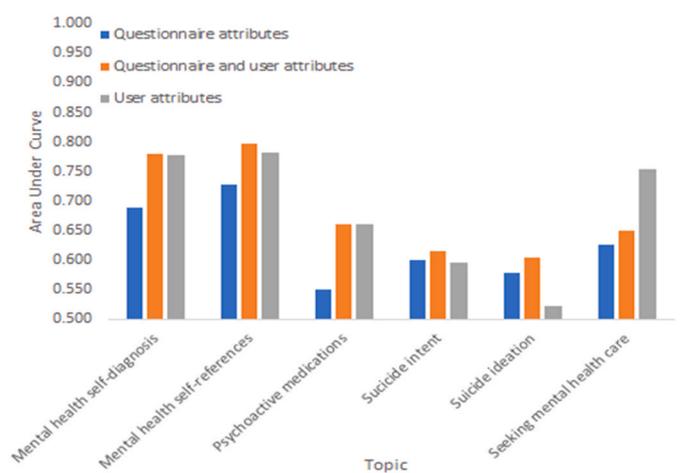


Fig. 3. Area Under Curve (AUC) for predicting respective outcomes. Independent attributes are the attributes of the screening tools, user attributes, or both.

of future outcomes, with the screening tool themselves having a relatively small effect. The exceptions to this observation were suicidal ideation, where screening tools had an extraordinarily large effect (though note the relatively low overall accuracy). For suicide intent, screening tool attributes were as predictive as user interests.

2.3. Specific online screening attributes predicting later mental health outcomes

The possible association between queries and the attributes of the screens selected by users were estimated by running, for each screening tool topic, a test between queries that appeared at least ten times and each of the attributes of the topic screens. None of the topic and attribute combinations demonstrated a statistically significant interaction ($p > 0.05$ after Bonferroni correction).

Table 2 shows the Cox proportional hazards model parameters for each of the outcomes. Future searches for self-diagnosis were more likely following the use of an anxiety, depression, or PTSD screening tool. Mental health self-references were associated with the use of anxiety, bipolar, depression, or eating disorder screening tools, as well as when the screening tool provides a message. Screening tools that suggested diagnosis were associated with a lower likelihood of future mental health searches. Psychoactive searches were associated with cases when the screening tools do not link to a screen, are not suggestive of diagnosis, or do not offer treatment online. Interestingly, suicidal intent was

strongly associated with screening tools that refer people to in-person care.

3. Discussion

The current study examined the ability to predict mental health outcomes based on exposure to online screening tools in a naturalistic sample of 126,060 participants. Supporting prior research examining the potential impact of screens on mental health outcomes (Batterham and Calear, 2017; Choi et al., 2018; Cronly et al., 2018; Drake et al., 2014; Fitzsimmons-Craft et al., 2019; King et al., 2015; McDonald et al., 2019; Pretorius et al., 2019; Rowlands et al., 2015), the results suggested that the type of content displayed on these online screens could independently predict future mental health outcomes. This evinces that the content of online screening tools has meaningful relationships to mental health outcomes, showing the largest impact on searches for mental-health self-references and mental health self-diagnosis. Nevertheless, the results also suggested that these models were capable of predicting the seeking of mental health care and potential indicators of psychotropic treatment. Surprisingly, the type of content on these online screens also predicted future searches for suicidal ideation and active suicidal intent. Taken together, these findings imply that online screens may heavily influence a variety of benefits (mental-health information seeking, seeking care, and care receipt), as well as risks (suicidal ideation and active suicidal intent).

Our results indicated that, overall, past search behaviors also predict outcomes. This is especially true for mental health diagnosis, self-reference, and seeking medical care. The hardest outcomes to predict were suicidal ideation and active suicidal intent. However, for these outcomes, the online screening tool attributes have the largest effect in improving prediction. This could suggest that the qualities of these online screens may produce the greatest changes in suicidal ideation and intent when an individual's level of baseline functioning and behavioral

characteristics are taken into account. This provides partial corroboration of the need to investigate the short-term and immediate effects of environmental risk for suicidal thoughts and behaviors (Franklin et al., 2017).

Notably, users who were directed to surveys that contained referrals to in-person care had a 71% higher likelihood of subsequent suicidal intent. This seems counterintuitive, as one would suspect that providing mental health resources of any sort would decrease suicidal intent. Although prior research has found referral periods to be critical in times of suicide risk (National Action Alliance for Suicide Prevention, n.d.), similar results have not been found in previous literature, and current literature on risk factors for suicide does not include in-person referrals (Steele et al., 2018). One might hypothesize a confounding effect accounting for such a finding; specifically, users with more severe illness who were more likely to have later suicide intent (regardless of intervention) were more likely to click on websites that issued referrals to in-person care. However, we have found that pre-existing data was not predictive of being exposed to website content, making this hypothesis less likely. Moreover, in-person referrals were most often provided as a static link on the screening tool website, and were not dynamically presented post-completion based on symptom severity. The referral initiation process may have been potentially overwhelming for some persons, thereby triggering decompensation. More research is needed to further address these issues (Jacobson et al., 2020a). We also found that providing suggestive diagnoses decreases the likelihood of mental health self-references searches by 22% for any given later time point, and decreases the likelihood of searching for psychoactive medications by 8% at any given later time point. Partially supporting prior work on feedback severity (Batterham et al., 2016), providing a suggestive diagnosis may thus be potentially harmful to information-seeking and the potential receipt of in-person treatment. It is possible that providing a diagnosis from online screens is too stigmatizing and encourages a pattern of avoidance behaviors (Morgan et al., 2012). An alternative

Table 2
Cox-proportional hazards estimates of screening content predicting mental health outcomes.

Attribute	Mental health self-diagnosis	Mental health self-references	Psychoactive medications	Suicide intent	Suicide ideation	Seeking mental health care
Anxiety screening tool	1.485*	1.241*	1.074	1.259	1.297	1.109
Bipolar screening tool	1.482	1.245*	1.098	1.543	0.022	0.827
Depression screening tool	1.486*	1.240*	1.084	1.169	0.163	1.045
Eating disorder screening tool	1.496	1.341*	1.116	0.986	0.114	4.473*
Psychosis screening tool	1.268	1.118	1.074	1.270	0.562	1.224
PTSD screening tool	1.958*	1.162	1.097	1.220	0.405	1.416
Does this page contain screen?	0.752	0.952	0.869*	0.656	0.337	0.692
Does this page link to screen?	1.137	1.076	1.043	1.055	0.034	0.807
Suggestive diagnosis?	0.912	0.776*	0.916*	0.978	11.363	1.157
Is message provided?	1.267	1.397*	1.210	0.845	1.131	0.767
Internal digital treatment?	0.601	0.558	0.726*	0.361		0.171
Referrals to facilitate in-person patient care?	1.038	0.983	1.021	1.727*	4.569	0.716
Referrals to digital treatment (other than crisis referral)?	0.981	1.06	1.078	1.478	12.359	2.537
For digital treatment referrals, did referral involve human?	1.031	1.052	0.971	0.591	0.263	1.016
Referral to crisis services?	1.044	0.958	1.024	1.425	15.328	1.146
Time of the day	1.000	1.005	1.002	1.004	1.045	0.996
Day of the week	0.988	1.003	0.991*	1.061	1.124	0.965
MHA screener?	0.772	0.932	0.871	0.868	0.022	1.007
Previous searches	1.059*	1.003	1.005*	0.496		1.177

Note. This table depicts the hazard ratios for each respective outcome. Note that the hazards ratios reflect the effect of the given predictor on the outcome across time, such that hazards reflect the chance that the participant would search for a given mental health outcome at a later time. Each separate column depicts a separate outcome (mental health self-references, mental health self-diagnosis, seeking mental health care, psychoactive medications, suicidal intent, and suicidal ideation). The rows of this table each reflect the different multivariate predictors. The primary effects of interest were the screening content. For each outcome, we also controlled for the screening tool type (i.e. anxiety, bipolar, depression, eating disorder, psychosis, and posttraumatic stress disorder), the time of the day, the day of the week, the rate of prior searches for each outcome, and whether the screening tool was from the Mental Health America (the mental health screening domain with the largest user base). Stars denote statistically significant interactions ($P < 0.05$, with Bonferroni correction). Note that blank rows were removed when they were invariant for non-censored outcomes.

explanation of the current results is that users may find sites that provide a suggestive diagnosis as more definitive and therefore may feel less need to search for mental health terms. Screens providing a message increases the likelihood of searches with mental health self-references at any later time point by 39.7%. Messages may be important in providing context in how to interpret information, thereby enabling more thorough follow-up and subsequent information-seeking regarding their own mental health.

The study has strength in its large naturalistic sample constituency, allowing for population-level research with strong ecological validity. Nevertheless, this focus on large ecological validity comes at the expense of greater internal validity, meaning that the current study raises many questions which should be further interrogated in future endeavors using large, naturalistic, randomized controlled trials on populations. Regarding the query labeling of predictive models, we utilized a range of terms to index psychiatric conditions, but these terms were not exhaustive and could have included other lay descriptions of psychiatric conditions, including prominent symptoms (e.g., “worry”, “hallucinations”, “sadness”). Another limitation of the current study was that it was not possible to determine whether an individual actually completed a screen, and what feedback the user might have received. Thus, the effects observed could be influenced by other unknown qualities of the screening websites themselves that were not known. It is important to note that, although the directionality of the results is implied, causality cannot be ascertained given the lack of experimental manipulations. Additionally, some of the AUCs were marginally better than chance (e.g. questionnaire attributes alone predicting psychoactive medication). Our work does not directly assess the mental outcomes of the users, but rather uses search behaviors as a surrogate for potential mental health

outcomes. Consequently, further research is required to examine the potential to test these outcomes in large randomized controlled trials with longitudinal behavioral follow-up periods.

Conflicts of interest

The authors have no conflicts of interest to declare.

CRediT authorship contribution statement

Nicholas C. Jacobson: Writing - review & editing, contributed to the initiation of study design, feature codification, literature searches, results interpretation, manuscript write-up, and manuscript review. **Elad Yom-Tov:** Formal analysis, contributed to the study design, data analysis, result interpretation, figure creation, and manuscript write-up. **Damien Lekkas:** Writing - review & editing, contributed to literature searches, feature codification, figure creation, and manuscript write-up, and manuscript review. **Michael Heinz:** Writing - review & editing, contributed to feature codification, results interpretation, and manuscript write-up and review. **Lili Liu:** contributed to feature codification, and manuscript write-up. **Paul J. Barr:** Writing - review & editing, contributed to the study design, manuscript write-up, and manuscript review.

Acknowledgement

This work was supported by the National Institute of Mental Health (R01 MH123482) and the National Institute of Drug Abuse (P30 DA029926).

Appendix A

Anxiety screening tool
Anxiety test
Assessment questions for psychosis
Bipolar screening
Bipolar test
Depression screen
Depression screening
Depression screening tool
Depression test
Do i have a mental illness
Eating disorder screening
Eating disorder test
Fmla and mental health
Fmla for mental health
Fmla for mental health reasons
Fmla mental health
How to get a service dog
How to obtain a service dog
Mental health america
Mental health assessment tools
Mental health fmla
Mental health screening
Mental health screening questionnaire
Mental health screening tools
Mental health symptom checker
Mental health symptoms diagnosis tool
Mentalhealthamerica.net
Mha screening
Mhascreening.org
Online depression screening
Partnership for prescription assistance
Psychosis questionnaire

Psychosis screening questionnaire
 Psychosis test
 Ptsd test
 Screening tool for psychosis.

Appendix B

Table A1
 Labeling Agreement of Website/Screen Features

Website/Screen Feature	Labeler Agreement (Kappa)
Does this page contain a screen?	0.74
Does this page link to a screen?	0.53
Does the screen provide a suggestive diagnosis?	0.78
Is a message provided with the diagnosis?	0.71
Does the page contain internal digital treatment?	0.49
Are there referrals to facilitate in-person care?	0.52
Are there referrals to digital treatment (other than crisis referral)?	0.32
Does referral to digital treatment involve a human?	0.51
Is there a referral to crisis services?	0.56

References

- Batterham, P.J., Calear, A.L., 2017. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. *JMIR Ment Health* 4, e26. <https://doi.org/10.2196/mental.7722>.
- Batterham, P.J., Calear, A.L., Sunderland, M., Carragher, N., Brewer, J.L., 2016. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. *BJPsych Open* 2, 67–73. <https://doi.org/10.1192/bjpo.bp.115.001552>.
- Bloom, D.E., Cafiero, E.T., Jane-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettnner, K., Rosenberg, L., Seligman, B., Stein, A.Z., Weinstein, C., 2011. *The Global Economic Burden of Non-communicable Diseases*. World Economic Forum, Geneva.
- Bose, J., 2017. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2017 National Survey on Drug Use and Health 124*.
- Carter, J.C., Olmsted, M.P., Kaplan, A.S., McCabe, R.E., Mills, J.S., Aimé, A., 2003. Self-help for bulimia nervosa: a randomized controlled trial. *Aust. J. Pharm.* 160, 973–978. <https://doi.org/10.1176/appi.ajp.160.5.973>.
- Choi, I., Milne, D.N., Deady, M., Calvo, R.A., Harvey, S.B., Glozier, N., 2018. Impact of mental health screening on promoting immediate online help-seeking: randomized trial comparing normative versus humor-driven feedback. *JMIR Ment Health* 5, e26. <https://doi.org/10.2196/mental.9480>.
- Cronly, J., Duff, A.J., Riekert, K.A., Perry, I.J., Fitzgerald, A.P., Horgan, A., Lehane, E., Howe, B., Chroinin, M.N., Savage, E., 2018. Online versus paper-based screening for depression and anxiety in adults with cystic fibrosis in Ireland: a cross-sectional exploratory study. *BMJ Open* 8, e019305. <https://doi.org/10.1136/bmjopen-2017-019305>.
- Drake, E., Howard, E., Kinsey, E., 2014. Online screening and referral for postpartum depression: an exploratory study. *Community Ment. Health J.* 50, 305–311. <https://doi.org/10.1007/s10597-012-9573-3>.
- Duda, R.O., Hart, P.E., Stork, D.G., 1973. *Pattern Classification and Scene Analysis*. Wiley, New York.
- Fitzsimmons-Craft, E.E., Balantekin, K.N., Eichen, D.M., Graham, A.K., Monterubio, G.E., Sadeh-Sharvit, S., Goel, N.J., Flatt, R.E., Saffran, K., Karam, A.M., Firebaugh, M.-L., Trockel, M., Taylor, C.B., Wilfley, D.E., 2019. Screening and offering online programs for eating disorders: reach, pathology, and differences across eating disorder status groups at 28 U.S. universities. *Int. J. Eat. Disord.* 52, 1125–1136. <https://doi.org/10.1002/eat.23134>.
- Franklin, J.C., Ribeiro, J.D., Fox, K.R., Bentley, K.H., Kleiman, E.M., Huang, X., Musachio, K.M., Jaroszewski, A.C., Chang, B.P., Nock, M.K., 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol. Bull.* 143, 187–232. <https://doi.org/10.1037/bul0000084>.
- Gulliver, A., Griffiths, K.M., Christensen, H., 2010. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatr.* 10, 113. <https://doi.org/10.1186/1471-244X-10-113>.
- Hassem, T., Laher, S., 2019. A systematic review of online depression screening tools for use in the South African context. *S. Afr. J. Psychiatr.* 25, 1–8. <https://doi.org/10.4102/sajpsychiatry.v25i0.1373>.
- Iragorri, N., Spackman, E., 2018. Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis. *Publ. Health Rev.* 39, 17. <https://doi.org/10.1186/s40985-018-0093-8>.
- Jacobson, N.C., Bentley, K.H., Walton, A., Wang, S.B., Fortgang, R.G., Millner, A.J., Coombs, G., Rodman, A.M., Coppersmith, D.L., 2020a. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. *Bull. World Health Organ.* 98, 270–276. <https://doi.org/10.2471/BLT.19.237107>.
- Jacobson, N.C., Lekkas, D., Price, G., Heinz, M.V., Song, M., O'Malley, A.J., Barr, P.J., 2020b. Flattening the mental health Curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. *JMIR Ment Health* 7, e19347. <https://doi.org/10.2196/19347>.
- King, C.A., Eisenberg, D., Zheng, K., Czyz, E., Kramer, A., Horwitz, A., Chermack, S., 2015. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. *J. Consult. Clin. Psychol.* 83, 630–636. <https://doi.org/10.1037/a0038805>.
- Magaard, J.L., Seeralan, T., Schulz, H., Brütt, A.L., 2017. Factors associated with help-seeking behaviour among individuals with major depression: a systematic review. *PLoS One* 12. <https://doi.org/10.1371/journal.pone.0176730>.
- McDonald, M., Christoforidou, E., Van Rijsbergen, N., Gajwani, R., Gross, J., Gumley, A. I., Lawrie, S.M., Schwannauer, M., Schultze-Lutter, F., Uhlhaas, P.J., 2019. Using online screening in the general population to detect participants at clinical high-risk for psychosis. *Schizophr. Bull.* 45, 600–609. <https://doi.org/10.1093/schbul/sby069>.
- Morgan, L., Knight, C., Bagwash, J., Thompson, F., 2012. Borderline personality disorder and the role of art therapy: a discussion of its utility from the perspective of those with a lived experience. *Int. J. Art Ther.* 17, 91–97. <https://doi.org/10.1080/17454832.2012.734836>.
- National Action Alliance for Suicide Prevention, n.d. *Best Practices in Care Transitions for Individuals with Suicide Risk: Inpatient Care to Outpatient Care*. Education Development Center, Inc.
- Pretorius, C., Chambers, D., Cowan, B., Coyle, D., 2019. Young people seeking help online for mental health: cross-sectional survey study. *JMIR Ment Health* 6, e13524. <https://doi.org/10.2196/13524>.
- Rosenblum, S., Yom-Tov, E., 2017. Seeking web-based information about attention deficit hyperactivity disorder: where, what, and when. *J. Med. Internet Res.* 19, e126. <https://doi.org/10.2196/jmir.6579>.
- Rowlands, I.J., Loxton, D., Dobson, A., Mishra, G.D., 2015. Seeking health information online: association with young Australian women's physical, mental, and reproductive health. *J. Med. Internet Res.* 17, e120. <https://doi.org/10.2196/jmir.4048>.
- Shidhaye, R., Lund, C., Chisholm, D., 2015. Closing the treatment gap for mental, neurological and substance use disorders by strengthening existing health care platforms: strategies for delivery and integration of evidence-based interventions. *Int. J. Ment. Health Syst.* 9, 1–11. <https://doi.org/10.1186/s13033-015-0031-9>.
- Singal, A.G., Higgins, P.D.R., Waljee, A.K., 2014. A primer on effectiveness and efficacy trials. *Clin. Transl. Gastroenterol.* 5, e45. <https://doi.org/10.1038/ctg.2013.13>.
- Stahl, S.M., Grady, M.M., Muntner, N., 2017. *Prescriber's Guide: Stahl's Essential Psychopharmacology, sixth ed.* Cambridge University Press, Cambridge, UK.
- Steele, I.H., Thrower, N., Noroian, P., Saleh, F.M., 2018. Understanding suicide across the lifespan: a United States perspective of suicide risk factors, assessment &

- management. *J. Forensic Sci.* 63, 162–171. <https://doi.org/10.1111/1556-4029.13519>.
- Thompson, A., Issakidis, C., Hunt, C., 2008. Delay to seek treatment for anxiety and mood disorders in an Australian clinical sample. *Behav. Change* 25, 71–84. <https://doi.org/10.1375/bech.25.2.71>.
- Wang, P.S., Berglund, P., Olfson, M., Pincus, H.A., Wells, K.B., Kessler, R.C., 2005. Failure and delay in initial treatment contact after first onset of mental disorders in the national comorbidity survey replication. *Arch. Gen. Psychiatr.* 62, 603–613. <https://doi.org/10.1001/archpsyc.62.6.603>.
- Webb, C.A., Rosso, I.M., Rauch, S.L., 2017. Internet-based cognitive behavioral therapy for depression: current progress & future directions. *Harv. Rev. Psychiatr.* 25, 114–122. <https://doi.org/10.1097/HRP.0000000000000139>.
- Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., Burstein, R., Murray, C.J., Vos, T., 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* 382, 1575–1586. [https://doi.org/10.1016/S0140-6736\(13\)61611-6](https://doi.org/10.1016/S0140-6736(13)61611-6).
- Wilhelm, S., Weingarden, H., Ladis, L., Braddick, V., Shin, J., Jacobson, N.C., 2020. Cognitive-behavioral therapy in the digital age: presidential address. *Behav. Ther.* 51, 1–14. <https://doi.org/10.1016/j.beth.2019.08.001>.
- Yom-Tov, E., Gabrilovich, E., 2013. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale Analysis of web search queries. *J. Med. Internet Res.* 15, e124. <https://doi.org/10.2196/jmir.2614>.