

Real-World Outcomes Associated with a Digital Self-Care Behavioral Health Platform

This article was published in the following Scient Open Access Journal:

Annals of Clinical Research and Trials

Received September 18, 2017; Accepted October 09, 2017; Published October 31, 2017

Krista Schladweiler^{1*}, Abigail Hirsch², Ed Jones³, and Luke B. Snow⁴

¹Vice President, Health Economics and Outcomes Research, myStrength Inc., 1875 Lawrence Street, Suite 550, Denver, CO 80202 USA

²Chief Clinical Officer, myStrength, Inc., 1875 Lawrence Street, Suite 550, Denver, CO 80202 USA
³Vice President, Strategic Planning, myStrength, Inc., 1875 Lawrence Street, Suite 550, Denver, CO 80202 USA

⁴Business Intelligence Developer, myStrength, Inc., 1875 Lawrence Street, Suite 550, Denver, CO 80202 USA

Abstract

Objective: To evaluate the real-world effectiveness of managing depression on the population-level using myStrength, a digital self-care behavioral health and well-being platform.

Methods: A series of effect size calculations were undertaken to estimate the impact of myStrength on symptom burden reduction among commercially-insured adults showing signs of depression. The depression scale of the DASS was periodically administered to myStrength users to measure self-reported change in depression symptom severity.

Results: A total of 2,138 commercially-insured symptomatic adults were included in the analysis, and 1,142 of them met the criteria for clinical depression at baseline, while the rest were at risk for depression. The myStrength estimated effect size for the at-risk population was 0.66 [95% CI: 0.60, 0.73]. Narrowing the population to focus exclusively on those users meeting the threshold for clinical depression, the myStrength estimated effect size increased to 1.02 [95% CI: 0.95, 1.13]. For both groups, the majority of symptom burden reduction occurred during the first 14 days of myStrength utilization.

Conclusions: Among a real-world, commercially-insured, adult population with some degree of depression, myStrength was shown to have a comparable effect size to that of traditional psychotherapy. Depression management may benefit from expanding treatment offerings to include digital platforms.

KeyWords: Depression, Digital behavioral health, Effect size, Health promotion

Introduction

The literature suggests that about one of every five adults aged 18 or older in the United States (18.5%, 43.8 million) struggled with a psychological or behavioral health condition in the past year [1]. Taking into account the natural course of depression, decades of research have shown that symptoms can last anywhere from two months to multiple years, with the average depressive episode persisting for five to six months [2]. While 70-85% of adults with depression experience periods of remission, depressive episodes can re-occur, especially among individuals with greater severity and co-occurring conditions [2,3]. According to the Centers for Medicare & Medicaid Services-maintained National Health Expenditure Accounts, mental health concerns such as anxiety and depression are the costliest medical conditions to treat, exceeding \$200 billion in 2013 [4].

Research has shown that less than half of adults dealing with behavioral health conditions seek treatment or receive adequate treatment [5]. Often those adults with the most-debilitating conditions, such as severe depression, are the ones who forgo treatment [6]. Untreated/under-treated psychological distress also carries significant direct and indirect costs associated with disability [7,8], lost workplace productivity [9], and increased total healthcare utilization and costs [10].

myStrength is an engaging and empowering self-care platform designed to help close the behavioral healthcare treatment gap [11]. Developed to fulfill unmet consumer needs, extend access, and improve outcomes, myStrength delivers evidence-based care available from anywhere 24/7/365. By making Cognitive Behavioral Therapy (CBT) available to users in a digital format, myStrength is increasing access to this well-researched and widely-adopted approach to addressing psychological distress and mental well-being [12]. CBT has even been shown to be a viable alternative

*Corresponding author: Krista Schladweiler, Vice President, Health Economics and Outcomes Research, myStrength, Inc., 1875 Lawrence Street, Suite 550, Denver, CO 80202 USA, Tel: 302-983-2954, Fax: 720-452-2980, E-mail: kschladweiler@mystrength.com

to antidepressant medication [13]. By applying the same CBT techniques used during face-to-face therapy sessions, myStrength empowers users to become more aware of how they feel and equips them with the skills to address their behavioral health needs on their own or in conjunction with a therapist.

As both researchers and healthcare providers strive to evaluate population health interventions, such as myStrength, a combination of Randomized Controlled Trials (RCT) and real-world analyses are required to close the efficacy-effectiveness gap [14]. RCTs are the methodological “gold standard” to measure treatment efficacy, given their high internal validity, leading to precise and accurate estimates. However, studies with high internal validity often lack generalizable findings, leading to low external validity. Real-world evidence generated from observational, retrospective studies, addresses this concern by quantifying outcomes in a naturalistic setting among actual users with varying sociodemographic and clinical characteristics, and most importantly, real-world adherence rates.

myStrength has previously evaluated the efficacy of its depression program. The RCT study findings validated the platform’s favorable and rapid impact on clinical outcomes [15]. Building on those research findings, myStrength sought to evaluate the effectiveness of the depression program compared to the existing standards of care for symptom management using an effect size model. Effect size ratios normalize the impact of an intervention on a desired outcome, allowing for an “apples to apples” comparison. In this analysis, we will compare the magnitude of real-world myStrength outcomes to effect sizes reported in the literature for outpatient psychotherapy and pharmacotherapy.

When evaluating the value of new resources such as digital tools, it is important to have a benchmark for comparison for how the tool is functioning in the real world relative to the other best available options. Psychotherapy outcomes are one appropriate benchmark as we have six decades of research for comparison. A landmark summary of this research, *The Great Psychotherapy Debate*, published in two editions in 2001 and 2015, found that psychotherapy, measured across many clinical models of intervention involving populations at-risk of depression, achieves, on average, a large effect size of 0.8 [16]. We can also benchmark against antidepressant medication interventions, especially for populations with more severe depression where pharmacotherapy may be most effective and has been most studied.

Our first goal is to assess, in a real-world context, what effect size can be expected from myStrength for two populations: a symptomatic/treatment-recommended population and a clinical depression/treatment strongly-encouraged population. A secondary goal is to investigate average symptom improvement in relation to the length of time engaging with the myStrength platform.

Methods

This observational, longitudinal study was designed to quantify the normalized effect size achieved for commercially-insured adults using myStrength’s self-help resources. myStrength is a web and mobile-based behavioral health platform, responsive to individual users’ interests and areas of focus. Grounded in evidence-based approaches, myStrength offers a personalized population health approach to managing depression, among

other behavioral health concerns, as well as to enhancing overall well-being. myStrength resources include, but are not limited to: computerized CBT programs, mood trackers, mindfulness exercises, sharing of community and personal inspirations, as well as a searchable library of over 1,600 mental health and wellness/well-being resources. myStrength users can go at their own pace and/or use the platform under the guidance of a mental health professional.

In addition to the myriad of resources available, myStrength users also complete symptom severity assessments when registering to use the platform (baseline) and are prompted to complete repeat assessments periodically over time. For the purposes of this analysis, the assessment of interest was the depression subscale score (range 0 to 42) of the Depression, Anxiety, and Stress Scale 21 (DASS). The DASS is a reliable and validated 21-question self-report scale designed to measure symptom severity [17]. Users with baseline depression scores greater than or equal to 10 were classified as having symptomatic depression that may benefit from treatment. Users with baseline depression scores exceeding 20 were deemed to meet the criteria for clinical depression where treatment would be strongly recommended.

The study population was sourced from two large, nationally-representative commercial insurers in the United States. The analysis time period spanned from January 1, 2014 to January 31, 2017. To be eligible for the study, myStrength users had to be 18 years of age or older, registered for myStrength during the study time period, and must have completed a baseline and at least one follow-up depression assessment. If users completed multiple follow-up assessments, the last depression score was analyzed.

The difference between baseline and last assessment quantified the change in symptom severity score. The effect size was calculated as the mean change in depression score divided by the pooled difference in standard deviation from baseline to last assessment (Cohen’s d).

Statistical Analysis

Data analyses were conducted using STATA v14.2 [18]. Descriptive statistics were performed to understand baseline characteristics of the study population. Variables of interest included age, gender, and myStrength usage. Cohen’s d was calculated for the entire symptomatic population and for the sub-population meeting the criteria for clinical depression. In addition, DASS depression scores were plotted by duration of use of the platform to last assessment available to look at how length of site usage impacted outcomes.

Ethics

Approval for the observational study has been reviewed and granted by Solutions IRB Institutional Review Board #2017/03/1. myStrength data storage is compliant with HIPAA guidelines.

Results

Symptomatic depression population

Of the 2,138 myStrength users identified as being symptomatic and possibly benefiting from treatment, 76.6% were female, with a mean age of 44.6 years, while the 45-54 age group represented the largest concentration of symptomatic users at 31.7% (Table 1). As shown in Table 2, these users accessed myStrength an

	Symptomatic (n = 2,138)		Clinical Depression (n = 1,142)	
	User Count	Percent/Mean	User Count	Percent/Mean
Gender (Female)	1,633	77.1%	858	76.6%
Gender (Male)	476	22.5%	255	22.8%
Gender (Unspecified/Other)	10	0.5%	7	0.6%
Average Age	2,119	44.6 years	1,130	43.8 years
18-24	82	3.9%	52	4.6%
25-34	392	18.5%	242	21.4%
35-44	500	23.6%	252	22.3%
45-54	672	31.7%	342	30.3%
55-64	432	20.4%	216	19.1%
65 and over	41	1.9%	26	2.3%

Table 1: Descriptive characteristics of myStrength users with some level of depression.

	Symptomatic (n= 2,138)	Clinical Depression (n = 1,142)
Average baseline depression score (median)	23.4 (22.0)	30.4 (30.0)
Standard deviation	8.9	6.2
Average last depression score (median)	16.5 (14.0)	20.1 (18.0)
Standard deviation	11.8	12.4
Paired t-test (mean, [95% CI])	6.97, [6.48-7.45]	10.24, [9.55-10.94]
p-value	<0.0001	<0.0001
Effect size (d, [95% CI])	0.66, [0.60, 0.73]	1.02, [0.95, 1.13]
Average logins in first 30 days	4.0	4.1
Average total study period logins	11.4	11.5

d = Cohen's d; CI = confidence interval.

Table 2: Baseline and last assessment depression scores and myStrength utilization data.

average of 4 times during their first 30 days of use, the period during which the most significant symptom change often occurs [15]. Over the study period, users logged in to the platform an average 11.4 times.

The baseline mean depression score was 23.4, corresponding to severe depression according to the DASS rating system (median = 22; standard deviation = 8.9). At last assessment, the average depression scores decreased to 16.5, thereby reducing depression symptoms to moderate intensity (median = 14; standard deviation = 11.8; $p < 0.0001$). The mean myStrength effect size for this population who would likely benefit from treatment was 0.66 [95% CI: 0.60, 0.73].

Clinical depression population

Stratifying the population to focus on those users meeting the criteria for clinical depression (n=1,142), the demographic profile was largely consistent with that of the symptomatic, treatment-recommended population (Table 1). myStrength users with clinical depression exhibited similar utilization and satisfaction patterns to the total symptomatic population (Table 2).

At baseline, this subgroup met the DASS criteria for extremely severe depression, on average, with an initial depression score of 30.4 (median = 30; standard deviation = 6.2). At last assessment, the average depression scores decreased to 20.1, demonstrating even greater symptom burden reduction among users with clinical depression for whom treatment would be strongly encouraged (median = 18; standard deviation = 12.4; $p < 0.0001$). The mean myStrength effect size for users with likely clinical depression was 1.02 [95% CI: 0.95, 1.13].

Rate of improvement

Figure 1 displays the average depression scores during the

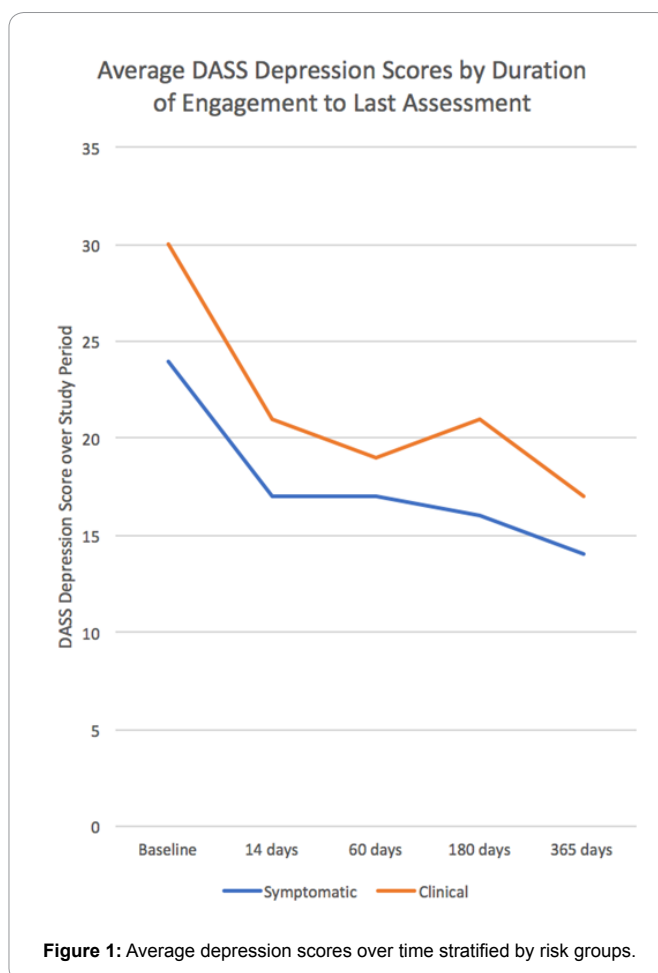


Figure 1: Average depression scores over time stratified by risk groups.

study period stratified by risk population. Users were prompted to complete assessments at baseline and subsequently on/around days 14, 60, 180 and 365. Both the symptomatic and clinical depression populations experienced the most significant decrease in depression scores within the first 14 days of using the myStrength platform. For users whose last available assessments were during this period, depression scores for symptomatic users decreased an average 7 points, generally reducing their severity category to moderate. Similarly, the clinical depression population was downgraded from the severe/extremely severe categories to being on the cusp of the moderate depression range, with only two weeks of platform usage. Both populations continued to experience further reduction in symptom burden over the course of the following weeks and months, but not to the same magnitude.

Limitations

These study findings must be interpreted in the context of several limitations. Given the retrospective, observational nature of the study, the analysis was limited to the variables already collected. The statuses of potentially confounding covariates such as concurrent psychotherapy and/or antidepressant medication are unknown. However, the robust sample size is intended to minimize any heterogeneity with regard to these covariates. Building on these initial findings, future research will solicit and include concurrent behavioral health treatments to further evolve the measurement of myStrength effectiveness.

Study findings are generalizable to those myStrength users who completed a baseline and at least one follow-up depression assessment, and may not be representative of the effect experienced by more infrequent myStrength users. These parameters were not only necessary to facilitate the calculation of effect size, but were appropriate given this study's focus on the population of active users of the myStrength platform.

The analysis universe is limited to commercially-insured adults, and therefore additional research is needed to quantify the myStrength effect size on different populations (i.e. adolescents, Medicaid and Medicare recipients).

Discussion

This observational, real-world study was designed to accomplish two main goals. The first goal was to provide documentation of real-world effect sizes, for a population of digital behavioral health users with depression. In doing so, we facilitate benchmarking across different behavioral health management strategies. By choosing to calculate effect size, the methodological sophistication is moved beyond a vague estimate of "does the intervention work," to a standardized and well-accepted statistical model for understanding "how well does the intervention work?" [19].

Digital behavioral health platforms have significantly increased access to care and thus warrant empirical evaluation as to their effectiveness with respect to impacting the spectrum of symptom severity. myStrength was found to have a moderate to large effect on reducing depression symptom burden, with the greatest effect found among users for whom treatment would most likely be needed to manage their symptoms. By further-segmenting the study population to hone in on those with clinical

depression, we were able to show a marked increase in effect size correlating with increased symptom severity. Platforms deemed to have moderate to large effect sizes have the potential to positively impact the societal burden of depression [20].

In addition to demonstrating the robust effectiveness of myStrength, this research also accomplished the secondary goal of identifying the inflection point at which the greatest magnitude of change occurs. Having the ability to substantially reduce the duration of a depressive episode can have far-reaching ramifications in terms of quality of life, workplace productivity, and the likelihood of experiencing a future episode [3]. These real-world findings offer another proof point demonstrating myStrength's rapid improvement in depression symptom burden [15].

To leverage the benchmarking value of effect size and to put these findings into perspective, comparisons to the literature for alternative behavioral health treatments, such as psychotherapy and antidepressant medication, can be made. As noted in his groundbreaking meta-analysis, Wampold found that psychotherapy has a large effect size of 0.8 [16]. The myStrength intervention is 82.5% (0.66/0.8) as effective as psychotherapy for adults experiencing symptoms of depression. A more recent meta-analysis of 11 high-quality randomized controlled trials found that the psychotherapy effect size may be grossly overstated in the literature and more accurately estimated to be a much smaller effect size ($d=0.22$) [21]. In either case, digital tools would seem to be worth analyzing within the research literature alongside traditional care to understand whether this is an effective modality.

Similarly, myStrength effect size findings can be compared to the literature on antidepressant medication. The clinical depression population identified in this research study includes individuals who would likely be offered antidepressant medication to manage more severe symptoms, and thus makes for a more "apples to apples" comparison to antidepressant medication effectiveness. The psychopharmacology literature spans the continuum of touting the relative favorability of antidepressant medication for moderate ($d = 0.53$) and severe depression ($d = 0.81$) to negating those more-robust findings as being overstated and highly-biased ($d = 0.3$) [22-25]. Some research even suggests that the effect size for antidepressant medication is largely driven by the placebo effect [26]. Therefore, a clear point of comparison with the myStrength results is difficult to establish. While antidepressant medication has its place for adults dealing with significant symptom burden, this work suggests that enhancing treatment with a digital program may help to bring immediate relief.

Digital behavioral health tools such as myStrength increase access to evidence-based self-care resources without the often-associated stigma. In this technological age when new digital tools are constantly being developed, it is even more important to demonstrate the real-world performance of these programs as compared to other treatment alternatives. Digital tools deemed to have robust clinical efficacy which can be observed on the population-level have the potential to revolutionize the way behavioral health care is delivered.

Acknowledgements

This research was funded by myStrength. AH and LBS

performed all analyses. KS and EJ drafted the manuscript with significant input from all authors. All authors reviewed and approved the manuscript for publication.

References

1. Substance Abuse and Mental Health Services Administration, Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings, NSDUH Series H-49, HHS Publication No. (SMA) 14-4887. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.
2. Lehmann HE. Clinical evaluation and natural course of depression. *J Clin Psychiatry*. 1983;44(5 Pt 2):5-10.
3. Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. *J Affect Disord*. 2014;152-154:65-75.
4. Roehrig, C. Mental disorders top the list of the most costly conditions in the United States: \$201 billion. *Health Aff (Millwood)*. 2016;35(6):1130-1135.
5. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):629-40.
6. Pratt LA, Brody DJ. Depression in the U.S. household population, 2009-2012. NCHS Data Brief No 172 Hyattsville MD National Center for Health Statistics. 2014.
7. Katon W, Sullivan M, Russo J, Dobie R, Sakai C. Depressive symptoms and measures of disability: A prospective study. *J Affect Disord*. 1993;27(4):245-254.
8. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. *JAMA*. 1989;262(7):914-919.
9. The Impact Mental Health Has On Workplace Productivity. Telligen. 2016; 19.
10. Katon W, Von Korff M, Lin E, et al. Distressed high utilizers of medical care. DSM-III-R diagnoses and treatment needs. *Gen Hosp Psychiatry*. 1990;12(6):355-362.
11. myStrength.
12. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognit Ther Res*. 2012;36(5):427-440.
13. Driessen E, Hollon SD. Cognitive behavioral therapy for mood disorders: Efficacy, Moderators and Mediators. *Psychiatr Clin North Am*. 2010;33(3):537-555.
14. Daniel Blumenthal, Kristina Yu-Isenberg, John Yee, and Anupam Jena. Real-World Evidence Complements Randomized Controlled Trials In Clinical Decision Making. *Drugs and medical innovation*. 2017;27..
15. Hirsch A, Luellen J, Holder JM, et al. Managing depressive symptoms in the workplace using a web-based self-care tool: A pilot randomized controlled trial. *JMIR Res Protoc*. 2017;6(4):e51.
16. Wampold B, Imel Z. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*, Second Edition, Routledge, New York, 2015.
17. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol*. 2005; 44(Pt 2):227-239.
18. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
19. Coe R. It's the effect size, stupid: What effect size is and why it is important. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 2002:12-14.
20. Fairburn CG and Patel V. The impact of digital technology on psychological treatments and their dissemination. *Behav Res Ther*. 2017;88:19-25.
21. Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychol Med*. 2010;40(2):211-223.
22. Woo J-M, Kim W, Hwang T-Y, et al. Impact of depression on work productivity and its improvement after outpatient treatment with antidepressants. *Value Health*. 2011;14(4):475-482.
23. Vohringer PA, Ghaemi SN. Solving the antidepressant efficacy question: Effect sizes in major depressive disorder. *Clin Ther*. 2011;33(12):B49-B61.
24. Fournier MA, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA*. 2010;303(1):47-53.
25. Khan A, Brown WA. Antidepressants versus placebo in major depression: An overview. *World Psychiatry*. 2015;14(3):294-300.
26. Kirsch I. Antidepressants and the placebo effect. *Z Psychol*. 2014;222(3):128-134.