EHealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders

Burger, Franziska; Brinkman, W.P.; Neerincx, M.A.

Publication date
2018

Document Version
Final published version

Published in
annual review of cybertherapy and telemedicine

Citation (APA)
Burger, F., Brinkman, W. P., & Neerincx, M. A. (2018). EHealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders. *annual review of cybertherapy and telemedicine*, *16*, 18-24.

EHealth4MDD: A database of e-health systems for the prevention and treatment of depressive disorders

Franziska BURGER[a,1], Mark A. NEERINCX[b], and Willem-Paul BRINKMAN[a]
[a] *Delft University of Technology, Netherlands*
[b] *TNO Perceptual and Cognitive Systems*

Abstract. To date, meta-analyses of e-mental health systems for major depressive disorder (MDD) have largely overlooked the technological side of interventions. This warranted the creation of an open access database, EHealth4MDD, for the systematic study of the technological implementation in relation to intervention content, study design, and study outcomes. E-health systems were identified by conducting an exhaustive search on PubMed, Scopus, and Web of Science in 2017. The 5379 retrieved records yielded 267 systems. One coder extracted information from the records on 45 variables, organized into 14 tables in EHealth4MDD. A sample of each high-inference variable was double coded by a second coder to assess reliability. Percent agreement was satisfactory given that coders received no training and the number of possible categories was large. Furthermore, scales were developed to rate the degree of technological sophistication of system functions for each of five function types. Four of these scales demonstrated concurrent validity, as evidenced by the substantial to strong correlations observed when comparing the scales with the results of an unlabeled ordering task. For researchers in both computer science and clinical psychology, the database presents a useful tool to systematically study e-mental health interventions for depression.

Keywords. e-mental health systems, major depressive disorder, systematic review tool, relational database, reference database, information retrieval, clinical psychology, software engineering

1. Introduction

Although depression can be treated effectively, more than half of the approximately 300 million people worldwide suffering from the illness are receiving inadequate or no treatment [1]. E-mental health presents a promising direction in overcoming many of the barriers to and shortcomings of face-to-face treatment [2].

To date, both systematic reviews and meta-analyses in the field have largely focused on delivery aspects (e.g. guidance as a factor of influence [3]) and therapy aspects (e.g. limiting the scope to a specific therapeutic approach [4]) in relation to outcomes while neglecting influences of technology. To the best of our knowledge, only Zhao et al. [5] systematically reviewed the presence of certain technological features (communication tools, instructional ICT features, and self-monitoring tools) in psychoeducational e-mental health systems for depression, reaching the conclusion that most interventions lack in technological sophistication.

[1] Franziska Burger, Interactive Intelligence Group, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands; E-mail: f.burger91@gmail.com.

As has been pointed out in recent surveys, the field would benefit from a clearer picture of the features of e-mental health systems for depression that contribute to outcomes and those that are superfluous [2]. This warranted the creation of an open access database of e-mental health systems for the treatment and prevention of major depressive disorder (MDD), enabling the systematic examination of the composition of these systems and how this relates to their evaluation context and dropout rates.

2. Method

2.1 Search Strategy

Considered for inclusion was primary research published in English describing e-mental health support systems, i.e. interventions with therapeutic content delivered on information and communication technology platforms, for the prevention or treatment of major depressive disorder in adults developed and evaluated between 2000 and 2017. To ensure quality, only support systems having been scientifically evaluated with end users were included. The time frame was lower-bounded to give an accurate overview of the state-of-the-art: in earlier systems the technological sophistication is limited by the availability of technology rather than being a design choice. Additionally, we excluded research on systems (1) serving only as a medium between therapist and user or between users (2) targeting children, targeting women with postpartum, perinatal, or prenatal depression, targeting caregivers or family members of people with depression, targeting a comorbid psychotic condition, (3) aiming to reduce stigma associated with depression, (4) diagnostic tools or decision aids, (5) lacking psychotherapeutic functionality (e.g. only supporting adherence to antidepressant treatment), (6) having a very narrow scope (i.e. system developed for one specific patient with very specific combination of conditions). The full search-query is given on the database website (see Footnote 2 below).

An exhaustive search was conducted for articles and papers published up until 2017 describing eHealth interventions for depression on Scopus, PubMed, and Web of Science. A total of 5379 records were retrieved from the three databases (5359) and research syntheses in the field (20). All records were filtered first on title, then on abstract, and finally on full article by the first author, C1. Due to resource restrictions, a sample of the records at each stage was double coded by a second, independent coder, C2, a computer science student. Table 1 presents sample sizes and agreement scores at each stage. Sample sizes were determined by trading off the available resources against the margin of error of the percent agreement using the methods proposed by Gwet [6].

2.2 Data extraction

Included articles were coded by C1 on a total of 45 variables and entered into a relational database. This database consists of 14 tables grouped into three larger clusters. The *systems* cluster details the e-mental health systems, their versions, their functionality, and their therapeutic purposes (four tables). In this cluster, systems were characterized on a macro-level (year of completion; whether its purpose is to prevent, treat, or monitor; whether it is guided, unguided, or an adjunct to face-to-face therapy; etc.) and on a micro-level, i.e. their functions. An instantiation of a function is its concrete implementation in a system. Functions were of two types: support functions and intervention functions. Support functions aim at increasing adherence to the intervention. They were again categorized into four subtypes: support functions for treatment planning (e.g. scheduling of sessions), treatment execution (e.g. reminders), monitoring (e.g. monitoring of symptoms), and social support (e.g. therapist support). To determine possible support functions in the domain, adherence strategies as defined in, for example, [7] were considered. Intervention functions support patient activities aimed at reducing depression symptoms. They are linked to specific classes of therapeutic interventions. The second cluster of tables in the database is the *evaluations* cluster, detailing the empirical studies of the systems in the systems cluster, their design, the employed measurement instruments, and dropout rates per study arm (four tables). Finally, the *publications* cluster details articles describing systems and their evaluations and the authors of these articles (three tables). The remaining three tables link systems to evaluations, systems to publications, and publications to evaluations.

Of the 45 variables, 41 were identified as low-inference (e.g. intended duration of the intervention) as they could be extracted directly from the literature, and 4 were identified as high-inference, since extraction required interpretation of the literature. The four high-inference codes were (1) identifying functions in a system description (e.g. SuperBetter implements the intervention function *3 good things*), (2) classifying an instantiation as a particular function (e.g. “Participants were instructed to follow one module per week” describes *Tunneling*), (3) linking an intervention function to a therapeutic intervention class (e.g. *3 good things* is a technique in Positive Psychology), and (4) assigning a degree of technological sophistication to an instantiation. For each of the high-inference variables, a random sample was selected and double coded by a second coder to assess reliability.

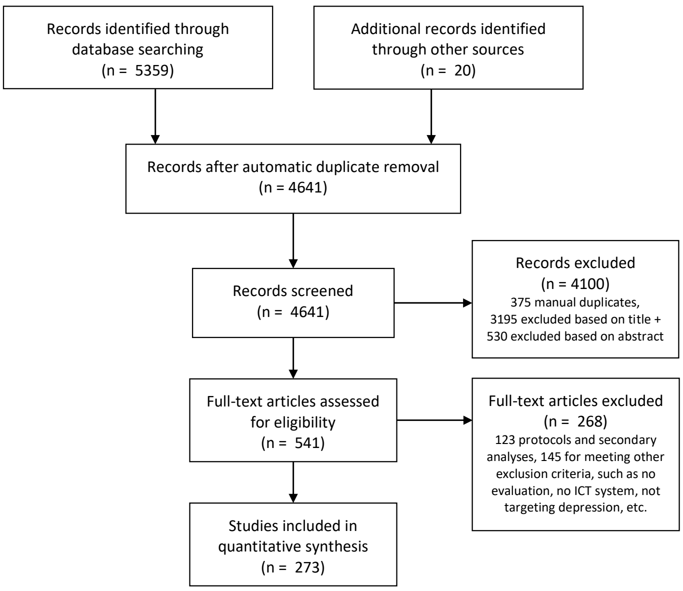

Figure 1. PRISMA diagram detailing the literature screening process of C1.

To *identify functions in a system description*, a graduate student in clinical psychology, C3, was provided with articles describing systems and the complete list of 184 functions. The percentage agreement was calculated as the percentage of overlap in all assigned functions of C1 and C3. To *classify an instantiation of a function*, C3 was again provided with the list of 184 functions as well as with a list of 125 snippets of text from articles describing functions. If C1 and C3 appointed the same function to the description of an instantiation, it was scored as agreement. In *linking a therapeutic intervention class to an intervention function*, C3 was provided with a list of all intervention functions and a set of 25 possible therapeutic intervention classes. While C1 always only selected one therapeutic intervention class, C3 was permitted to select multiple. Coder agreement was calculated by coding as agreement whenever C1's class was a subset of the class(es) assigned by C3. C3 received no training for the tasks other than a detailed coding manual. Furthermore, C1 *rated all instantiations of functions on their degree of technological sophistication* with the respective e-mental Health Degree of Technological Sophistication scale (eHDTS) developed specifically for this purpose[1]. Reliability was examined by regarding agreement in a sample that was recoded by C2 (Table 1).

Concurrent validity of the eHDTS scales was examined by correlating eHDTS scores assigned to instantiations of functions with knowledge of the scale levels (informed scores) with rank scores obtained when coders with computer science or related degrees were asked to sort the same instantiations according to their intuitive understanding of "technological sophistication" (naïve scores). Each of five coders, C4-8, received only instantiation descriptions of a specific category (e.g. intervention, treatment planning, monitoring). C2, however, received a large sample of component descriptions taken from all the categories to allow for the examination of comparability of the different eHDTS scales. At least one week after the naïve sorting, the coders were asked to rate the same sample component on their respective eHDTS. Spearman correlations were calculated to examine agreement between informed and naïve scores within coders (Intra-Coder Correlation, Table 2). All raw data and analysis scripts can be accessed online[2].

3. Results

3.1 Reliability Analyses

Both the screening procedure and high-inference codes were subjected to a reliability analysis. Table 1 presents the agreement scores obtained between the coders. In the literature screening process, coders agreed in approximately 80% of cases. Inter-coder reliability was moderate to substantial according to the classification proposed by Landis and Koch [8]. Since agreement and reliability could only be assessed on a

[1] The exact definitions for the levels of the five scales, eHDTS_I (intervention), eHDTS_Sp (planning), eHDTS_Se (execution), eHDTS_Sm (monitoring), and eHDTS_Ss (social) as well as the specific search queries, detailed descriptions of all coded variables, a diagram of the database structure, diagrams of the methodology of the reliability and validity analyses, and querying functionality for database content can be found on the database website http://insyprojects.ewi.tudelft.nl:8888/.

[2] Data and analyses can be accessed at the 4TU.Center for Research Data national research data archive under the following doi: 0.4121/uuid:7e7e91ab-7afc-4b48-8915-e2bc80b23c99 or for quick access https://tinyurl.com/y7k25uqp.

sample, particularly *false positives* (C1 excluded while C2 included, indicating that other relevant articles of the population may have been missed entirely by C1 and hence may not be in the database) had to be examined closely. Four reasons could be identified: mistakes by C1 (full-paper: 1 record), mistakes by C2 (title: 2 records), C2 misunderstanding an exclusion criterion (title: 4 records, abstract: 2 records, full-paper: 3 records), and precaution on the part of C2 as he was instructed to include records when in doubt (title: 5 records, abstract: 3 records). Cautiously included records were unanimously excluded by re-evaluation at the next filtering stage. False negatives were not analyzed further as they did not pose a threat to database content.

Table 1. Results of the reliability analyses conducted for the literature screening process and selected variables coded for in the database. For assigning a degree of technological sophistication to a function instantiation, a weighted Cohen's kappa was calculated since the scale is ordinal.

	Coders	**Population**	**Sample**	**Percent Agreement, 95% CIs**	**Cohen's kappa, 95% CIs**
Literature Screening					
Title Filter	C1, C2	4266	100	.81 [.71, .87]	.50 [.31, .69]
Abstract Filter	C1, C2	1071	44	.80 [.64, .89]	.58 [.34, .82]
Full Article Filter	C1, C2	541	25	.84 [.60, .92]	.69 [.42, .95]
Coding					
Identify functions of system	C1, C3	273	10	.37 [.30, .43]	-
Classify function of instantiation	C1, C3	2224	125	.58 [.49, .66]	.57 [.48, .65]
Map intervention functions to therapeutic intervention class	C1, C3	141	141	.81 [.73, .87]	-
Assign degree of technological sophistication to instantiation	C1, C2	2224	132	.48 [.39, .55]	.60 [.51, .69]

3.2 Scale validation

To determine concurrent validity of the different eHDTS scales, we correlated the informed scores of the coders with their own naïve scores (intra-coder correlation) as well as with the informed scores of C1 (inter-coder correlation). Table 2 shows the Spearman correlations for each scale.

Table 2. Results of the scale validation conducted to assess concurrent validity of the degree of technological sophistication scales. Confidence intervals were obtained by bootstrapping.

	Coders	**Population**	**Sample**	**Intra-Coder Correlation, 95% CIs**	**Inter-Coder Correlation, 95% CIs**
Intervention	C1, C4	1344 (60%)	27	.59 [.21, .80]	.81 [.58, .92]
Treatment Planning	C1, C5	29 (1.3%)	20	.82 [.62, .92]	.70 [.21, .92]
Treatment Execution	C1, C6	445 (20%)	27	.52 [.08, .78]	.67 [.31, .86]
Monitoring	C1, C7	140 (6.3%)	29	.27 [-.14, .61]	.52 [.07, .78]
Social Support	C1, C8	266 (12%)	29	.52 [.11, .79]	.65 [.24, .87]
All types mixed	C1, C2	2224 (100%)	117	.47 [.31, .60]	.58 [.43, .70]

3.3 Distribution over system functions

Analyses concerning the content of the database exceed the scope of this work. We therefore only briefly describe the distribution over function instantiations here. As can be seen in the Population column of Table 2, approximately 60% of function instantiations across systems are of the intervention type. Of the support instantiations, 3% support the user in establishing adherence strategies initially, while more than half aim at reeling the user back in. In line with research indicating that adherence to the systems is higher when human support is included, one third of support instantiations strive to provide human contact.

4. Discussion and Conclusion

The EHealth4MDD database is a systematic inventory of e-mental health systems for the treatment and prevention of depression. It contains 267 such systems with a total of 2224 function instantiations. Approximately 60% of these functions are of a psychotherapeutic nature while 40% aim to support the user in adhering to the intervention. In the literature screening process for populating the database, moderate to substantial reliability was obtained. Double coding of high-inference codes yielded satisfactory percent agreement scores in light of the vast number of possible categories. Therefore, the findings show clear consistency between coders and, as common in high-inference coding, some degree of individual subjectivity. Significant correlations of naïve with informed ratings indicate that four of the five different eHDTS scales capture the intuitive understanding of “technological sophistication” as held by those with a computer science or related degree. Furthermore, correlations between pairs of coders using the scales were significant and high on the same four scales, indicating that technological sophistication can be reliably assessed with the scales. However, the database is not without limitations as a single coder coded most data and reliability of this coding could only be assessed on samples. This resulted in large differences between coders in exposure to the data and therefore possibly lower scores of inter-coder agreement and reliability than might otherwise be expected.

For researchers in both computer science and clinical psychology, the database presents a useful tool for the systematic study of e-mental health interventions for depression. It allows for a better understanding of system composition and of how functionalities contribute to clinical outcomes. Since the database is open access and implemented as a standard MySQL database, it can be linked with other databases, for example, databases of clinical trials. Furthermore, the accessibility allows the research community to contribute to the maintenance of the database.

References

[1] World Federation for Mental Health, Depression: A global Crisis, *World Health Orgaization. Federation for Mental Health.*, 2012.

[2] J. Schröder, T. Berger, S. Westermann, J. P. Klein, and S. Moritz, Internet interventions for depression: new developments, *Dialogues Clin. Neurosci.*, **18** (2016),. 203.

[3] D. Richards and T. Richardson, Computer-based psychological treatments for depression: A systematic review and meta-analysis, *Clinical Psychology Review.*, **32** (2012) 329–342.

[4] P. S. Foroushani, J. Schneider, and N. Assareh, Meta-review of the effectiveness of computerised CBT in treating depression, *BMC Psychiatry*, **11** (2011), 11-131

[5] D. Zhao, M. L. A. Lustria, and J. Hendrickse, Systematic review of the information and communication technology features of web-and mobile-based psychoeducational interventions for depression, *Patient Education and Counselling.*, 2017.

[6] K. L. Gwet, *Handbook of Inter-Rater Reliability: the definitive guide to measuring the extent of agreement among raters*. 2010.

[7] H. Oinas-Kukkonen and M. Harjumaa, Persuasive systems design: Key issues, process model, and system features," *Communications of the Association for Information Systems*, **24** (2009), 28

[8] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, 1977.