

Clinically Meaningful Use of Mental Health Apps

PAC Lab Group Meeting - 01/14/2020

Citation:
Zhang R., Nicholas J., Knapp A. A., Graham A. K., Gray E., Kwasny M. J., Reddy M., Mohr D. C. (2019). Clinically Meaningful Use of Mental Health Apps and its Effects on Depression: Mixed Methods Study. J Med Internet Res, 21(12), e15644.

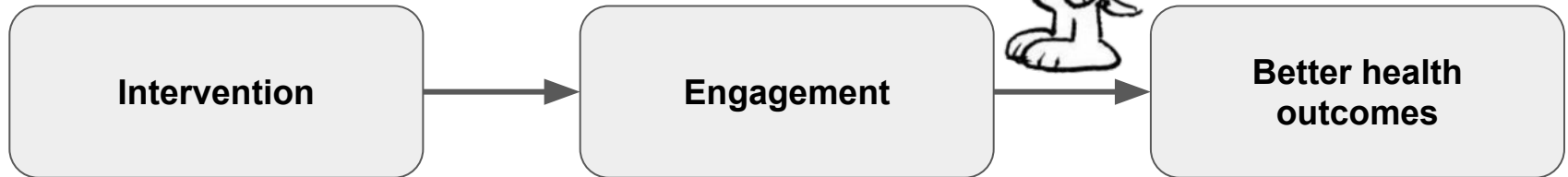
Table of Contents

Slide	Topic
3-5	Introduce study and discussion questions
6-11	Review methodology
12-14	Findings
15	Revisit discussion questions

Key research questions prompting paper

- Can we create a categorization of the types of user behaviors in a suite of mental health apps for depression and anxiety?
- How are different types of app usage related to improvements in symptoms of depression and anxiety?
- Can we differentiate more nuanced app use from generic app use (i.e. intensity and duration of app use), and examine how these different use metrics influence outcomes (lower depression and anxiety)?

Relevance to PAC Lab



Discussion questions for PAC Lab

1. Reproducibility in metrics - how do we take these metrics, or equivalently developed “clinically meaningful metrics”, and make them reproducible (can be used for different applications)?
2. If this were closed-loop (give human feedback), would improved quality of life metrics (e.g. decreased sleep) be better for correlating meaningful usage, rather than PHQ-9?
3. How do we argue that “clinically meaningful” metrics are better than generic app usage, if both are significantly correlated with PHQ-9 reduction? Is having a more nuanced understanding of engagement the core question, and does this question matter?
4. What other factors (age, normal phone behaviors, comorbidities) need to be controlled when defining meaningful smartphone-app engagement metrics? Does not controlling for these factors reduce meaning in the study? I think yes.
5. Why do we think anxiety was not significantly correlated? Paper speculation - motivation is important, anxiety is a more activating condition. Should we strive for MH app-usage metrics that work across multiple conditions?
6. Philosophical question: self-tracking was the most influential at all levels of intensity (logging behaviors + reviewing)...is this more of a reflection of the person’s motivation, or the intervention? How do we design interventions and control for motivation in measuring intervention success?

Methods - IntelliCare Platform

IntelliCare platform consists of 12 clinical apps, each targeting a specific behavioral or psychological treatment strategy to improve symptoms of depression and anxiety.

Example App	Behavioral strategy	Description
Social Force	Social support	Prompts the user to identify supportive people in their lives, and provides encouragement for the user to get back in touch with those positive people.
iCope	Proactive Coping	Allows the user to send oneself inspirational messages and reassuring statements, written in their own words, to help the user get through tough spots or challenging situations.

Original IntelliCare paper:

Mohr, D. C., Tomasino, K. N., Lattie, E. G., Palac, H. L., Kwasny, M. J., Weingardt, K., ... Schueller, S. M. (2017). IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. *Journal of medical Internet research*, 19(1), e10. doi:10.2196/jmir.6645

Methods - Inclusion Criteria

Inclusion criteria as follows:

1. Aged 18 years or older,
2. Reported elevated symptoms of depression (Patient Health Questionnaire-9 [PHQ-9] ≥ 10) or anxiety (generalized anxiety disorder-7 [GAD-7] ≥ 8),
3. Resided in the United States
4. Could speak and read English
5. Had an Android phone with data and text plans

301 eligible participants recruited for 8 week trial.

Methods - Treatments

Randomized to 1 of 4 treatment groups (2x2 factorial design), based upon two types of treatment:

Coaches - Participants assigned to the coaching condition received 8 weeks of coaching aimed to support engagement.

Recommendations - Participants assigned to the recommendation condition received recommendations for new apps weekly.

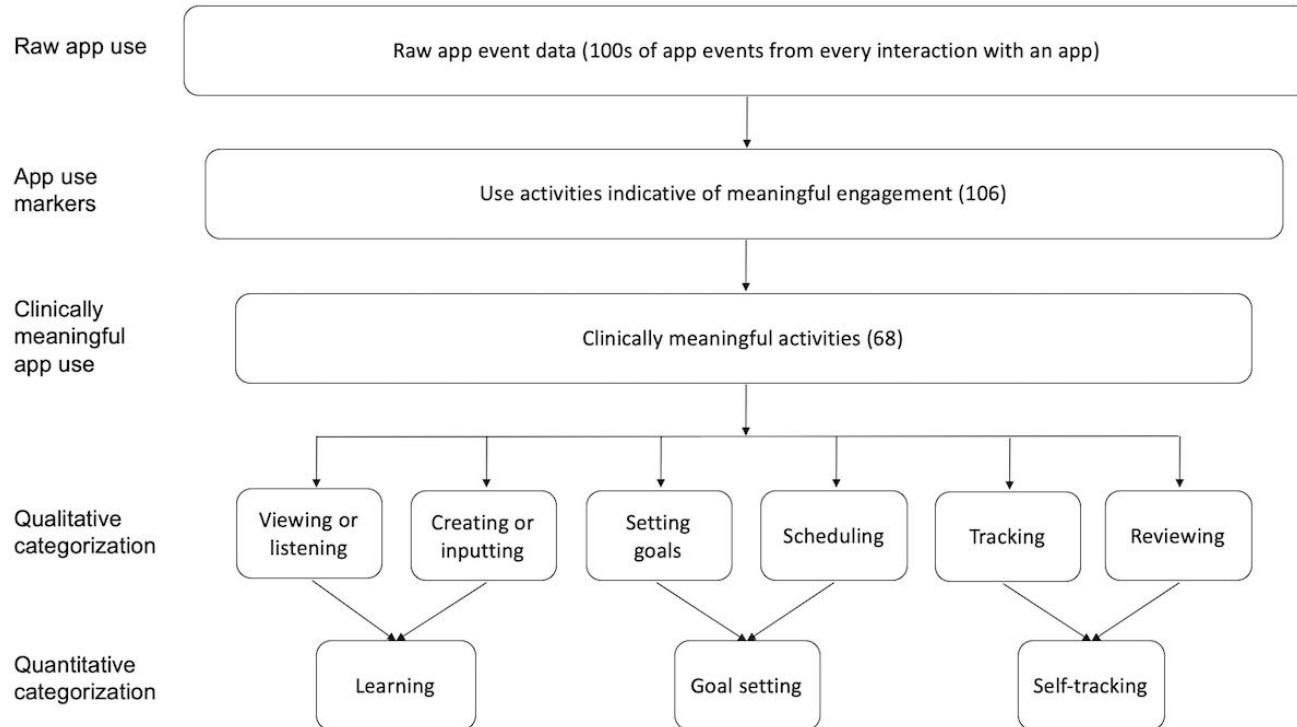
Methods - Treatments

Randomized to 1 of 4 treatment groups (2x2 factorial design), based upon two types of treatment:

Coaches - Participants assigned to the coaching condition received 8 weeks of coaching aimed to support engagement.

Recommendations - Participants assigned to the recommendation condition received recommendations for new apps weekly.

Methods - Defining clinically meaningful usage



Methods - Categorization

Quantitative	Qualitative	Description
Learning	Viewing/listening	Reading/watching/listening to content from the app (e.g. playing an exercise video, viewing a coping card, and listening to relaxing audio)
	Creating/inputting	Creating and editing content for the purpose of learning and cultivating a skill (e.g. identifying a coping activity and creating a positive or self-affirming statement)
Goal setting	Setting goals	Selecting, editing, or adding self-identified or assigned goals (e.g. adding or deleting a checklist item and selecting a weekly goal)
	Scheduling	Scheduling activities or changing reminders to fit one's schedule (e.g. scheduling an upcoming exercise and changing the reminder time)
Self-tracking	Tracking	Keeping track of one's own performance or status through checking off, rating, or logging personal activities and moods, including facts and reasons (e.g. checking a completed activity, rating a level of stress, and creating a sleep log)
	Reviewing	Reviewing one's own content and progress (e.g. reviewing past activity and lessons)

Paper findings - clinically meaningful usage

Covariate	Model 1 ^a		Model 2 ^b		Model 3 ^c	
	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value
Intercept	1.86 (0.93)	.047	1.57 (0.93)	.09	2.16 (0.91)	.02
Coached	-0.78 (0.52)	.13	-0.42 (0.56)	.46	-0.12 (0.56)	.83
Full Hub	-0.21 (0.50)	.67	-0.28 (0.51)	.59	-0.05 (0.51)	.93
PHQ9_baseline	0.54 (0.05)	<.001	0.56 (0.06)	<.001	0.55 (0.06)	<.001
Learning_minimal intensity^d						
Learning_low intensity	0.64 (0.73)	.39	— ^e	—	—	—
Learning_moderate intensity	-2.17 (0.71)	.002	—	—	—	—
Learning_high intensity	-1.22 (0.73)	.09	—	—	—	—
Goal setting_minimal intensity^d						
Goal setting_low intensity	—	—	-0.62 (0.76)	.41	—	—
Goal setting_moderate intensity	—	—	-2.08 (0.76)	.007	—	—
Goal setting_high intensity	—	—	-0.76 (0.76)	.32	—	—
Self-tracking_minimal intensity^d						
Self-tracking_low intensity	—	—	—	—	-2.46 (0.78)	.002
Self-tracking_moderate intensity	—	—	—	—	-1.94 (0.76)	.01
Self-tracking_high intensity	—	—	—	—	-1.92 (0.73)	.009

^a $R^2=0.307$; Adjusted $R^2=0.292$.

^b $R^2=0.281$; Adjusted $R^2=0.265$.

^c $R^2=0.289$; Adjusted $R^2=0.274$.

^dValues of reference group.

^eNot applicable.

Table 1. Regression models of 3 clusters of clinically meaningful activities predicting depression outcome.

Paper findings - meaningful vs. generic usage

Covariate	Model 1 ^a		Model 2 ^b		Model 3 ^b	
	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value
Intercept	2.25 (0.92)	.02	2.24 (0.92)	.02	1.71 (0.93)	.07
Coached	-0.34 (0.53)	.52	-0.58 (0.51)	.26	-0.57 (0.54)	.23
Full Hub	-0.11 (0.51)	.82	0.26 (0.54)	.63	0.02 (0.53)	.98
PHQ9_baseline	0.55 (0.06)	<.001	0.55 (0.05)	<.001	0.54 (0.06)	<.001
Meaningful use_minimal intensity						
Meaningful use_low intensity	-2.00 (0.74)	.007	— ^c	—	—	—
Meaningful use_moderate intensity	-2.07 (0.74)	.006	—	—	—	—
Meaningful use_high intensity	-2.05 (0.74)	.006	—	—	—	—
Generic app use_minimal intensity^d						
Generic app use_low intensity	—	—	-1.44 (0.72)	.047	—	—
Generic app use_moderate intensity	—	—	-2.38 (0.73)	.001	—	—
Generic app use_high intensity	—	—	-2.45 (0.76)	.001	—	—
Generic app use_minimal duration^d						
Generic app use_low duration	—	—	—	—	-0.32 (0.75)	.68
Generic app use_moderate duration	—	—	—	—	-1.52 (0.76)	.045
Generic app use_high duration	—	—	—	—	-1.24 (0.78)	.12

^a $R^2=0.288$; Adjusted $R^2=0.273$.

^b $R^2=0.295$; Adjusted $R^2=0.279$.

^c $R^2=0.274$; Adjusted $R^2=0.258$.

^dValues of reference group.

^eNot applicable.

Table 2. Regression models of total meaningful app use, generic app use, and duration of app use predicting depression outcome.

Paper findings - still anxious (no associations with GAD-7)



Revisit discussion questions

1. Reproducibility in metrics - how do we take these metrics, or equivalently developed “clinically meaningful metrics”, and make them reproducible (can be used for different applications)?
2. If this were closed-loop (give human feedback), would improved quality of life metrics (e.g. decreased sleep) be better for correlating meaningful usage, rather than PHQ-9?
3. How do we argue that “clinically meaningful” metrics are better than generic app usage, if both are significantly correlated with PHQ-9 reduction? Is having a more nuanced understanding of engagement the core question, and does this question matter?
4. What other factors (age, normal phone behaviors, comorbidities) need to be controlled when defining meaningful smartphone-app engagement metrics? Does not controlling for these factors reduce meaning in the study? I think yes.
5. Why do we think anxiety was not significantly correlated? Paper speculation - motivation is important, anxiety is a more activating condition. Should we strive for MH app-usage metrics that work across multiple conditions?
6. Philosophical question: self-tracking was the most influential at all levels of intensity (logging behaviors + reviewing)...is this more of a reflection of the person’s motivation, or the intervention? How do we design interventions and control for motivation in measuring intervention success?