

Usable Data Visualization for Digital Biomarkers: An Analysis of Usability, Data Sharing, and Clinician Contact

Luke Scheuer John Torous

Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA

Keywords

Mhealth · Digital health · Apps · Health · Visualization

Abstract

Background: While digital phenotyping smartphone apps can collect vast amounts of information on participants, less is known about how these data can be shared back. Data visualization is critical to ensuring applications of digital signals and biomarkers are more informed, ethical, and impactful. But little is known about how sharing of these data, especially at different levels from raw data through proposed biomarkers, impacts patients' perceptions. **Methods:** We compared five different graphs generated from data created by the open source mindLAMP app that reflected different ways to share data, from raw data through digital biomarkers and correlation matrices. All graphs were shown to 28 participants, and the graphs' usability was measured via the System Usability Scale (SUS). Additionally, participants were asked about their comfort sharing different kinds of data, administered the Digital Working Alliance Inventory (D-WAI), and asked if they would want to use these visualizations with care providers. **Results:** Of the five graphs shown to participants, the graph visualizing change in survey re-

sponses over the course of a week received the highest usability score, with the graph showing multiple metrics changing over a week receiving the lowest usability score. Participants were significantly more likely to be willing to share Global Positioning System data after viewing the graphs, and 25 of 28 participants agreed that they would like to use these graphs to communicate with their clinician. **Discussion/Conclusions:** Data visualizations can help participants and patients understand digital biomarkers and increase trust in how they are created. As digital biomarkers become more complex, simple visualizations may fail to capture their multiple dimensions, and new interactive data visualizations may be necessary to help realize their full value.

© 2022 The Author(s).

Published by S. Karger AG, Basel

Background

The role of digital data and biomarkers for healthcare continues to expand. Smartphones and other connected devices are able to collect rich, real-time, and temporally dense data that have accelerated interest and potential in behavioral biomarkers. From simple longitudinal surveys (<https://www.zotero.org/google-docs/?5jhI5l> [1]) to in-

teractive assessments of reaction time [2, 3], digital data from smartphones offer a new window into health. Many of these digital signals, such as GPS, that can be used to infer circadian routines, or wearables to understand fatigue [4, 5], are ubiquitous and often already collected by smartphones, which are owned by the majority of the population [6] and those with mental illness [7, 8].

While the potential of digital biomarkers is already well known [9–13], challenges to their uptake include growing concerns around data sharing and perceived lack of clinical value. Indeed, several studies confirm that usage rates of these mental health apps drop to less than 5% within 10 days [14, 15], and apps that collect data without visualizing it for users are often found to be unengaging [16]. Many people are reluctant to share their digital signals because of privacy concerns [17] and not understanding what the data are used for. Data visualization offers a solution, in that it can help people learn how their raw data are used, how that raw data can be transformed into privacy-preserving digital biomarkers, and how those digital biomarkers relate to their health. Given the vast amount of temporal data generated by digital devices and the early state of research on these biomarkers, visualization is even more important as it offers a more accessible and interpretable tool than summary statistics.

Yet, current research on data visualization and its impact on trust or engagement remain sparse. Engagement research to date on mental health apps has identified a need to provide users with personalized content available across a range of devices [18]. A 2020 review article by Polhemus et al. [16] on the status of the visualization landscape highlighted the need for graphs that individuals with mental illness can use both on their own and in concert with physicians or other care workers but noted that most research studies focused on one particular app or product instead of more generalizable knowledge [16]. None to date have explored interactive visualizations, which may be particularly important for sharing complex temporal data gathered across numerous sensors (e.g., GPS trends in home time vs. accelerometer-derived sleep and their combined relationship with mood). While numerous papers have examined engagement features of their own app and many call for good design and co-creation – few offer specific and generalizable principles. Thus, this study focuses on three particularly relevant issues as informed by recent literature: how simplicity and interactivity affect a graph's usability, how educating users about data affects their willingness to share it, and what graphs users want to use by themselves and which they want to use with clinicians or other providers.

We used five graphs already piloted in patient-facing studies in our lab and used by clinicians in our clinic to investigate how different levels of analysis, interactivity, and graph design affect the usability of graphs, the alliance between user and graph, and whether visualizations could change how comfortable users were with sharing different forms of data. We explore how users can best understand their digital biomarkers and the difference between un-interactive graphs versus more interactive visualizations utilizing tooltip hovering features.

Methods

Data Collection

Participants for this study were recruited from a larger study investigating engagement with the mindLAMP app [9, 10]. mindLAMP is a smartphone-based app, which both collects many commonly used digital biomarkers, such as GPS, accelerometer, or step count, and can be used to remotely administer surveys and common cognitive games [19], in a research or clinical setting [20]. No features unique to mindLAMP were used for this study to ensure our results remained broadly applicable. Eligible participants for the larger study were 18 years of age or older and reported moderate symptoms of stress as measured by the Perceived Stress Scale (PSS) [21]. Twenty-eight participants partook in a structured interview with Luke Scheuer (L.S.) and completed three different measures during the study visit.

First and before looking at any visualizations, participants were asked to rate how comfortable they were sharing five different forms of digital data collectible by a standard smartphone: keylogging or content data from texts and emails (keylogging), metadata concerning the number of texts or emails sent (metadata), GPS location data (GPS), accelerometer data (accelerometer) or data from surveys or questionnaires administered digitally (surveys), using a 5-point scale of “Strongly Disagree,” “Disagree,” “Neither Agree nor Disagree,” “Agree,” or “Strongly Agree”; this scale was used for all subsequent measures as well. This survey was repeated at the end of the visit with the goal of assessing how visualizations may influence data sharing as shown below in Figures 1 and 2.

Next, static images of five graphs of varying complexity and analysis level were shown, one at a time, to the participants on a computer, along with a brief explanation: a data quality graph (data quality), a graph showing the changes in two survey scores over a week (survey responses), a set of graphs showing home time, step, and screen use data derived from analyzing passive data (analyzed passive), a summary graph showing weekly change in multiple metrics (summary), and a correlation graph comparing multiple metrics (correlation) (Table 1). For each graph, the participants took the System Usability Scale (SUS), a widely used, easily administered, and normatively scored measure of usability in digital products [22, 23].

Finally, participants were asked to respond to Digital Working Alliance Inventory (D-WAI) [24] questions, to describe how they would feel about using these graphs in a clinical setting. Participants were also asked if they felt the graphs could provide new insight into their problems or help them communicate better with their clinicians, as well as if they had any qualitative comments about the graphs shown to them.

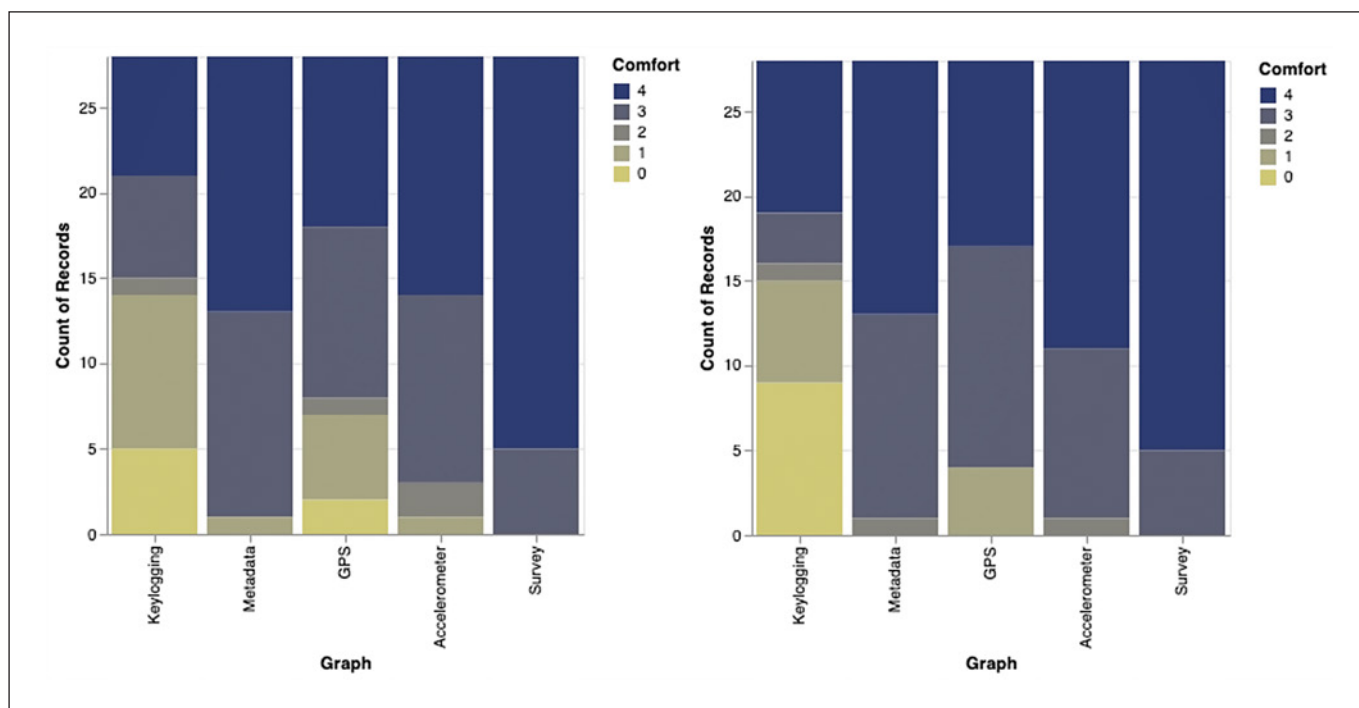


Fig. 1. Stacked bar chart of comfort levels. This chart shows the distribution of various participants' comfort levels sharing each type of data. Lower scores reflect less comfort, while higher scores reflect more comfort.

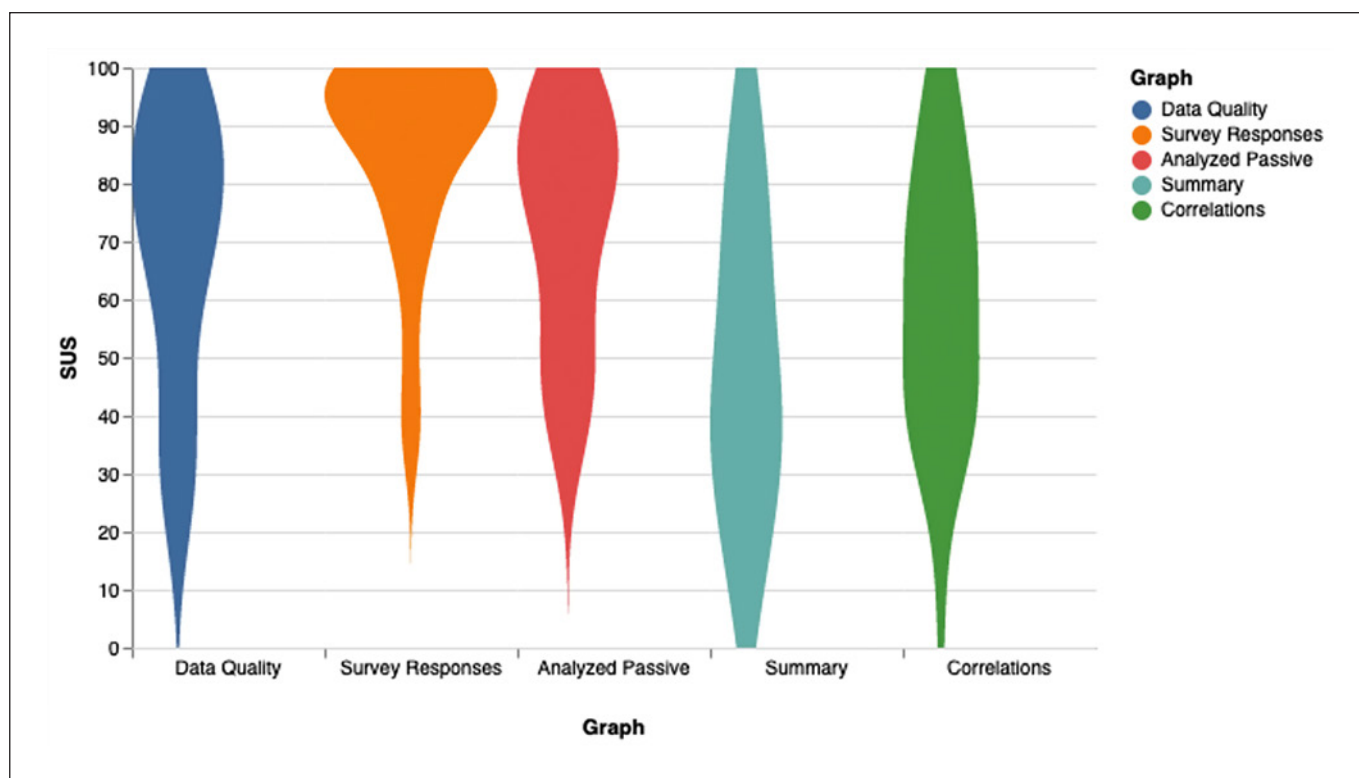


Fig. 2. Distribution of SUS scores by the graph.

Table 1. Graphs shown to each participant

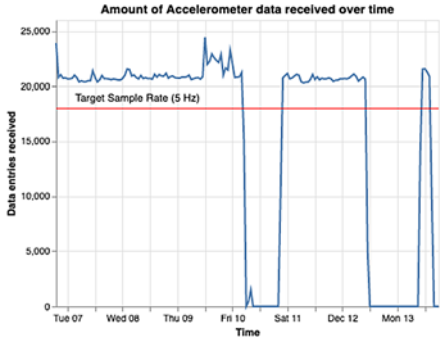
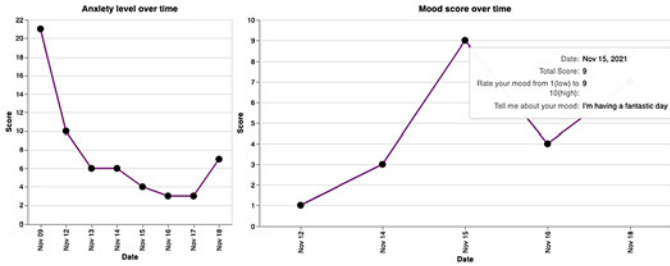
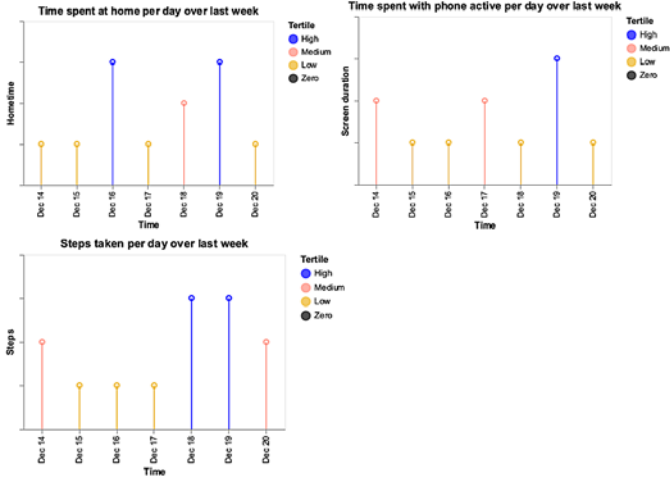
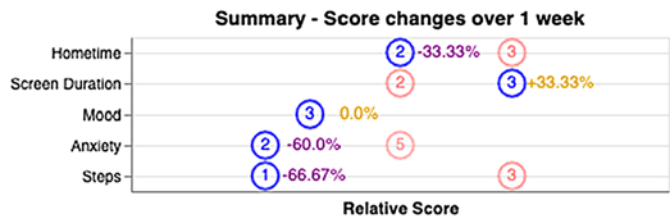
Graph(s)	Description
	<p>A – data quality: this first graph is a measurement of the amount of data (here, accelerometer) collected by your phone each hour over the course of a week. You could use this data to make sure that you were collecting enough data to get useful information from any analyses</p>
	<p>B – survey responses*: the second graph is an example of survey data collected over the course of around a week. You could see how your score changes from day to day. By hovering over a specific data point, you can see your specific responses to each question, as shown on the right</p>
	<p>C – analyzed sensor data: these graphs are examples of data generated by analyzing data from sources such as GPS or an accelerometer. Here, you could see how your steps, time spent at home, and time spent using a phone have changed over the course of a week. Of note, all these scores are relative to themselves, not any standard measure – as such, when a measurement is “high” that means it is high for you</p>
	<p>D – summary: this graph is a summary chart intended to show how multiple types of data change over the course of a week. Here, red points are “old,” and blue points are “new,” with the percentage change over a week, shown next to the newer point. For example, here home time has decreased by about 1/3rd over a week. Mood is unchanged here</p>

Table 1 (continued)

Graph(s)	Description
	<p>E – correlation*: this graph is a correlation chart intended to show how some kinds of data change together. Higher correlations, meaning scores increase together, are shown in blue, with lower correlations, meaning scores change in opposite directions, shown in red. For example, home time and anxiety are negatively correlated – meaning as you spend more time at home, you are less anxious. However, home time and screen time are positively correlated – meaning more time spent at home is associated with more screen time. Like the second graph, hovering over a specific box would bring up a “tooltip” with additional info; here, the exact value of the correlation, the two variables being compared, and a brief interpretation</p>

The left column of this table shows images of the five graphs shown to participants in the order they saw them. The right column shows the description read to each participant as they saw the graph for the first time. Asterisks mark graphs with an example of a tooltip.

Table 2. Comfort means and standard deviations

	Mean (SD), pre	Mean (SD), post
Keylogging (0–4)	2.04 (1.52)	1.89 (1.72)
Metadata (0–4)	3.46 (0.69)	3.5 (0.58)
GPS (0–4)	2.75 (1.32)	3.1 (0.99)
Accelerometer (0–4)	3.36 (0.78)	3.57 (0.57)
Survey (0–4)	3.82 (0.39)	3.82 (0.39)
Combined (0–20)	15.42 (3.46)	15.89 (3.31)

Results

Comfort

Comfort levels for each form of digital data were scored on a 0–4 scale: a score of 0 represents the least comfort sharing data; a score of 4, the most. Thus, total comfort sharing data were scored on a 0–20 scale. A one-way between subjects ANOVA analysis on comfort sharing data was found to be significant for both pre- ($F [4, 23] = 12.98$, p value < 0.001) and post-viewing ($F [4, 23] = 17.98$, p value < 0.001) conditions (Table 2).

Additionally, pre- and post-comfort were compared by way of a two-sided paired t test for all forms of data asked about. The only statistically significant change was

an increase in the comfort of GPS ($p = 0.039$). Also of note is a statistically insignificant decrease in comfort with sharing keylogging data ($p = 0.211$) and close to statistically significant increase in comfort sharing accelerometer data ($p = 0.084$).

SUS Individual

SUS results for each graph were scored on a 0–100 scale, with lower scores indicating lower usability and higher scores reflecting higher usability. Average scores for each graph were mapped onto the scoring system suggested by Bangor et al. [23], yielding the ratings shown in Table 3 below.

A one-way between subjects ANOVA analysis showed significant differences between graph types ($F [4, 23] = 10.92$, $p < 0.001$). Post hoc comparisons between different graphs showed that the survey response graph was rated significantly more usable than all four other graphs (vs. data quality: $p < 0.001$; vs. analyzed passive: $p < 0.001$; vs. summary: $p < 0.001$; vs. correlations: $p < 0.001$). Analyzed passive graphs were more usable than the summary ($p < 0.001$) or correlation ($p = 0.001$) graphs. The data quality graph was also more usable than the summary ($p < 0.001$) or correlation ($p = 0.017$) graphs, and the correlation graph was only significantly more usable than the summary graph ($p = 0.048$).

Table 3. Mean and standard deviation in the SUS score, converted to adjective rating 21

	Mean score	Standard deviation	Adjective rating
Data quality	69.29	23.44	OK (51–72)
Survey responses	85.80	17.63	Excellent (85+)
Analyzed passive	72.86	20.34	Good (72–85)
Summary	49.02	25.90	Poor (39–51)
Correlations	59.38	23.12	OK (51–72)

Table 4. D-WAI and added questions' mean scores by item. Asterisks indicate the additional items not present in the traditional D-WAI

Statement	Mean (SD)
I trust these graphs to guide me towards my personal goals	3.14 (0.74)
I believe the graphs would help me to address my problem	3.25 (0.83)
The graphs encourage me to accomplish tasks and make progress	3.11 (0.86)
I agree that using the graphs are important for my goals	3.11 (0.86)
The graphs are easy to use and understand	2.82 (0.92)
The graphs support me to overcome challenges	2.57 (1.05)
These graphs give me a new way to look at my problems*	3.50 (0.68)
These graphs would help me communicate better with my clinician*	3.38 (0.86)

Digital Working Alliance Inventory

Digital Working Alliance scores, as well as the two added questions, were converted from a scale of “Strongly Disagree,” “Disagree,” “Neither Agree nor Disagree,” “Agree,” or “Strongly Agree” to a 0–4 point scale (Table 4).

Discussion

Data visualization remains a promising if largely unexplored means to help users better understand, engage with, and benefit from digital health data. Our findings indicate that effective data visualizations can change people's willingness to share data, inform how data are shared today, and suggest new ways of communicating with clinicians.

We found data visualization to be a potentially useful method to help people both better understand what data they are sharing and increase their comfort doing so. After showing participants graphs which incorporated a measure derived from GPS, we observed a statistically significant increase in willingness to share GPS data. Sharing information in this manner can also help patients understand what they do not wish to share. For example, we found participants were not more willing to share their keylogging data (the content of their text and email messages) after viewing the graphs – in fact, average comfort decreased, although not significantly.

In general, users found simple graphs – ones that showed raw data such as the number of steps taken per day or survey scores – more usable (Tables 1, 2). More complicated graphs that integrated and analyzed several data metrics, like the summary, which showed weekly changes in all variables, or correlation charts, which showed how different metrics changed together, were rated less usable. In other words, performing more analyses on data to tie together several streams of data did not increase usability for patients. Our results suggest also that interactivity features can make a noticeable difference in usability. Two of the graphs shown, survey responses and correlations, contained an example of tooltips, whereby a user could hover over specific sections of a graph and receive more information; for instance, a record of their responses to particular questions or a brief interpretation of what a correlation value means (Table 1). Compared to graphs of similar complexity, the graphs with tooltips received higher usability scores (Table 2). This suggests that interactivity through a system like tooltips can add depth to simple data more detailed and complex data more understandable.

Participants were able to identify graphs they valued both for use on their own, as well as graphs that they wanted to use in concert with a clinician. After adjusting for differences in scoring, the graph system used for this study scored similarly on the D-WAI to the average D-WAI ratings found by another study for people's most commonly used meditation apps, 33.87 versus 30.58 [25]. While simpler graphs were rated as more independently

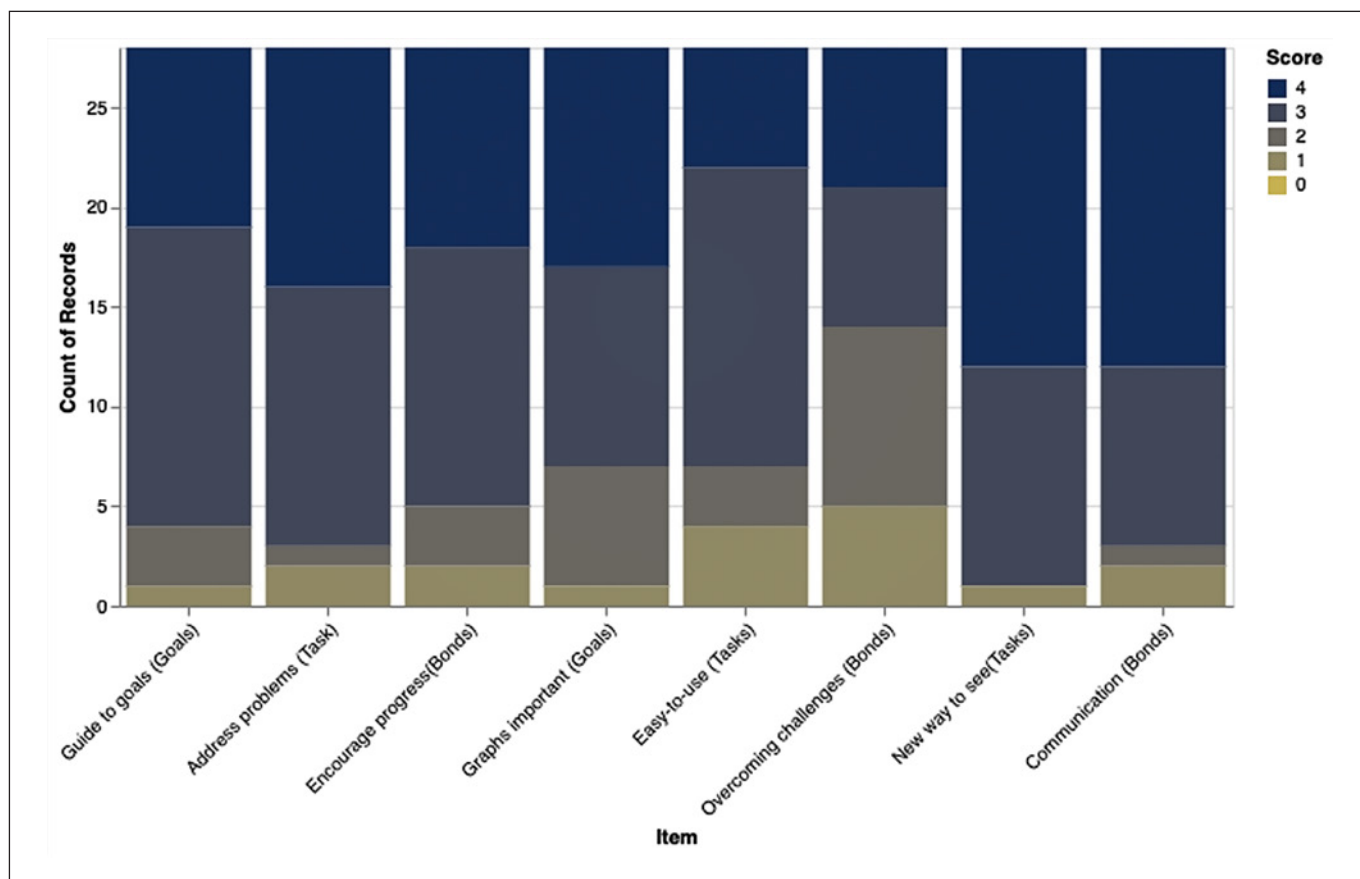


Fig. 3. The first six items from left to right represent the traditional D-WAI items, with “New way to see” and “Communication” representing “These graphs give me a new way to look at my problems” and “These graphs would help me communicate better with my clinician,” respectively.

usable than complex ones, most participants saw the value of all the graphs – of the 28 participants surveyed, 27 agreed that using the graphs would give them “a new way to look at their problems,” and 25 agreed that the graphs would help them “communicate better with (their) clinician” (Fig. 3). This indicates that users still saw the value in the complex summary or correlation graphs they rated as less usable and understood the intermediary role that physicians, other clinicians, and digital navigators could play to help patients get the greatest value out of their data in a clinical setting.

Though we did ask participants if they had qualitative feedback about any of the visualizations shown, the small sample size made it difficult to draw out any new conclusions, and most, if not all, feedback is already detailed above: the intuitive nature of simpler visualizations like the survey responses graph, the overly complex nature of the summary graph in particular, and how tooltips increased usability.

Our study has some limitations. First, it had a relatively small sample size reflecting the nature of a pilot study. The fact that participants in this study were all sampled from a larger study where an entry criterion was scoring moderately or above on PSS is an advantage – as the studied issues are particularly relevant to a clinical population; however, the fact that the larger study examined the uses of digital technology may mean that participants were more willing to share data and use technology than may be true for the average patient. Additionally, since the images for this study were static, future studies should let participants interact with the graphs as they would in a clinical setting – either on their own phone or computer and be able to directly interact with the images. Finally, though the topic of data visualization and return of patient data is of broad interest across many, if not all, health settings, and this study focused primarily on a mental health context, meaning additional work would be required to extend our conclusions to other settings; and

though the SUS is used widely for assessing mental health app usability, using additional usability scales and qualitative questions in the future could ensure we capture a more holistic portrait of visualization usability.

This work also suggests next steps. First, we identified the important role clinicians could play in helping patients to understand data when it must be presented in a more complicated way, like a summary or correlation graph. The next step, then, is to meet with clinicians and learn what they value in data visualizations and if they see value in having graphs available to them and their patients. Second, usability scores for graphs with tooltips were around 10 points higher compared to graphs of similar complexity that did not utilize tooltips. This suggests measuring the usability of single graphs with and without tooltips to further tease out the direct effects of interactivity, which might help make more complicated graphs like the summary and correlation charts usable by those without clinical backgrounds. Investigating both of these issues will provide valuable information and help create a system of visualizations that support both the patient and clinician.

Conclusion

As digital biomarkers continue to expand their role in healthcare, visualization is important to increase their trust, uptake, and impact. Our pilot study explored data visualization for digital phenotyping data and found that simple graphs are valuable today and more interactive visualizations hold currently unexplored potential for using these biomarkers in clinical settings.

References

- 1 Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29, 190 data points from 278 apps. *Acta Psychiatr Scand*. 2021 Aug;144(2):201–10.
- 2 Gansner M, Nisenson M, Carson N, Torous J. A pilot study using ecological momentary assessment via smartphone application to identify adolescent problematic internet use. *Psychiatry Res*. 2020 Nov;293:113428.
- 3 Henson P, Torous J. Feasibility and correlations of smartphone meta-data toward dynamic understanding of depression and suicide risk in schizophrenia. *Int J Methods Psychiatr Res*. 2020 Jun;29(2):e1825. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/mpr.1825>.
- 4 Wisniewski H, Henson P, Torous J. Using a smartphone app to identify clinically relevant behavior trends via symptom report, cognition scores, and exercise levels: a case series. *Front Psychiatry*. 2019 Sep;2310:652.
- 5 Luo H, Lee PA, Clay I, Jaggi M, De Luca V. Assessment of fatigue using wearable sensors: a pilot study. *Digit Biomark*. 2020;4(Suppl 1):59–72.
- 6 Pew Research Center. *Demographics of mobile device ownership and adoption in the United States (Internet)*. Pew Research Center; 2021. [cited 2022 Feb 15]. Available from: <https://www.pewresearch.org/internet/fact-sheet/mobile/>.
- 7 Iliescu R, Kumaravel A, Smurawska L, Torous J, Keshavan M. Smartphone ownership and use of mental health applications by psychiatric inpatients. *Psychiatry Res*. 2021 May 1;299:113806.
- 8 Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. *JMIR MHealth UHealth*. 2014 Jan 21;2(1):e2.
- 9 Vaidyam A, Halamka J, Torous J. Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. *JMIR MHealth UHealth*. 2022 Jan 7;10(1):e30557.
- 10 Bilden R, Torous J. Global collaboration around digital mental health: the LAMP consortium. *J Technol Behav Sci*. 2022 Jan;7(2):227–33.

Statement of Ethics

This study protocol was reviewed and approved by BIDMC IRB approval number #2021P000949 with a waiver on informed consent. This study was granted a waiver of written informed consent.

Conflict of Interest Statement

J.T. has cofounded a mental health technology company called Precision Mental Wellness, unrelated to this work.

Funding Sources

There are no funding sources for this

Author Contributions

Conception and design: J.T. and L.S. Administrative support: J.T. Provision of study material or patients: N/A. Collection and assembly of data: L.S. Data analysis and interpretation: J.T. and L.S. Manuscript writing: J.T. and L.S. Final approval of manuscript: J.T. and L.S.

Data Availability Statement

Survey results can be shared upon reasonable request.

- 11 Gansner M, Nisenson M, Lin V, Carson N, Torous J. Piloting smartphone digital phenotyping to understand problematic internet use in an adolescent and young adult sample. *Child Psychiatry Hum Dev*. 2022 Jan. Online ahead of print.
- 12 Henson P, Pearson JF, Keshavan M, Torous J. Impact of dynamic greenspace exposure on symptomatology in individuals with schizophrenia. *PLoS One*. 2020;15(9):e0238498.
- 13 Melcher J, Lavoie J, Hays R, D'Mello R, Rauseo-Ricupero N, Camacho E, et al. Digital phenotyping of student mental health during COVID-19: an observational study of 100 college students. *J Am Coll Health*. 2021 Mar 26; 1–13.
- 14 Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J Med Internet Res*. 2019 Sep 25; 21(9):e14567.
- 15 Jaworski BK, Taylor K, Ramsey KM, Heinz A, Steinmetz S, Pagano I, et al. Exploring usage of COVID coach, a public mental health app designed for the COVID-19 pandemic: evaluation of analytics data. *J Med Internet Res*. 2021 Mar 1;23(3):e26559.
- 16 Polhemus A, Novák J, Majid S, Simblett S, Bruce S, Burke P, et al. Data visualization in chronic neurological and mental health condition self-management: a systematic review of user perspectives. *JMIR MHealth UHealth*. 2022;28(9):e25249.
- 17 Parker L, Halter V, Karliychuk T, Grundy Q. How private is your mental health app data? An empirical study of mental health app privacy policies and practices. *Int J Law Psychiatry*. 2019 May–Jun;64:198–204.
- 18 Balaskas A, Schueller SM, Cox AL, Doherty G. The functionality of mobile apps for anxiety: systematic search and analysis of engagement and tailoring features. *JMIR MHealth UHealth*. 2021 Oct 6;9(10):e26712.
- 19 Torous J, Vaidyam A. Multiple uses of app instead of using multiple apps: a case for rethinking the digital health technology toolbox. *Epidemiol Psychiatr Sci*. 2020;29:e100. Available from: <https://www.cambridge.org/core/journals/epidemiology-and-psychiatric-sciences/article/multiple-uses-of-app-instead-of-using-multiple-apps-a-case-for-rethinking-the-digital-health-technology-toolbox/CE90D6BCD02AB3BC1324DB9082225325>.
- 20 Rauseo-Ricupero N, Henson P, Agate-Mays M, Torous J. Case studies from the digital clinic: integrating digital phenotyping and clinical practice into today's world. *Int Rev Psychiatry*. 2021 May 19;33(4):394–403.
- 21 Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983 Dec;24(4):385.
- 22 Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Human Computer Interact*. 2008 Jul 29; 24(6):574–94.
- 23 Bangor A, Kortum PT, Miller JT. Determining what individual SUS scores mean: adding an adjective rating scale. *J User Experience*. 2009;114–23. Available from: <https://uxpa-journal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/>.
- 24 Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. *BJ-Psych Open*. 2019 Jan 29;5(1):e15.
- 25 Goldberg SB, Baldwin SA, Riordan KM, Torous J, Dahl CJ, Davidson RJ, et al. Alliance with an unguided smartphone app: validation of the digital working alliance inventory. *Assessment*. 2021 May 18;107319112110153. Online ahead of print.