Designing a Visual Analytics Tool to Support Analysis Tasks of Digital Mental Health Interventions: A Proof-of-Concept Study

Gyuwon Jung, Heejeong Lim, Kyungsik Han, Hyungsook Kim, Uichin Lee

Submitted to: JMIR Human Factors
on: July 31, 2024

Table of Contents

Designing a Visual Analytics Tool to Support Analysis Tasks of Digital Mental Health Interventions: A Proof-of-Concept Study

Gyuwon Jung[1] MS; Heejeong Lim[1] BS; Kyungsik Han[2, 3] PhD; Hyungsook Kim[4, 5] PhD; Uichin Lee[1] PhD

[1]School of Computing KAIST Daejeon KR
[2]Department of Data Science Hanyang University Seoul KR
[3]Department of Artificial Intelligence Hanyang University Seoul KR
[4]Hanyang Digital Healthcare Center Hanyang University Seoul KR
[5]Department of Cognitive Sciences, School of Intelligence Hanyang University Seoul KR

Corresponding Author:
Uichin Lee PhD
School of Computing
KAIST
291 Daehak-ro, Yuseong-gu
Daejeon
KR

Abstract

Background: Digital Health Interventions (DHIs) are widely used to manage users' health in everyday life through digital devices. The use of DHIs generates various data, such as records of intervention usage and the status of target symptoms, providing researchers with data-driven insights for improving these interventions even after deployment. Although DHI researchers have investigated this data, existing analysis practices have been carried out in a fragmented manner, limiting the comprehensive understanding of the data.

Objective: We proposed an analysis task model to help DHI researchers analyze observational data from a holistic perspective. This model was then used to prototype an interactive visual analytics tool. Our objective is to evaluate the model's suitability for DHI data analysis and explore task support through a visual analytics tool.

Methods: We constructed an analysis task model based on data analysis practices from existing DHI research. Moreover, we designed 'Maum Health Analytics,' an initial prototype of an interactive visual analytics tool that supports the tasks included in the proposed model. To investigate whether our model adequately covers the DHI data analysis process, we conducted a preliminary user study with five groups of DHI researchers (n=15). During this process, we had them use Maum Health Analytics within given data analysis scenarios, providing analyzed results from in-the-wild data collected in a non-experimental setting through a mobile DHI service targeting depressive symptoms. After using the analytics tool, we interviewed the DHI researchers to determine whether the analysis tasks were appropriate and how the information provided by the tool could be utilized in practice.

Results: Our analysis task model was created using three key components (i.e., user grouping criteria) for DHI data analysis: user characteristics, user engagement with DHIs, and the effectiveness of DHIs on the target symptom via pre-post comparisons. Furthermore, the prototype of interactive visual analytics was designed, with each feature mapped one-to-one to an analysis task described in the model. From the interview sessions, DHI researchers valued group-level analysis that enabled identifying users who need care, improving intervention content and recommendations, and understanding the effectiveness of DHIs in connection with user characteristics and engagement levels. They also noted several benefits of the model and tool, such as simplifying analysis tasks and supporting communication among diverse experts.

Conclusions: We proposed an analysis task model and designed an interactive visual analytics tool to support DHI researchers. Our user study showed that the model allows a holistic investigation of DHI data by integrating key analysis components, and the prototype tool simplifies analytic tasks and enhances communication among researchers. As DHIs grow, our model and tool could effectively meet the data analysis needs of researchers and improve efficiency.

(JMIR Preprints 31/07/2024:64967)
DOI: https://doi.org/10.2196/preprints.64967

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Designing a Visual Analytics Tool to Support Analysis Tasks of Digital Mental Health Interventions: A Proof-of-Concept Study

Abstract

Background: Digital Health Interventions (DHIs) are widely used to manage users' health in everyday life through digital devices. The use of DHIs generates various data, such as records of intervention usage and the status of target symptoms, providing researchers with data-driven insights for improving these interventions even after deployment. Although DHI researchers have investigated this data, existing analysis practices have been carried out in a fragmented manner, limiting the comprehensive understanding of the data.
Objective: We proposed an analysis task model to help DHI researchers analyze observational data from a holistic perspective. This model was then used to prototype an interactive visual analytics tool. Our objective is to evaluate the model's suitability for DHI data analysis and explore task support through a visual analytics tool.
Methods: We constructed an analysis task model based on data analysis practices from existing DHI research. Moreover, we designed 'Maum Health Analytics,' an initial prototype of an interactive visual analytics tool that supports the tasks included in the proposed model. To investigate whether our model adequately covers the DHI data analysis process, we conducted a preliminary user study with five groups of DHI researchers (n=15). During this process, we had them use Maum Health Analytics within given data analysis scenarios, providing analyzed results from in-the-wild data collected in a non-experimental setting through a mobile DHI service targeting depressive symptoms. After using the analytics tool, we interviewed the DHI researchers to determine whether the analysis tasks were appropriate and how the information provided by the tool could be utilized in practice.
Results: Our analysis task model was created using three key components (i.e., user grouping criteria) for DHI data analysis: user characteristics, user engagement with DHIs, and the effectiveness of DHIs on the target symptom via pre-post comparisons. Furthermore, the prototype of interactive visual analytics was designed, with each feature mapped one-to-one to an analysis task described in the model. From the interview sessions, DHI researchers valued group-level analysis that enabled identifying users who need care, improving intervention content and recommendations, and understanding the effectiveness of DHIs in connection with user characteristics and engagement levels. They also noted several benefits of the model and tool, such as simplifying analysis tasks and supporting communication among diverse experts.
Conclusions: We proposed an analysis task model and designed an interactive visual analytics tool to support DHI researchers. Our user study showed that the model allows a holistic investigation of DHI data by integrating key analysis components, and the prototype tool simplifies analytic tasks and enhances communication among researchers. As DHIs grow, our model and tool could effectively meet the data analysis needs of researchers and improve efficiency.

Keywords: Digital Health Interventions; Visual Analytics; Data Analysis Tasks; User Characteristics; User Engagement; Effectiveness

Introduction

Background

Digital devices such as smartphones and wearables are extensively used in healthcare to deliver interventions beyond traditional medical settings [1]. These interventions, known as Digital Health Interventions (DHIs), are designed to address various health issues and promote healthy behaviors like physical activity and smoking cessation, or manage chronic conditions such as depression and diabetes in daily life [2, 3].

DHIs utilize various intervention strategies, including behavior change techniques (BCTs) which are proven effective in altering health behaviors [4-6]. Such techniques include monitoring user behavior, setting goals, providing social support, and incorporating gamification. Furthermore, DHIs can be classified into different categories such as digital health, digital medicine, or digital therapeutics, based on clinical evidence and real-world outcomes [7]. Previous studies [8-11] have investigated the delivery procedures of DHIs, identified barriers in their delivery path, and determined the opportune moments for providing DHIs. All of these studies share the common objective of effectively improving user behavior or symptoms targeted by DHIs.

DHIs are commonly designed as mobile applications, such as those in mobile health (mHealth), enabling individuals to install and use them on their smart devices.
When users engage with the features and contents offered by DHIs, a variety of data is collected. For instance, several types of log data are passively gathered during user interactions with DHIs. Most of these log data include detailed records of user activities including the frequency of DHI app usage, the individual DHI content accessed, and user interactions such as taps, swipes, and text entries within particular intervention contents [12]. In addition, some data are actively provided by users through manual inputs, which are essential for capturing information that may be challenging to track solely via passive log data. These data often include basic user information, such as socio-demographic details, and periodic self-reports on their physical or mental health status.

Previous studies have conducted various analyses to better understand the data collected from DHIs and assess either the user engagement with the intervention contents or the effectiveness of the interventions in supporting target users. Researchers studying DHIs basically have explored user engagement with the interventions, assessing either subjective experiences about how immersed users are or objective behaviors such as usage frequency [11]. When quantitatively assessing engagement, researchers can analyze macro indicators such as the number of logins to the DHI, frequency of content access, time spent, and sequence of content use, as well as micro indicators like the number of clicks and swipes [13-16]. A recent study [17] suggested similar metrics for measuring engagement, categorizing them into individual-level metrics such as launch counts, usage durations, and long-term usage patterns, and population-level metrics, including the ratio of users who open the app at least once and the number of completers.

In addition, researchers have evaluated the effectiveness of DHIs on target symptoms by comparing health status before and after the usage of DHIs in natural settings. The evaluation of DHIs typically progresses from measuring efficacy with a small group of users in controlled environments (e.g., randomized controlled trials) to assessing effectiveness with a large group of users in uncontrolled environments [1]. By evaluating effectiveness without predefined treated and control groups, researchers can observe whether DHIs can yield the intended health changes even in the real world, where numerous confounding factors exist. These metrics (i.e., user engagement and DHI effectiveness) can indicate "how actively users have utilized DHIs" and "whether they have experienced improved health conditions," making them reasonable criteria for evaluating DHIs. One relevant work is the SilverCloud platform proposed by Doherty et al. [18], which explores how to encourage users to engage more actively with online mental health interventions. Using this platform, researchers have suggested approaches for visualizing log data to understand the engagement levels of individuals or groups of users [19] and predicted clinical outcomes of the interventions by employing machine learning techniques [20].

Nevertheless, existing analyses of DHI data have often been conducted separately based on individual researchers' interests, which limits the ability to understand the data from various perspectives. As a result, researchers may miss meaningful relationships among the various indicators that can be extracted from DHI data. Moreover, the lack of comprehensive analysis of DHI usage patterns and health status changes makes it challenging to determine how to improve DHIs according to the needs and preferences of specific groups of users. Given that the data analyses are repeated throughout the development and evaluation lifecycle of DHIs [21], it is necessary to integrate diverse analytical approaches that researchers can perform with DHI data.

Objective

Therefore, this study proposes an analysis task model to help DHI researchers analyze data from a holistic perspective, enabling them to uncover interactions and patterns that are not visible when these factors are analyzed independently. We constructed the model based on data analysis practices employed in existing DHI research, particularly building upon the work of Moshe et al. [22], which reviewed multiple studies on DHIs and their data analysis practices.
These practices indicate the insights that researchers typically aim to derive from DHI data.

The model consists of three main components: (1) user characteristics, (2) user engagement with DHIs, and (3) effectiveness of DHIs on the target symptom (i.e., changes in target symptoms). It includes individual analysis tasks for each component as well as tasks that explore the relationships between these components. Ultimately, the analysis task model would allow researchers to examine which factors specifically influence DHI usage behavior and effectiveness.
Furthermore, based on their understanding of DHI data, researchers may redesign DHIs for specific groups of users or optimize interventions to maximize DHI effectiveness. To further understand how the analysis task model proposed in this study can be utilized, we designed an initial prototype of an interactive visual analytics tool named 'Maum Health Analytics,' which is capable of performing the analyses included in the model. This prototype was provided to researchers with practical experience in DHI data analysis, and we allowed them to use it within a given data analysis scenario. In this process, to illustrate realistic DHI usage behavior and symptom changes, we provided analysis results based on in-the-wild data collected from 'Maum Health,' a mobile DHI service targeting depressive symptoms.

This study aims to evaluate whether the proposed analysis task model is suitable for DHI data analysis and to explore how these tasks can be supported through a visual analytics tool. Our preliminary user study identified several useful aspects of Maum Health Analytics from a Human-Computer Interaction perspective. We found that our tool helped DHI research teams investigate the user engagement and effectiveness of Maum Health from various angles, gain insights for recommending intervention content, and better understand users, especially those who need more care. Moreover, we observed that Maum Health Analytics could facilitate communication between various stakeholders, simplify the repetitive DHI analysis tasks, and potentially integrate with existing analytics tools to maximize its utility.

Overall, our contribution can be summarized as follows:
(1) We proposed an analysis task model for DHIs, aimed at assisting researchers in understanding DHI data from diverse perspectives.
(2) We designed an interactive visual analytics tool based on the proposed model and explored its feasibility and design implications to better support researchers in analyzing DHI data.

In the Methods section, we explain how the analysis task model and interactive visual analytics tool were built, followed by describing the procedure of the preliminary user study with the prototype of the analytics tool. Then, we detail the model and analytics tool we developed in the Results section, along with the user study results, including both quantitative and qualitative evaluations of the prototype. Finally, we discuss the analysis task model and interactive visual analytics for DHI researchers and present design implications.

Methods

Analysis Task Model

The motivation to design a visual analytics tool for DHIs initially stemmed from the authors' own experiences in analyzing data to enhance DHI services. Previously, we developed a mobile DHI service called ‘Maum Health’ to assist individuals struggling with depression, encouraging them to incorporate it into their daily routines for mental health management. Within this service, users could assess their depression symptoms using the CES-D-10 questionnaire [23] and access intervention contents such as art therapy (named “Mandala”), physical activity promotion (named “Geunsimtapa”), and cognitive-emotional games (named “Finding Blue”) to alleviate symptoms. Following a 3-month deployment of the DHI, we analyzed depression states and content usage logs stored on the server to extract data-driven insights for enhancing the intervention service.

As a multidisciplinary DHI research team comprising intervention content designers, clinicians, and system developers, our analysis addressed several research questions, including differences in depression changes based on user characteristics (e.g., by initial depression level) and the association between DHI usage and changes in depression. During the analysis of DHI data, we made three key observations regarding the analysis tasks.

First, DHI researchers primarily focus on exploring user engagement with DHIs and the effectiveness of those interventions when evaluating existing DHI services. They determine that the DHI service is well-designed if it is used more actively and if it leads to improvements in target symptom levels compared to before its use. In addition, given that the DHI was deployed to assess whether it supports achieving the intended outcome (i.e., improving depressive state) in the real world, researchers usually measure effectiveness of DHI instead of its efficacy. Unlike efficacy, which is typically measured from clinical trials involving random user assignment to control and treatment groups (controlled, research setting), effectiveness is evaluated using data from observational studies, allowing researchers to observe how interventions perform in natural settings (uncontrolled, non-research setting) [1]. This measurement is relevant to ‘real-world data (RWD)’ in the healthcare and medical domain, where data is collected without randomly assigning subjects to specific treatment conditions, and researchers conduct observational studies [24, 25].

Also, researchers attempt to investigate whether certain groups of users would exhibit higher (or lower) levels of the above metrics, aiming to examine the relationships between user groups and the metrics. They analyze engagement and effectiveness by user groups specified based on certain criteria and compare them across different groups [8]. While these analyses are valuable, they are often conducted separately, limiting a comprehensive understanding of how different factors such as user characteristics, engagement levels, and effectiveness interact. This fragmented approach can result in missing critical insights that could be gained from a more holistic perspective.

To address this gap, we constructed an analysis task model that aims to integrate these typical analysis task components and facilitate a more thorough understanding of DHI data. Specifically, this model was built upon Moshe et al. [22], a meta-analytic review that assessed multiple studies on

DHIs developed to address depressive symptoms. This review outlined four factors that previous studies have analyzed regarding their association with the effectiveness of DHIs: (1) characteristics of participants, (2) presence of guidance in using DHIs, (3) engagement with DHIs, and (4) study design and quality. Since several existing studies have conducted similar analyses [22], we determined that these analysis tasks are common and should be integrated into our model. However, we excluded two factors from our model: (2) the presence of guidance in using DHIs and (4) study design and quality. This decision was made because our DHI service did not involve any human expert support and we did not aim to demonstrate the efficacy of DHIs through clinical trials.

As a result, we formed our analysis task model with three key analysis components: (1) user characteristics, (2) user engagement with DHIs, and (3) effectiveness of DHIs on the target symptom. Based on these key components, we identified several important analysis tasks by referring to the practices of existing studies. These tasks investigate each component and the interrelationships among them. For example, the model includes not only the exploration of overall user engagement but also the examination of user engagement within specific groups defined by their characteristics or the effectiveness of DHIs, indicating how these components interact. The details of the proposed analysis task model for DHI data analysis and the individual tasks are provided in the Results section.

Interactive Visual Analytics

We designed ‘Maum Health Analytics,’ an initial prototype of an interactive visual analytics tool, to facilitate researchers in easily conducting DHI data analysis tasks from various perspectives. This prototype was formed based on the analysis task model proposed in the previous section, with each feature offered in the tool one-to-one mapped to an analysis task described in the model.

The primary goal of this research is to investigate whether our proposed model adequately covers the DHI data analysis process used by DHI researchers. Therefore, we designed this tool as an instance to evaluate that. As an initial study, we decided to create a medium-fidelity prototype capable of performing the tasks proposed by the model, rather than developing a complete system.

This prototype was designed using Figma, enabling users to explore DHI data through various features (i.e., each analysis task in the proposed model) within predefined data analysis scenarios. The development of the prototype followed an iterative design process. We began with creating an initial prototype to map out the features and functionalities aligned with the individual tasks in the proposed analysis task model. This prototype was then iteratively reviewed and refined based on feedback from domain experts who had experience with DHI data analysis. In the final prototype, we included actual analysis results from the in-the-wild data gathered through our previously deployed DHI (i.e., Maum Health). This provided users with a realistic context when interacting with the prototype during the preliminary user study.

Preliminary User Study

Maum Health Dataset

As mentioned above, we decided to provide the analysis results derived from data collected through our DHI service, 'Maum Health,' when evaluating the interactive visual analytics. Here, we provide a brief description of 'Maum Health' and the data it collected.

Maum Health

Maum Health is a DHI service developed to improve depressive symptoms, offered as a mobile application. Similar to typical mobile health applications, Maum Health provides various intervention contents beneficial for depressive symptoms, including art therapy (named ‘Mandala’), physical activity (named ‘Geunsimtapa’), and a cognitive-emotional screening game (named ‘Finding Blue’). Each intervention content consists of sessions composed of activities that users can perform on their own. For example, users can color the provided drawings, do walking and stretching exercises, and play interactive games when engaging with each content. Moreover, it assesses users' depression levels every two weeks using CES-D-10, a well-known survey in the mental health domain. The result is converted to a 100-point scale, where a higher score indicates more severe depression.

Dataset

Maum Health was provided as a recommended resource for individuals visiting public counseling centers for mental health. As a result, this DHI was used by 529 people over approximately three months, starting from September 2022. During this period, self-reported data entered by users and log data automatically recorded based on the use of intervention contents were collected.

The self-report data in Maum Health included one-time user basic information and periodically entered depressive symptom levels. When users first registered for Maum Health, they provided demographic information such as gender and age, along with information that might be related to depressive symptoms (e.g., marital status, alcohol/smoking experience, etc.). Moreover, during their use of Maum Health, users reported their depressive symptom levels bi-weekly through the CES-D-10 survey.

As users engaged with the three different intervention contents available on Maum Health, log data were recorded automatically. Whenever a user finished a session with specific content, timestamps for the start and end points of that session were labeled. Furthermore, for each session, a completion rate was recorded to indicate how well the user performed the given activities. We used features from both self-report and log data in designing Maum Health Analytics, as illustrated in Table 1.

Table 1. Features of Maum Health data used in the design and evaluation of Maum Health Analytics

Analysis components	Categories	Features
User characteristics	Basic information	Gender, age group, initial depression level
	Additional information	Martial status, cohabitant, occupation, education, economic status, drinking, smoking, army experiences, handedness
	Medical history	Depression history, medication, physical illness
User engagement with DHIs	Mandala (art therapy)	Total launch counts, total usage time, average completion rate
	Geunsimtapa (physical	Total launch counts, total usage time,

	activity promotion)	average completion rate
	Finding blue (cognitive-emotional games)	Total launch counts, total usage time, average completion rate
Effectiveness of DHIs	Depression level	CES-D-10 score

Considering potential quality issues with the in-the-wild data, we established inclusion and exclusion criteria for the data to be used when evaluating Maum Health Analytics. We selected users who had at least two depression score records, allowing us to observe changes in depression levels. Among them, we included users with a gap of 2-4 weeks between evaluations, as they were considered to maintain their DHI use well enough. As a result, we used the data from 173 out of 529 users. Also, recognizing that the initial usage time of Maum Health could vary among users, we considered the relative usage period starting from each user's first day of use.

Study Procedure

Participants

To evaluate Maum Health Analytics, we recruited five groups of experts with experience in analyzing DHI data as participants. Each group comprised 3-5 experts, including clinicians, intervention content designers, and system developers, who worked as a team. The details of these expert groups are shown in Table 2, and we will refer to these experts as ‘DHI researchers’ hereafter in this paper.

Table 2. The composition of participants by DHI research team

Group ID	Participant ID	Expertise of the group
A	A1, A2, A3	Clinicians, Intervention content designers
B	B1, B2	Clinicians, Intervention content designers
C	C1, C2, C3, C4, C5	Intervention content designers
D	D1, D2, D3	Clinicians, Intervention content designers, System developers
E	E1, E2	System developers

Evaluation

During the user study, we first briefly introduced our research and explained Maum Health Analytics, along with the dataset used for the evaluation. Then, we asked them to explore various features of our tool. Since our tool was a medium-fidelity prototype, the participants were only able to navigate the predefined functions, interactions, and corresponding analysis results.

In the evaluation session, we presented participants with specific DHI data analysis scenarios, enabling them to test all tasks supported by Maum Health Analytics. For example, if the scenario was "The expert investigates the distribution of user characteristics among those who frequently used

Mandala above average," the participants would follow several steps within the tool to analyze the data. Considering the limited functionality of the prototype, these scenarios were chosen based on cases with sufficient Maum Health data to fully showcase the results. We visualized analysis results from real-world DHI data to make the evaluation realistic, ensuring that participants could interact with and assess the prototype in a meaningful context.

After using Maum Health Analytics, we conducted an interview session and asked the following key questions: From the perspective of analyzing Maum Health data, (1) whether the analysis tasks defined in our tool are appropriate, and (2) how the information provided could be utilized in practice. We recorded all interview sessions with participants' consent and transcribed them to examine their responses thoroughly. Then, we performed an inductive analysis [26] while repeatedly reading the interview transcripts to identify key phrases, ideas, and themes. After that, we conducted affinity diagramming to group similar themes derived from the transcripts and reviewed the themes iteratively until all the researchers agreed on the final themes.

We also used the Post-Study System Usability Questionnaire (PSSUQ) [27] to evaluate the overall usability, which was designed specifically for scenario-based usability studies. In particular, it included a measure of information quality, allowing us to quantitatively assess whether the defined analysis tasks were suitable.

This study was approved by the Institutional Review Board (IRB) of a university, and we obtained written consent from all participants.

Results

Analysis Task Model

We formed our analysis task model with three components and their interrelationships: (1) user characteristics, (2) user engagement with DHIs, and (3) effectiveness of DHIs on the target symptom. The overall structure of the proposed model is illustrated in Figure 1, and further details for each component and analysis task are provided below.

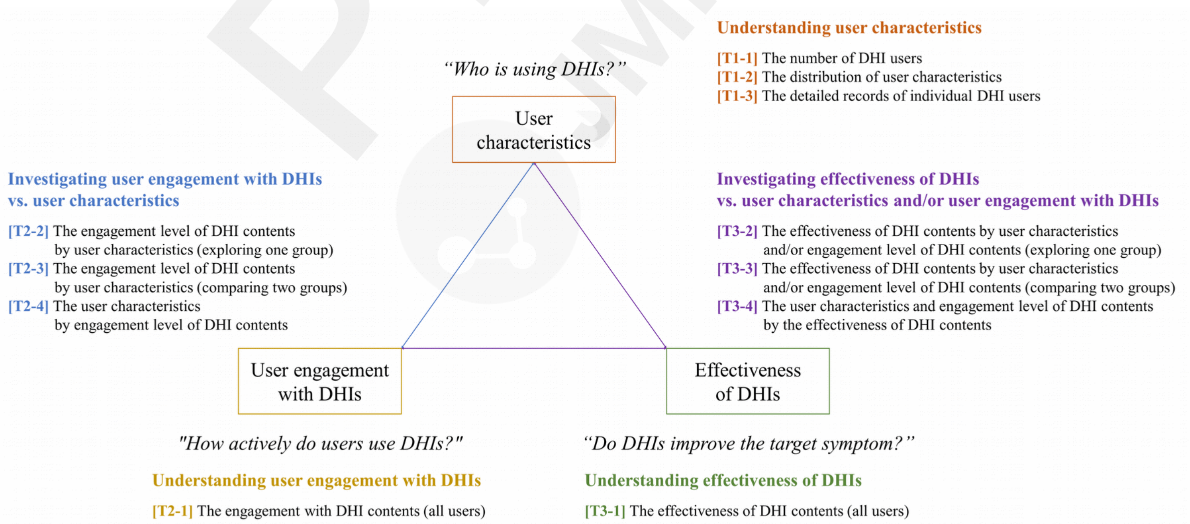


Figure 1. Overview of the proposed analysis task model for DHI data analysis

User Characteristics

Understanding the characteristics of DHI users is crucial because different user groups may lead to differences in DHI usage behaviors and changes in target symptoms. As demonstrated in a large-scale cross-study evaluation by Pratap et al. [28], user characteristics such as gender, age, and geographical or race/ethnic diversity may influence the sustained use of DHIs. Likewise, other studies showed that these characteristics could impact dropout rates from DHIs and changes in outcomes [29, 30].

User characteristics for DHI data analysis can vary, including demographics (e.g., age, gender, ethnicity), physical states (e.g., height, weight), psychological states (e.g., mood, depression, stress, motivation), and more [11]. In addition, factors such as personality traits, digital health literacy, and the availability of time and space for using DHIs could also be considered important constructs of user characteristics [9].

Regarding user characteristics, our model consists of three analysis tasks, which were identified based on a review of existing DHI research and the authors' deployment experiences (as illustrated in the Methods section). First, researchers track the changes in the number of DHI users over time (T1-1). This task is essential for understanding the overall adoption and sustained use of the DHI service. By monitoring the number of enrolled and active users, researchers can assess whether the DHI service is being effectively utilized. Next, the researchers examine the distribution of DHI users based on their characteristics (e.g., gender, age, and baseline symptom levels), as suggested in previous studies (T1-2). This task involves sorting and filtering users based on various characteristics to explore the composition of the user population. It helps in identifying trends and patterns in user engagement and the demographic reach of the DHI service. Finally, researchers review the detailed records of individual DHI users belonging to specific groups determined by user characteristics (T1-3). This task allows for a more granular analysis, providing insights into the behavior and engagement levels of users within specific segments of the population. By following these tasks, researchers can gain comprehensive insights into who is utilizing their DHI service, from a broad overview to detailed individual records.

User Engagement with DHIs

User engagement is an essential metric in DHI research, as highlighted in previous studies. This component is crucial not only for understanding the current activeness of users interacting with DHIs but also for exploring strategies to maintain and enhance their sustained use [31, 32]. Low engagement levels in DHIs are analogous to situations where patients do not take medications properly. Therefore, this metric should be closely monitored to assist DHI users in maintaining their usage until achieving the desired health outcome.

As reviewed by Pham et al. [33], previous studies have used various indicators to measure engagement levels, including the frequency of logins, accessed DHI features and modules, duration of DHI use, and more. Among these metrics, we included the frequency (i.e., launch counts) and duration (i.e., usage time) of DHI content usage in our analysis task model, as they were commonly used in existing studies [11, 34-36]. Furthermore, if there is a structured activity within the intervention content, its completion level can be measured and evaluated as a detailed user engagement metric [37].

For user engagement with DHIs, we include four analysis tasks in the model. First, researchers assess the engagement level of each DHI content across all users (T2-1). This task involves evaluating various engagement metrics for each DHI content, such as launch counts, usage time, and completion levels, as highlighted in previous studies [11]. This provides an overall understanding of how each content is being utilized.

However, considering that user characteristics may influence engagement with DHIs, researchers assess the engagement level of each DHI content for user groups specified by user characteristics (T2-2). This task allows for a segmented analysis based on attributes like age, gender, or baseline symptom levels, which helps in understanding how different user groups interact with the DHI contents. Furthermore, they compare the engagement level of each DHI content across different user groups (T2-3). For example, to determine whether age affects engagement levels, researchers can compare the distribution of engagement metrics among DHI users of different age groups. This comparative analysis helps identify patterns and variations in engagement across user characteristics.

This group-level analysis can also be conducted in reverse; researchers assess the user characteristics distribution for a user group where they attain a certain level of engagement (T2-4). By performing these analysis tasks, DHI researchers can evaluate the extent to which DHI contents are actively used, identify differences in engagement levels among various user groups, and understand the distribution of user characteristics within certain engagement levels.

Effectiveness of DHIs

Since the primary objective of DHIs is to improve the target symptoms, evaluating their effectiveness is necessary. To evaluate the effectiveness of DHIs in real-world settings, we analyze the changes in symptoms by comparing self-reported symptom levels of depression before and after the period of DHI usage. This pre-post comparison allows us to determine whether the intervention has led to a statistically significant improvement in the target symptoms, thus estimating the effectiveness of the DHIs. As explained above, effectiveness is measured in an uncontrolled setting to observe how DHI influences the target symptom in natural environments. For depression management, the following instruments are widely used [38, 39]: the 9-item Patient Health Questionnaire (PHQ-9) [40], the Center for Epidemiologic Studies Depression Scale (CES-D) [41], the Beck Depression Inventory-II (BDI-II) [42], and the Patient-Reported Outcomes Measurement Information System (PROMIS) [43].

Similar to the analysis of user engagement with DHIs, the analysis task model includes four tasks for evaluating the effectiveness of DHIs. First, researchers first assess the changes in depression levels across all users based on their self-reported depression states (T3-1). This task provides an overall measure of the DHI's impact on the target symptom.

Then, they assess the changes in depression levels for user groups (T3-2) and compare these changes across different user groups (T3-3). As illustrated earlier, user characteristics and engagement levels can influence the effectiveness of DHIs. Hence, user groups in tasks T3-2 and T3-3 can be specified based on user characteristics and/or the engagement level of each DHI content. For example, the effectiveness can be evaluated for a user group comprising females who utilized a specific DHI content for more than an hour, or it can be compared among user groups divided by age groups. This segmented and comparative analysis helps identify how the effectiveness of DHIs varies across user characteristics and engagement levels.

The group-level analysis can be conducted in reverse like task T2-4; researchers assess user characteristics and engagement levels distribution for a user group where they attain a certain change in depression levels (T3-4). This task is crucial for understanding which user characteristics and engagement behaviors are associated with significant changes in depression levels. All these tasks serve to assist researchers in exploring the effectiveness of DHIs from multiple perspectives. They allow for the analysis of effectiveness while considering moderating factors, enabling a nuanced understanding of how various user characteristics and engagement levels may influence outcomes. Moreover, they provide ways for tailoring DHIs to different users based on the observed effectiveness across diverse user groups.

Interactive Visual Analytics

Maum Health Analytics, the interactive visual analytics tool we designed, features three main pages. Each page corresponds to one of the three key analysis components: user characteristics, user engagement with DHIs, and the effectiveness of DHIs. On each page, researchers can select specific user group conditions they wish to investigate, and the tool interactively displays the analyzed results. These results are primarily presented through visual elements such as bar charts, line charts, tables, and more, aiding researchers in quickly grasping the overall trends or differences. Moreover, detailed analysis results, such as statistical testing, are provided together to assist researchers in interpreting these findings.

To enhance usability, we included a “tag” feature throughout the entire visual analytics tool. Tagging allows devising researcher-defined groups. For example, a researcher can create a tag for a user group characterized by high engagement levels and severe initial depression states. Once saved, these tags enable quick and easy access to researcher-defined user groups, functioning like a custom favorites list. This allows researchers to efficiently revisit and analyze the same user groups without having to redefine the conditions, streamlining the data analysis process with Maum Health Analytics.

Below, we provide brief explanations for each page of Maum Health Analytics.

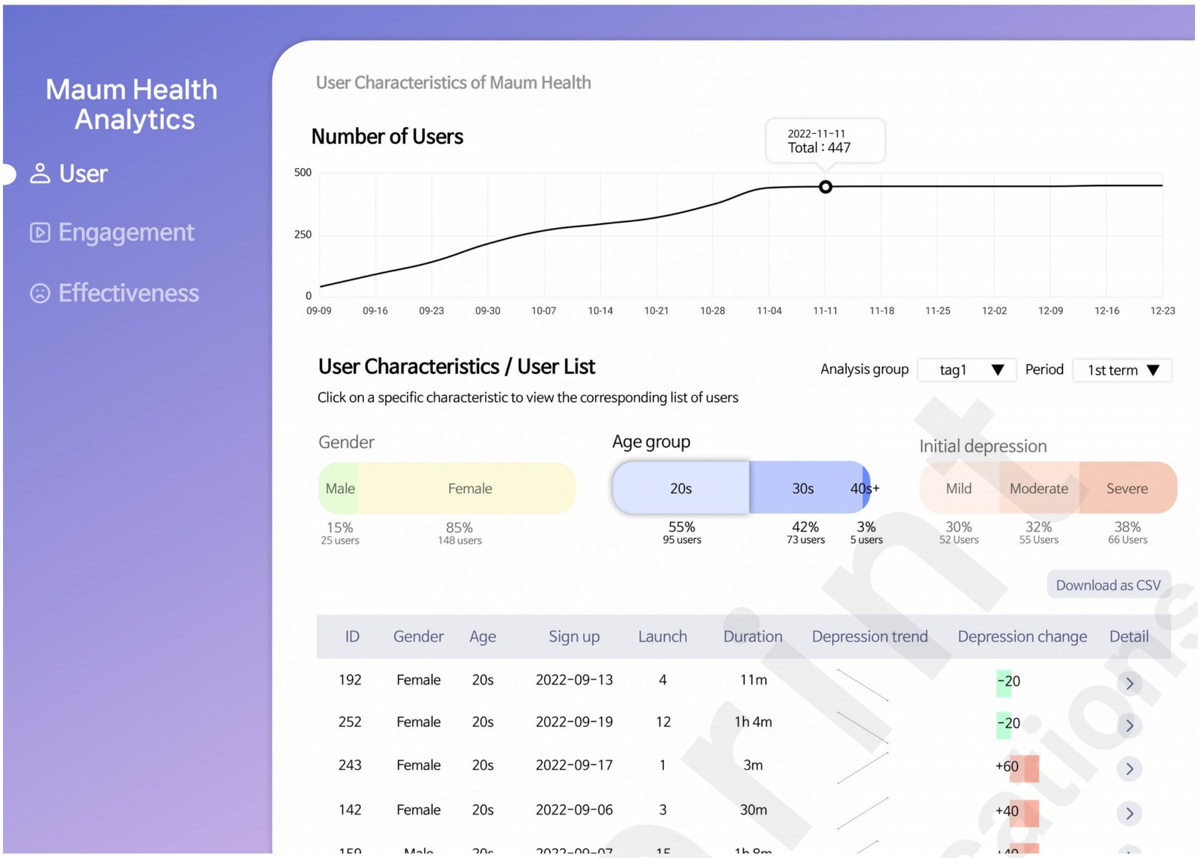


Figure 2. "User" page that supports tasks associated with user characteristics

The "User" page (Figure 2) supports tasks associated with user characteristics, presenting visualizations of the number of DHI users over time (T1-1) and the distribution of users by characteristics (T1-2). Researchers can select specific user characteristics to explore from the horizontal bar chart in the middle, and users matching those criteria are displayed in the table below accordingly. Furthermore, they can pick a specific user from the table to review detailed records (T1-3), including the user’s basic information, engagement levels with DHIs, changes in depression levels, and individual DHI content usage records.

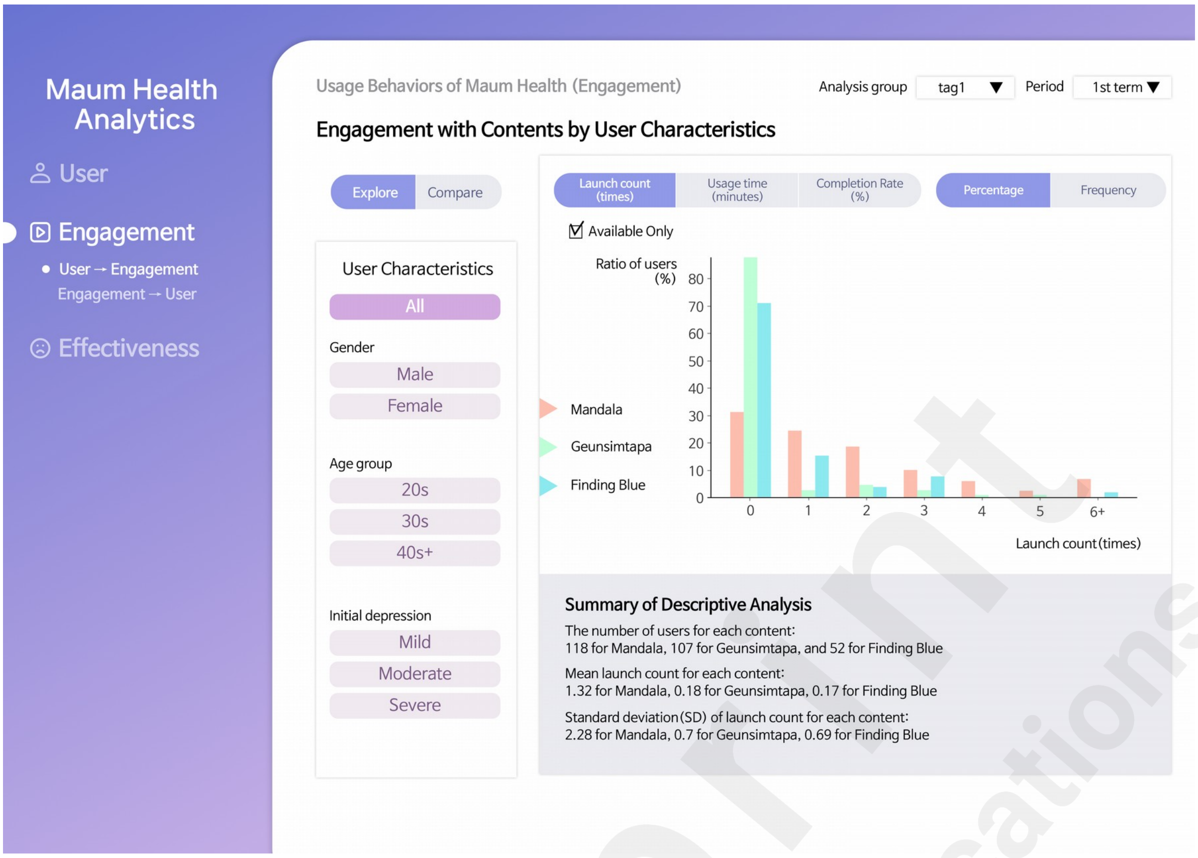


Figure 3. "Engagement" page that supports tasks associated with user engagement with DHIs

On the "Engagement" page (Figure 3), researchers can explore the engagement level of each DHI content across all users (T2-1) or within selected user groups (T2-2) specified by user characteristics. As shown in the figure, a histogram illustrates the distribution of engagement levels within the selected user group, providing insights into the overall usage behavior for each DHI content. Also, researchers are allowed to compare the engagement level across different user groups (T2-3) to identify which users are more engaged with each DHI content. Furthermore, user groups can be formed based on specific engagement levels to examine the distribution of user characteristics within those groups (T2-4).

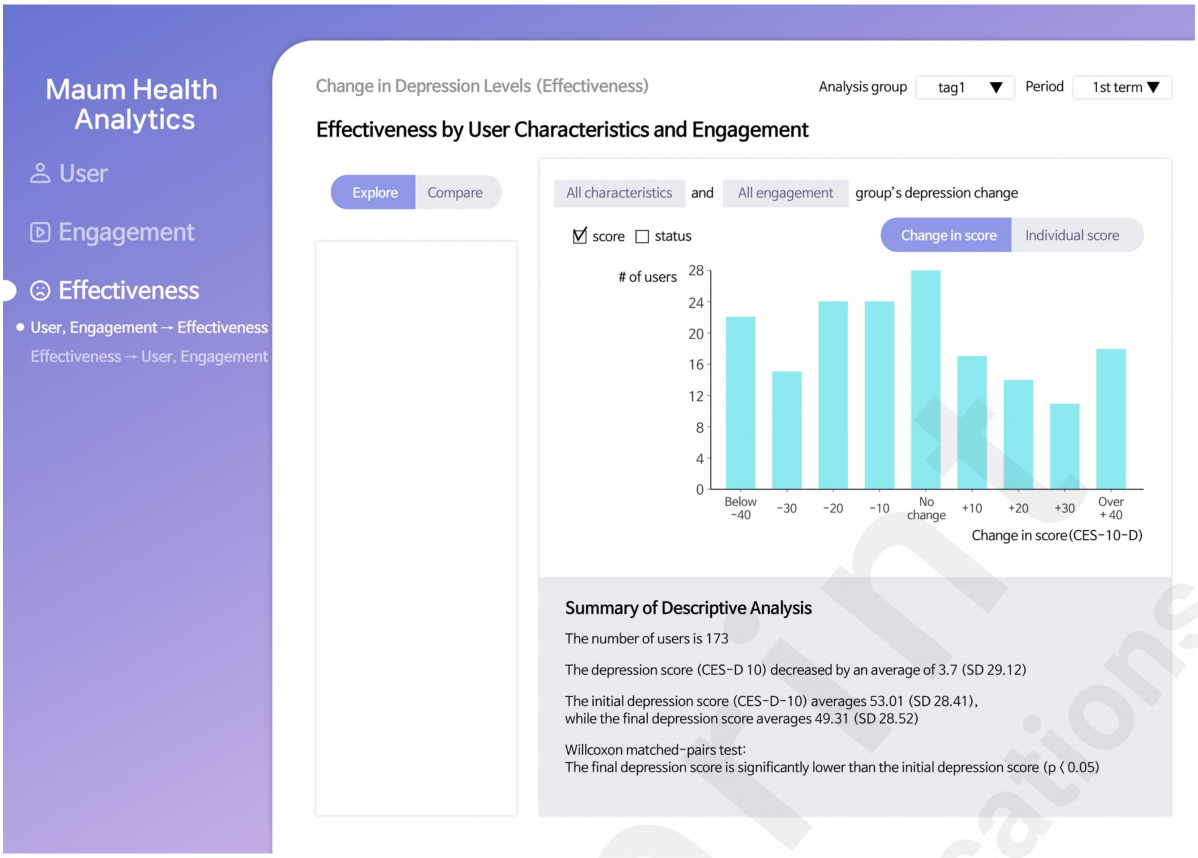


Figure 4. "Effectiveness" page that supports tasks associated with effectiveness of DHIs on the target symptom

Finally, researchers can explore the changes in depression levels on the "Effectiveness" page (Figure 4) either across all users (T3-1) or within selected user groups (T3-2). When specifying user groups for investigation, researchers are allowed to select user characteristics and/or user engagement with DHI contents, and the result is shown in a histogram. Similar to the Engagement page, researchers can compare the changes in depression levels across different user groups (T3-3). Additionally, they can create user groups based on the levels of depression change to understand the distribution of user characteristics and engagement levels within those groups (T3-4).

Further details about the Maum Health Analytics are provided in the Multimedia Appendix 1.

Preliminary User Study

From the preliminary user study with Maum Health Analytics described above, we evaluated both the analysis task model and the interactive visual analytics tool, using quantitative and qualitative methods.

Quantitative Evaluation

The PSSUQ questionnaire was composed of 3 subscales with 16 items that evaluated system usefulness, information quality, and interface quality. Each item is rated on a 7-point Likert scale (1: Strongly Agree, 4: Neutral, 7: Strongly Disagree) with lower scores representing better usability. When we calculated the PSSUQ score, we excluded two items (7 and 8) related to error recovery due to a lack of test cases.

For Maum Health Analytics, the mean of PSSUQ was 2.89 (SD: 1.02) and that of overall usability was 2.80 (SD: 1.57), meaning that participants mostly positively evaluated the usability of our analytics tools. Each subscale showed similar mean scores: system usefulness = 2.94 (SD: 1.01), information quality = 2.87 (SD: 1.21), and interface quality = 2.84 (SD: 1.31).

Qualitative Evaluation: Analysis Task Model

Our interviews revealed the participants' opinions on the analysis task model and its components proposed in this study.

User Characteristics

Participants found that tasks related to ‘user characteristics' enhanced their understanding of user groups. They noted the value in exploring individual usage patterns not visible in group-level analysis:
“When I discovered that females used Mandala more, I became curious about each female user’s usage patterns. Reviewing their specific usage logs could help me identify common characteristics of this user group” (A1).
They were keen on investigating representative users to gain better insights into the group. Previously, this was difficult, especially when activities from DHIs were recorded only on the user’s device, limiting access to the data:
“We could analyze DHI usage statistics by group, but it was challenging to check detailed usage logs since we had to take their phones for analysis” (A3).
With our tool, they could easily access these logs, allowing for an in-depth examination of the user group with sample cases.

In addition, the participants reported that our tool is useful for investigating DHI users (or user groups) needing management. They aimed to use our analytics tool to assess if DHI users met the minimum criteria for launch counts or usage time within a given period:
“For instance, if there is a weekly recommended use, I’d like to see how many users in the group meet that goal” (B1).
In the user characteristics section, participants skimmed the launch counts and usage time of a specific group and showed interest in analyzing individuals with low engagement.
Clinicians focused on managing DHI users to improve their target symptoms, not just on improving the DHI itself:
“I was wondering why this user didn’t use Maum Health contents that much, and why her status got worse even while using intervention contents” (A1).
They planned to intervene with low-engagement users to increase their engagement:
“Perhaps I need to decide how to manage the low engagement users, for instance, send a text message, make a call, and so on” (B1).
Likewise, encouraging DHI users to stay engaged was seen as crucial, and they believed our tool's insights could help improve DHI by providing such support.

Some participants mentioned that tags could help monitor DHI users needing support:
“In our study, there are individuals whom doctors have to care for more intensively. Previously, we had to find such users whenever we analyzed them. But here, I think it would be easier and more convenient because we can mark them with tags and monitor them accordingly” (C2).
E1 viewed tags as customizable user characteristics and noted that using them like favorites would make tracking certain users easier.

The participants also mentioned that our tool can track newly incoming users and retain existing ones. Monitoring the characteristics of both new and existing users is crucial to effectively provide intervention content:

“There is a difference between user groups by their retention rate. Based on that information, we could determine the frequency of exposing specific content for each user group, which might be available here, too” (D1).

However, participants emphasized the need to examine ‘active users’ separately, suggesting a point to be improved for our tool. Identifying active users is more critical for intervention content designers than knowing the total number of registered users:

“I’d like to see how many users are actively using DHI, not just registered. For designers, it is important to see whether they are using DHIs well enough” (C1).

To determine if an individual is an active user, they suggested including withdrawn records for each intervention session on the individual user's page:

“I think the quitting ratio while using content is also important. The intervention content may be interesting at the beginning, but it gets boring as time goes” (C3).

User Engagement with DHIs

Participants stated that Maum Health Analytics would help them identify specific user groups that are likely to benefit from certain intervention content and recommend its use effectively. In particular, they found it beneficial to know the distribution of characteristics among highly engaged users, which is related to task T2-4. For example, when extracting users who used Mandala above average, the proportion of females increased, which participants found useful for making a decision. Based on this information, they believed that they could further encourage highly engaged users. A1 noted:

“After examining the characteristics of users who use a particular content often, we can recommend it to similar users, such as those in the same age group. Maybe we can promote the DHI service by targeting similar groups of people through social media.”

They assumed similar engagement levels for users with similar characteristics and thought the tool could improve content recommendation methods, which currently rely only on depression levels:

“I think this tool can be used to make feedback more customized while monitoring the changes in DHI engagement” (A2).

Conversely, some participants believed users with low engagement should be further analyzed to design additional content for them. A3 said:

“If the proportion of females in users with high engagement is significant, we should create another content targeting men. Otherwise, the male user group may not be supported well and might leave the Maum Health app.”

However, B1 added that this recommendation method should also consider whether there is an actual improvement in depression, even with high DHI engagement.

Some participants mentioned that presenting the overall distribution of engagement with intervention content was useful and could help set guidelines for DHI use. They noted that interventions like traditional Cognitive Behavior Therapy (CBT) can be easily evaluated in terms of engagement because the number of sessions is predefined. However, setting evaluation criteria is challenging when users use the content independently without specific guidelines. They expected that with long-term DHI data accumulation, they could establish engagement guidelines using this tool:

“If the data becomes sufficient, values such as average or top 25% could be more meaningful. Observing the overall engagement level with this tool, we can provide engagement guidelines for each step or set a recommended use” (B1).

This suggests a method for determining the optimal dosage of DHI by leveraging data from our tool, where clear criteria are currently lacking.

The system developer group emphasized the need to carefully select usage metrics like launch counts and usage time when measuring user engagement:
"For example, there are cases when users stop using content when getting a call or eating. In those cases, they stayed in the content but there was no interaction recorded in the usage log" (E1).
Participants wanted to distinguish what constitutes 'effective' engagement, focusing on launch counts and usage time when users actively engaged with the DHI. As E2 noted:
"It is necessary to analyze how much users actively participated in the content before investigating its effectiveness. Then, we have to further examine what was behind if DHI turned out to be ineffective."

Unlike other expert groups, system developers primarily focused on improving usability and maintaining overall engagement levels. Therefore, they were less interested in examining whether the DHI effectively improved target symptoms. Instead, they prioritized analyzing collected content usage logs and determining metrics that better reflect user behavior.

Effectiveness of DHIs

Participants found Maum Health Analytics helpful for systematically evaluating the effectiveness of DHIs. Since DHI aims to manage the user's health, it is crucial to check its effectiveness by measuring changes in target symptoms. They found it meaningful to investigate changes in target symptoms based on engagement with each content. D1 noted:
"Of course, showing individual indicators can be a meaningful analysis for almost all apps. However, for DHIs, we should investigate which intervention features interact with each other and how they improve symptoms, requiring a relational analysis of multiple factors."
E1 added:
"For example, for game apps, the longer the usage time, the better. But in the case of DHIs, we need to see not only how long it has been used, but also how effective it is. I think this tool supports this process well enough."

The participants compared Google Analytics with our tool, highlighting a key difference in purpose. They explained that the key distinction of our tool is the inclusion of health-related indicators, making it more suitable for analyzing DHI services:
"Unlike Google Analytics, which mainly focuses on usage behavior and user retention, we can analyze DHI including the users' health indicators in this tool. This is critical for health-related services" (D1).

Participants reported that evaluating content effectiveness from various perspectives using the visual analytics tool was beneficial and might help answer research questions when designing and studying intervention content. C1 noted:
"Previously, it was inconvenient to repeatedly perform statistical analysis for each condition, so we sometimes skipped examining some questions in detail. Perhaps, we lost opportunities to find something new."
However, the tool allowed them to easily explore the effectiveness of DHIs from multiple angles. They appreciated features that compared effectiveness by user characteristics and engagement with the content, enabling them to explore the effectiveness across various user groups:
"Like the analysis in research, I was able to create control groups and experimental groups based on several criteria and easily compare the difference in depression change between the groups. I could see the relationship between engagement and effectiveness through this process" (C2).

They also liked the tag function, which enabled comparisons between user groups of interest. A2 responded:

"I like the tag because we can immediately see how much DHI has improved after updates if we label tags based on the version of DHI."
C1 added that exploring analysis results by tags would help the research team formulate and evaluate new hypotheses, similar to A/B testing.

In addition, participants found it useful to evaluate other indicators based on the level of depression change. Since DHI services are still in the early stages, there may be no clear evidence that increased usage leads to better outcomes. Therefore, B1 suggested it might be more insightful to investigate other indicators based on DHI effectiveness at this point.

Qualitative Evaluation: Interactive Visual Analytics

We also explored the participants' experiences with the interactive visual analytics tool itself during the interview sessions.

Simplifying the Analysis Tasks of DHI

Participants mentioned that using Maum Health Analytics would alleviate many cumbersome tasks in the existing data analysis process. They particularly appreciated not having to download the dataset whenever they analyze DHI data, recognizing this as the most practical feature:
"Our lab is developing a similar DHI app, but we have to download and analyze the entire dataset since there's no analytics tool like this. But here, we can check what happened to the user right away without such tasks" (A3).
Since most existing dashboards were used only for managing participants during experiments, participants liked that our tool enables efficient basic analysis of DHI.

In addition to simplifying analysis, the ability to view real-time results through our analytics tool was also highly valued by participants:
"I have experience in developing chatbot-based interventions, and I wanted to know in real-time whether users are using them well or how they respond. I was satisfied to see that information on this tool" (A3).
B1 noted that real-time information could be useful for various stakeholders:
"Real-time information may be meaningful to those who provide counseling services based on analyzed results, those who manage certain groups of users, and those who monitor the entire service."
Participants also appreciated the ability to extract and export specific groups of users for further analysis. A2 and A3 highlighted their interest in analyzing outliers in detail and liked that data for only those users could be downloaded separately. They added that this data could be further investigated using existing general-purpose analytic tools.

Supporting Communication Between Various Experts

Participants assessed that Maum Health Analytics would facilitate communication among stakeholders during the DHI improvement process. They thought this tool would be critical, especially when experts from different domains collaborate on developing and evaluating DHIs.
Previously, content designers had to create separate documents to explain analysis results to other domain experts, which was inconvenient:
"As a content designer, I always had to prepare additional materials about the analysis results to explain it to mental health counselors. However, with these visual explanations, I think they can understand the results easily without any further documents" (A3).
They added that those with limited knowledge and experience in data analysis could also perform various analyses using our tool:

“Since this tool showed the results visually, I thought it would be suitable for those who are not familiar with how to analyze data well. With this visual analytics tool, they would be able to answer various questions related to DHI usage” (A2).

In addition, if there is a division of roles between people who recruit DHI users and those who develop and analyze DHI services, this tool would allow them to quickly track who is currently using DHI.

“If someone else is recruiting DHI users, there will be difficulties in the analysis since we have to examine the data without understanding the subjects. This tool displays that information right away, making it easy to track DHI users’ information even if we don’t recruit the subjects by ourselves” (C1).

Providing Statistical Analysis and Interpretation of Results

Participants reported that Maum Health Analytics was useful for providing statistical analysis results when evaluating the engagement and effectiveness of DHIs. They emphasized the importance of statistical testing as an objective basis for presenting results:

“There must be people who think statistics are important. For example, in the public domain of managing these studies, it is important to leave a numerical rationale” (B1).

“If we say that this result is statistically significant, I think we can clearly communicate it to experts in other fields” (C3).

The statistical analysis information was meaningful, especially since most results are visually expressed, such as in graphs.

However, there was feedback that interpreting these results should include more insightful explanations for improving DHIs. Descriptive statistics such as mean and standard deviation can be roughly interpreted through graphs. Therefore, some participants suggested that presenting statistical testing results along with practical interpretations would be beneficial:

“When considering people who are not familiar with statistics may see statistical analysis, it would be nice to provide interpretations related to decision-making, such as what this result means and how to utilize it in updating DHIs” (C1).

Enhancing Continuity Between Analysis Tasks

Participants commented that transition between different analysis tasks should be improved by allowing the user group currently under analysis to be pinned. They wanted to pin their group of interest and see all analysis results for that group continuously:

“For example, if I went to another page after checking the engagement distribution of Mandala, I wanted to view the analysis results of the same user group first” (B1).

E1 particularly pointed out the need for a feature to save the specified user group:

“We first narrowed the user group down by adding more filters, but we had to find and select the conditions again if we moved to other pages. As a result, the analysis tasks seemed somewhat disconnected.”

They suggested that the ability to save and pin the specified user group could serve as a connector between analysis tasks, enabling a seamless experience of analyzing data from DHIs. This feature would ensure that researchers can maintain their focus on the same user group across different analysis tasks and pages.

Discussion

Principal Findings

This study proposed an analysis task model for practitioners to explore and analyze data generated from DHIs. Building upon previous studies related to DHIs, we developed a task model comprising key components, including user characteristics, user engagement with DHIs, and the effectiveness of DHIs. The model was composed of analysis tasks that explore not only each component individually but also their interrelationships. In addition, we designed an interactive visual analytics tool based on the proposed model to investigate how the model and tool could support DHI data analysis practices. To make the evaluation more realistic, we visualized analysis results from real-world data collected from a currently operational DHI service (named Maum Health) on the prototype and examined the DHI researchers' user experience while using it.

Below, we provide several discussions on the analysis task model and the interactive visual analytics tool, and suggest design implications for supporting researchers when analyzing DHI data.

Analysis Task Model

The analysis task model allowed the DHI research team to analyze data holistically, addressing the limitations of existing studies that analyzed each key component separately. Our model supported the analysis of relationships between key components by specifying user groups based on certain conditions (e.g., gender) and comparing target outcomes (e.g., changes in depression) between them. During this process, the research team could identify meaningful insights for improving existing DHIs, such as suggesting content to potential target users, setting criteria for DHI usage, and determining better interventions. This aligns with recent studies investigating which DHI usage behaviors (i.e., engagement) affect target outcomes [8, 44], highlighting the importance of further analyses of these relationships.

In real-world settings, DHIs can be distributed through app markets to many individuals with varying characteristics, contexts of use, and engagement levels [45]. Therefore, as suggested in our model, it is critical to evaluate DHI usage behaviors and changes in the target symptom for each group of similar users. Group-level analysis has been widely employed in previous studies dealing with large-scale health data [13, 22, 46]. These studies explored overall trends or patterns at the user group level and derived insights to address common issues within the group. Given that DHI data can be collected from a large number of users, DHI researchers may follow similar analysis practices.

At the same time, records of individual users could be provided as sample cases to complement the understanding of a specific user group. Since it is challenging to review all records, the analysis task model can be improved to suggest representative users of that group. For instance, they could be users showing mean or median engagement levels or those with the most frequent user characteristics. This approach helps researchers gain a more nuanced understanding of user behaviors and outcomes, allowing for more tailored and effective interventions.

Considering the researchers' feedback on our prototype that they would set guidelines after exploring engagement levels for each user group, the group-level analysis may also be utilized to determine the optimal dosage. In practice, determining the dosage of typical DHIs is difficult due to various behavioral and contextual factors influencing dose-response relationships [14, 16]. Our model could support researchers in investigating how much DHI usage is required to achieve a certain level of improvement in the target symptom. Based on this insight, they can further improve existing DHIs by adjusting the dosage or revising the difficulty of tasks to be performed in the intervention content while monitoring how these changes affect the effectiveness of DHIs.

Interactive Visual Analytics

The interactive visual analytics for DHIs proposed in our study demonstrated the potential to allow researchers to perform analysis tasks efficiently. As revealed in the study, researchers often repeated similar DHI data analyses, but there was a lack of research on visual analytics tools to streamline this process. Maum Health Analytics enabled researchers to analyze DHI data from diverse perspectives in real time through simple interactions.

The proposed analytics tool can be utilized at various stages of the development and evaluation lifecycle of DHI services, which typically follow an agile development process [21]. This process involves several stages of design and testing, with results from testing fed back into the previous cycle to improve DHIs continuously. Therefore, researchers need to keep monitoring and evaluating DHI data, and our tool could make this iterative process more efficient.

We also discovered that the interactive visual analytics tool could be useful for DHI research teams consisting of experts from diverse domains and backgrounds, such as clinicians, intervention content designers, and system developers. As shown in the study, their purposes for data analysis may vary due to differences in their backgrounds. For example, clinicians value whether there has been improvement in target symptoms, while system developers focus more on the log data generated from DHI.

In such situations, the analyses provided by our analytics tool allow researchers to explore unfamiliar aspects of the DHI data and extend their knowledge about the data. This facilitates communication between researchers from different fields, generating diverse approaches to better support users and improve existing intervention components. Furthermore, the analytics tool can be used to quickly share results with stakeholders outside the research team, such as funding bodies or external project managers, aiding their decision-making process.

Design Implications

Based on the findings, we suggest further exploring ways to quantify DHI usage behaviors and diversifying the metrics used to represent user engagement with the interventions. As reported in the study, some DHI users might simply run the intervention content without actively engaging with it. Relying solely on traditional high-level engagement metrics such as launch counts or usage time may not distinguish this kind of passive usage, which could negatively affect the quality of the analysis.

Therefore, we propose incorporating more detailed interaction metrics to better understand how actively users are engaging with the DHIs. For example, instead of just measuring how often a DHI is launched or how long it is used, we could track specific actions like touches, swipes, or clicks within the app to get a clearer picture of user engagement [12]. Moreover, if the DHI includes activities that users are supposed to perform, the system can track adherence to these activities and incorporate this data into the engagement metrics. By doing so, we can identify more meaningful (or effective) engagement and analyze its impact on improving the target symptoms. This approach would provide a more nuanced understanding of user behavior and the true effectiveness of the interventions.

In addition, we suggest design implications for interactive visual analytics tools for DHI researchers. Given that multiple user groups can be formed from the data, it is challenging to determine which groups should be investigated. Therefore, the tool can be designed to proactively propose important user groups that should be examined by researchers. For example, the analytics tool could recommend user groups with very high engagement or those with very low effectiveness, as these may indicate critical insights. By leveraging the data analysis history stored in the analytics tool, it can provide analysis results that researchers are interested in or guide them to discover other important user groups.

Furthermore, we can support user group selection through natural language, making it easier for researchers to locate the groups they want to explore. This could be a practical alternative, especially if numerous metrics are available for analyzing user engagement with DHIs and their effectiveness. At the same time, enabling researchers to explore various factors based on the specified user group (i.e., tag function in this study) would enhance the user experience of visual analytics by allowing for seamless DHI data analysis.

When designing the visual analytics tool, both exploratory and explanatory aspects should be carefully considered. As a multidisciplinary research team, members may have varying levels of data knowledge and experience in data analysis. Therefore, providing interpretations of the analyzed results is crucial for them to apply data-driven insights in practice. For example, our tool can be further improved by explaining statistical results and offering practical guidelines based on them.

With recent advances in large language models (LLMs), visual analytics could generate reports based on the data analysis history, enhancing researchers' understanding of the DHI data [47]. As suggested in a recent study [48], data presented in visual elements can also be accompanied by a data-driven narrative generated by LLMs, effectively aiding communication about the data. By integrating these features, the visual analytics tool can become a more powerful and user-friendly resource for multidisciplinary research teams, facilitating deeper insights and more effective decision-making in the development and evaluation of DHIs.

Limitations and Future Works

In this study, we conducted a preliminary user study on a visual analytics tool designed based on findings from our research experience and previous studies. The analytics tool presented in the study was a preliminary prototype, offering only limited interactions for exploratory purposes. In future work, it will be necessary to develop a fully functional system based on this prototype and evaluate it in comparison with existing analytics tools. During the evaluation process, it will also be essential to include DHI data generated over multiple periods to allow users to assess how engagement and effectiveness change over time, as frequently occurs in practice.

Investigating how this analytics tool and the foundational analysis task model are employed in real-world settings will be an important area of future research for effectively supporting DHI research teams. By addressing these aspects, we aim to create a comprehensive and robust analytics tool that significantly contributes to the field of digital health interventions.

Conclusions

We proposed an analysis task model to support DHI researchers by incorporating analysis practices from existing DHI research. Based on this model, we designed a prototype of interactive visual analytics to explore how our task model and tool can aid DHI data analysis tasks. The preliminary user study demonstrated that our model enables a holistic investigation of DHI data by integrating three key components and examining their relationships. Moreover, the analytics tool showed potential in simplifying repetitive tasks and facilitating communication among researchers with diverse backgrounds and interests.

As the use of DHIs continues to grow, research on enhancing data analysis and supporting decision-making for improving the existing DHIs is increasingly important. In this context, we expect that the model and tool proposed in our study would effectively meet the data analysis needs of DHI researchers and make the process more efficient.

Acknowledgments

This research was supported by the Bio & Medical Technology Development Program and the Basic Science Research Program of the National Research Foundation (NRF), funded by the Korean government (MSIT) (No. 2021M3A9E4080780) and by the NIPA (National IT Industry Promotion Agency), funded by the Korean government (MSIT) (No. H1001-24-1002).

Authors' Contributions

All authors participated in the conception of the study. GJ, HL, and UL developed the task analysis model and designed the interactive visual analytics, with contributions from KH and HK. KH and HK collected the Maum Health data, while GJ, HL, and UL analyzed and prepared it. The preliminary user study was conducted by GJ and HJ, with feedback from UL, KH, and HK. GJ prepared the manuscript, and all authors read, commented on, and contributed to the final manuscript.

Conflicts of Interest

None declared.

Abbreviations

BCT: Behavior Change Technique
CBT: Cognitive Behavior Therapy
DHI: Digital Health Intervention
LLM: Large Language Model
mHealth: Mobile Health

Multimedia Appendix 1

Description of Analysis Tasks Supported by Maum Health Analytics

References

1. World Health Organization., *Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment.* 2016.
2. Kowatsch, T., et al., *A design and evaluation framework for digital health interventions.* It-Information Technology, 2019. **61**(5-6): p. 253-263.
3. Murray, E., et al., *Evaluating Digital Health Interventions: Key Questions and Approaches.* Am J Prev Med, 2016. **51**(5): p. 843-851.
4. Dugas, M., G.D. Gao, and R. Agarwal, *Unpacking mHealth interventions: A systematic review of behavior change techniques used in randomized controlled trials assessing*

mHealth effectiveness. Digital Health, 2020. **6**.
5. Michie, S., et al., *The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions.* Annals of Behavioral Medicine, 2013. **46**(1): p. 81-95.
6. Michie, S., et al., *Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data).* Health Technology Assessment, 2015. **19**(99): p. 1-+.
7. Lee, U., et al., *Toward Data-Driven Digital Therapeutics Analytics: Literature Review and Research Directions.* Ieee-Caa Journal of Automatica Sinica, 2023. **10**(1): p. 42-66.
8. Alshurafa, N., et al., *Is More Always Better? Discovering Incentivized mHealth Intervention Engagement Related to Health Behavior Trends.* Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2018. **2**(4): p. Article 153.
9. Borghouts, J., et al., *Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review.* Journal of Medical Internet Research, 2021. **23**(3).
10. Choi, W., et al., *Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention.* Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2019. **3**(2): p. Article 39.
11. Perski, O., et al., *Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis.* Translational Behavioral Medicine, 2017. **7**(2).
12. Cole-Lewis, H., N. Ezeanochie, and J. Turgiss, *Understanding Health Behavior Technology Engagement: Pathway to Measuring Digital Behavior Change Interventions.* JMIR Form Res, 2019. **3**(4): p. e14052.
13. Chen, A.T., et al., *A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention.* Journal of Biomedical Informatics, 2019. **94**.
14. Donkin, L., et al., *A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies.* Journal of Medical Internet Research, 2011. **13**(3).
15. Fleming, T., et al., *Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety.* J Med Internet Res, 2018. **20**(6): p. e199.
16. Mclaughlin, M., et al., *Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis.* Journal of Medical Internet Research, 2021. **23**(2).
17. Lipschitz, J.M., et al., *Digital Mental Health Interventions for Depression: Scoping Review of User Engagement.* J Med Internet Res, 2022. **24**(10): p. e39204.
18. Doherty, G., D. Coyle, and J. Sharry, *Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, Association for Computing Machinery: Austin, Texas, USA. p. 1421–1430.
19. Morrison, C. and G. Doherty, *Analyzing Engagement in a Web-Based Intervention Platform Through Visualizing Log-Data.* Journal of Medical Internet Research, 2014. **16**(11).
20. Thieme, A., et al., *Designing Human-centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment.* Acm Transactions on Computer-Human Interaction, 2023. **30**(2).
21. Wilson, K., et al., *Agile research to complement agile development: a proposal for an mHealth research lifecycle.* Npj Digital Medicine, 2018. **1**.
22. Moshe, I., et al., *Digital Interventions for the Treatment of Depression: A Meta-Analytic*

Review. Psychological Bulletin, 2021. **147**(8): p. 749-786.
23. Andresen, E.M., et al., *Screening for Depression in Well Older Adults - Evaluation of a Short-Form of the Ces-D.* American Journal of Preventive Medicine, 1994. **10**(2): p. 77-84.
24. Liu, X., *Real-World Data for the Drug Development in the Digital Era.* Journal of Artificial Intelligence and Technology, 2022. **2**(2): p. 42-46.
25. Taur, S.R., *Observational designs for real-world evidence studies.* Perspect Clin Res, 2022. **13**(1): p. 12-16.
26. Patton, M.Q., *Enhancing the quality and credibility of qualitative analysis.* Health Services Research, 1999. **34**(5): p. 1189-1208.
27. Sauro, J. and J.R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*. 1 ed. 2012: Morgan Kaufmann. 312.
28. Pratap, A., et al., *Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants.* Npj Digital Medicine, 2020. **3**(1).
29. Karyotaki, E., et al., *Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis.* Psychological Medicine, 2015. **45**(13): p. 2717-2726.
30. Karyotaki, E., et al., *Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms A Meta-analysis of Individual Participant Data.* Jama Psychiatry, 2017. **74**(4): p. 351-359.
31. Baumel, A., et al., *Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis.* Journal of Medical Internet Research, 2019. **21**(9).
32. Eysenbach, G., *The Law of Attrition.* Journal of Medical Internet Research, 2005. **7**(1).
33. Pham, Q., et al., *A Library of Analytic Indicators to Evaluate Effective Engagement with Consumer mHealth Apps for Chronic Conditions: Scoping Review.* Jmir Mhealth and Uhealth, 2019. **7**(1).
34. Short, C.E., et al., *Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies.* J Med Internet Res, 2018. **20**(11): p. e292.
35. Sieverink, F., S.M. Kelders, and J.E. van Gemert-Pijnen, *Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence.* J Med Internet Res, 2017. **19**(12): p. e402.
36. Taki, S., et al., *Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index.* JMIR Mhealth Uhealth, 2017. **5**(6): p. e89.
37. Molloy, A. and P.L. Anderson, *Engagement with mobile health interventions for depression: A systematic review.* Internet Interv, 2021. **26**: p. 100454.
38. Choi, S.W., et al., *Establishing a Common Metric for Depressive Symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS Depression.* Psychological Assessment, 2014. **26**(2): p. 513-527.
39. Wahl, I., et al., *Standardization of depression measurement: a common metric was developed for 11 self-report depression measures.* Journal of Clinical Epidemiology, 2014. **67**(1): p. 73-86.
40. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* J Gen Intern Med, 2001. **16**(9): p. 606-13.
41. Radloff, L.S., *The CES-D Scale: A self-report depression scale for research in the general population.* Applied Psychological Measurement, 1977. **1**(3): p. 385-401.
42. Beck, A.T., et al., *Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients.* Journal of Personality Assessment, 1996. **67**(3): p. 588-597.
43. Pilkonis, P.A., et al., *Item Banks for Measuring Emotional Distress From the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, Anxiety,*

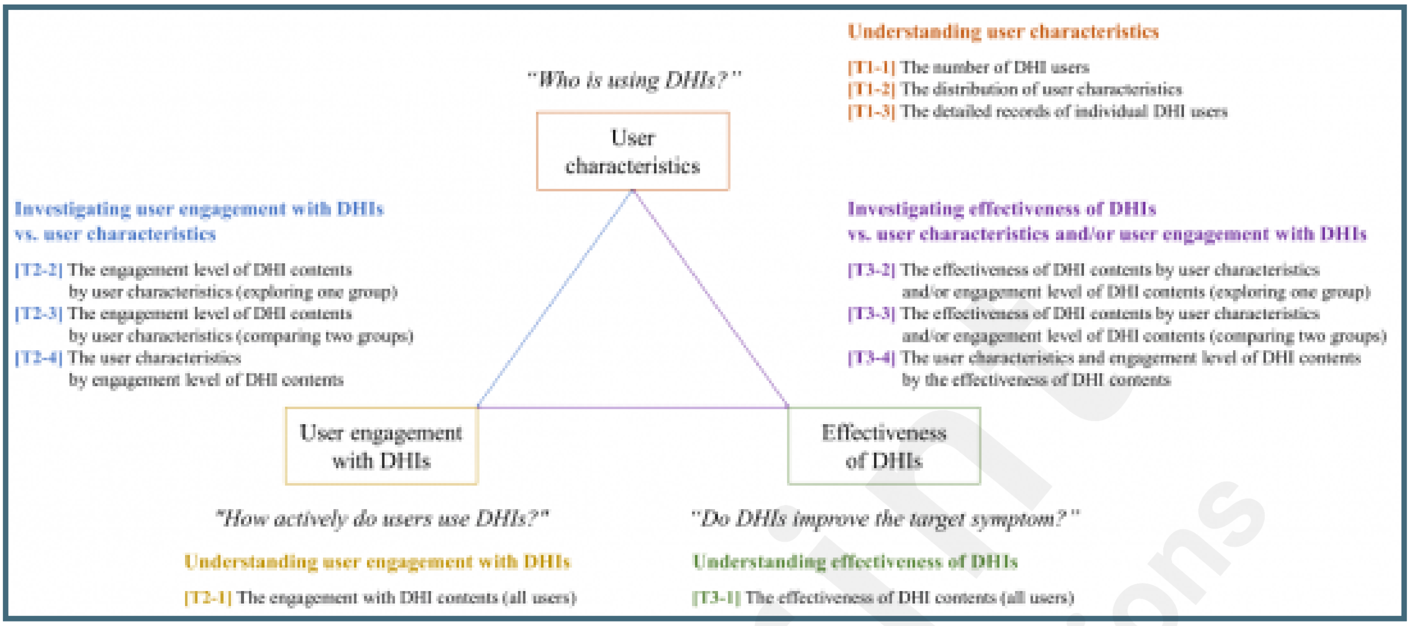
and Anger. Assessment, 2011. **18**(3): p. 263-283.

44. Strauss, G., et al., *Meaningful engagement: A crossfunctional framework for digital therapeutics.* Frontiers in Digital Health, 2022. **4**.
45. Grundy, Q., *A Review of the Quality and Impact of Mobile Health Apps.* Annual Review of Public Health, 2022. **43**: p. 117-134.
46. Kelders, S.M., E.T. Bohlmeijer, and J.E.W.C. Van Gemert-Pijnen, *Participants, Usage, and Use Patterns of a Web-Based Intervention for the Prevention of Depression Within a Randomized Controlled Trial.* Journal of Medical Internet Research, 2013. **15**(8).
47. Cao, S.X., Q. Chen, and N. Cao, *Visual narrative for data journalism based on user experience.* Journal of Visualization, 2024.
48. Zhao, Y., et al., *LEVA: Using Large Language Models to Enhance Visual Analytics.* IEEE Trans Vis Comput Graph, 2024.

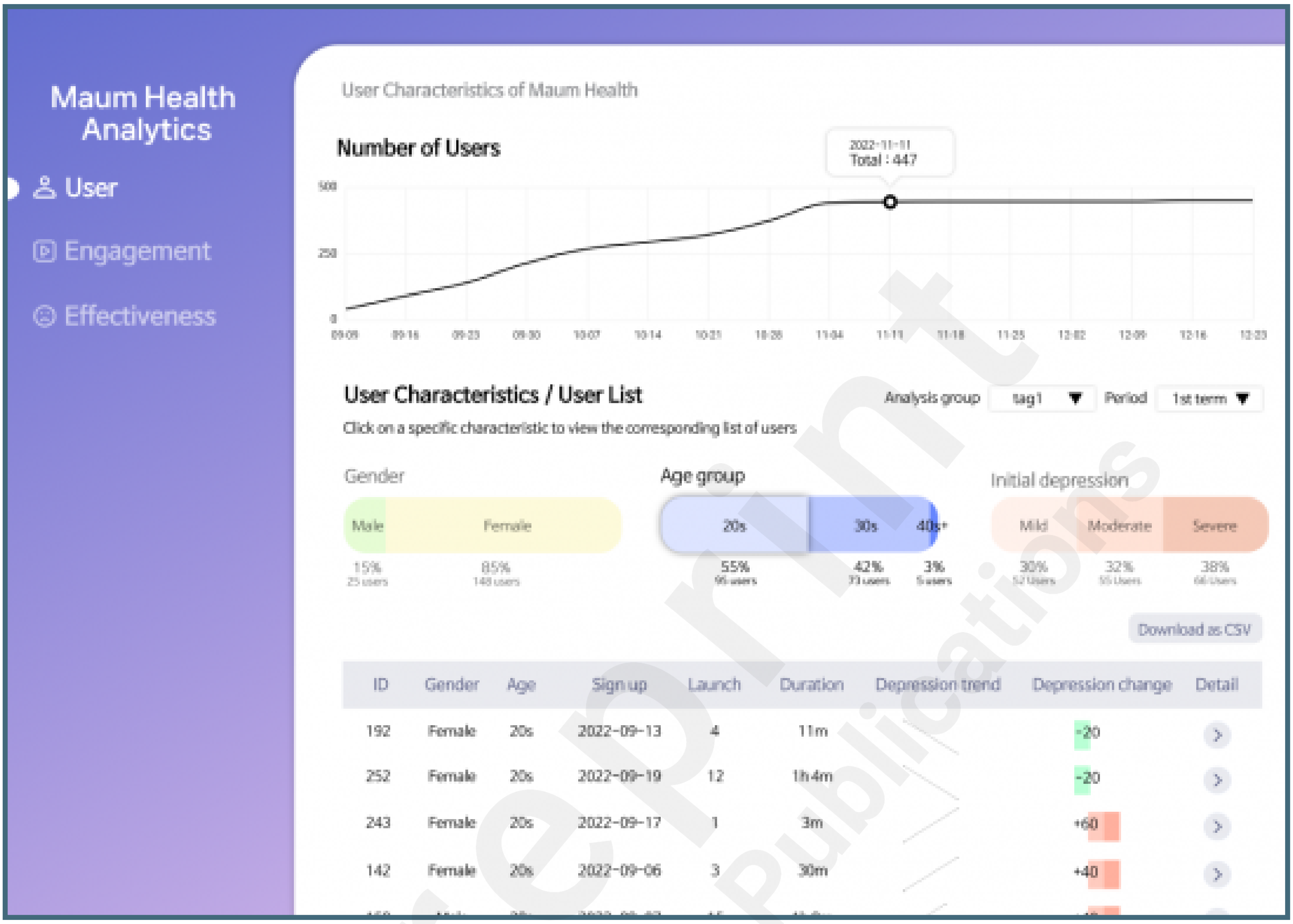
Supplementary Files

Figures

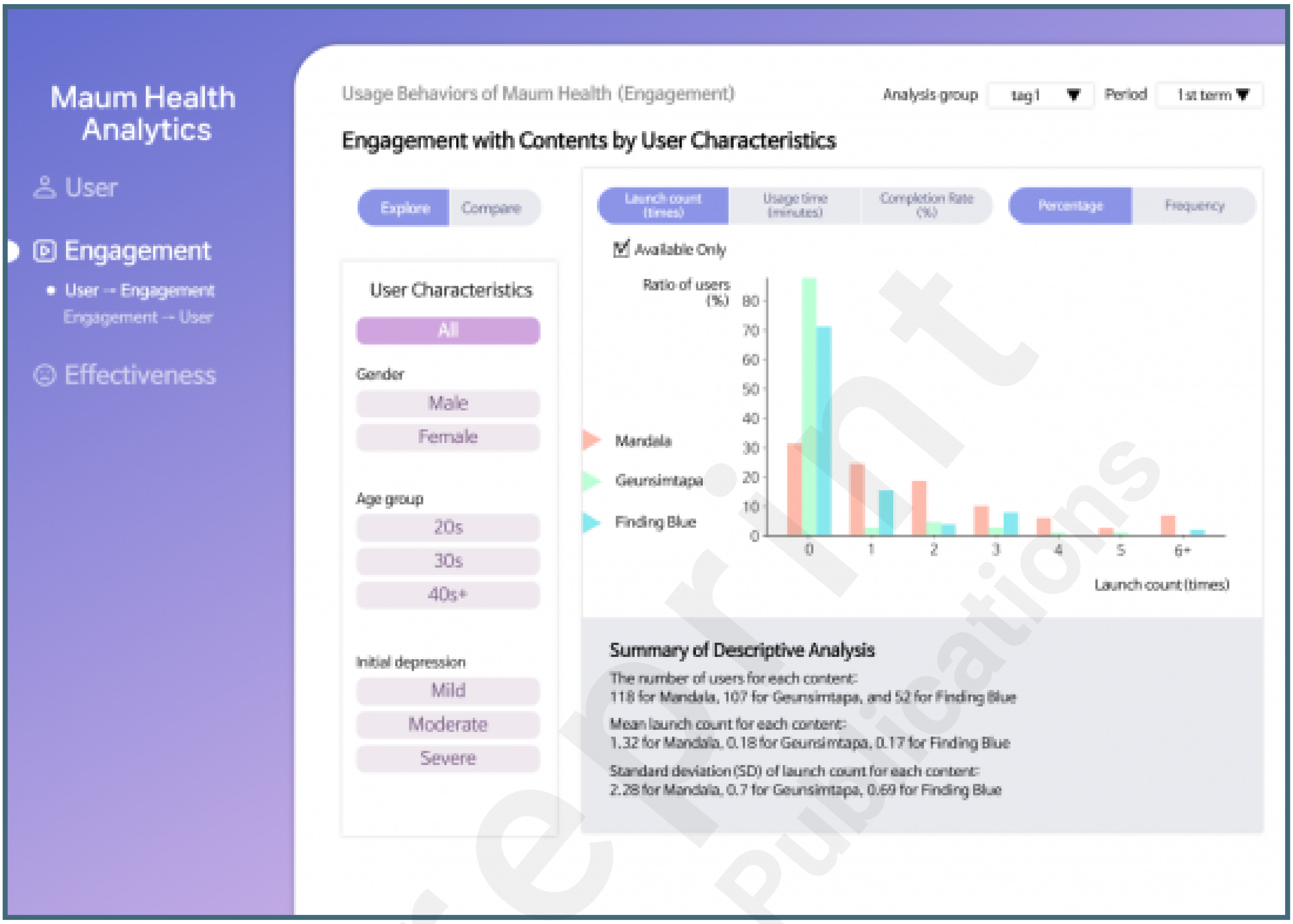
Overview of the proposed analysis task model for DHI data analysis.



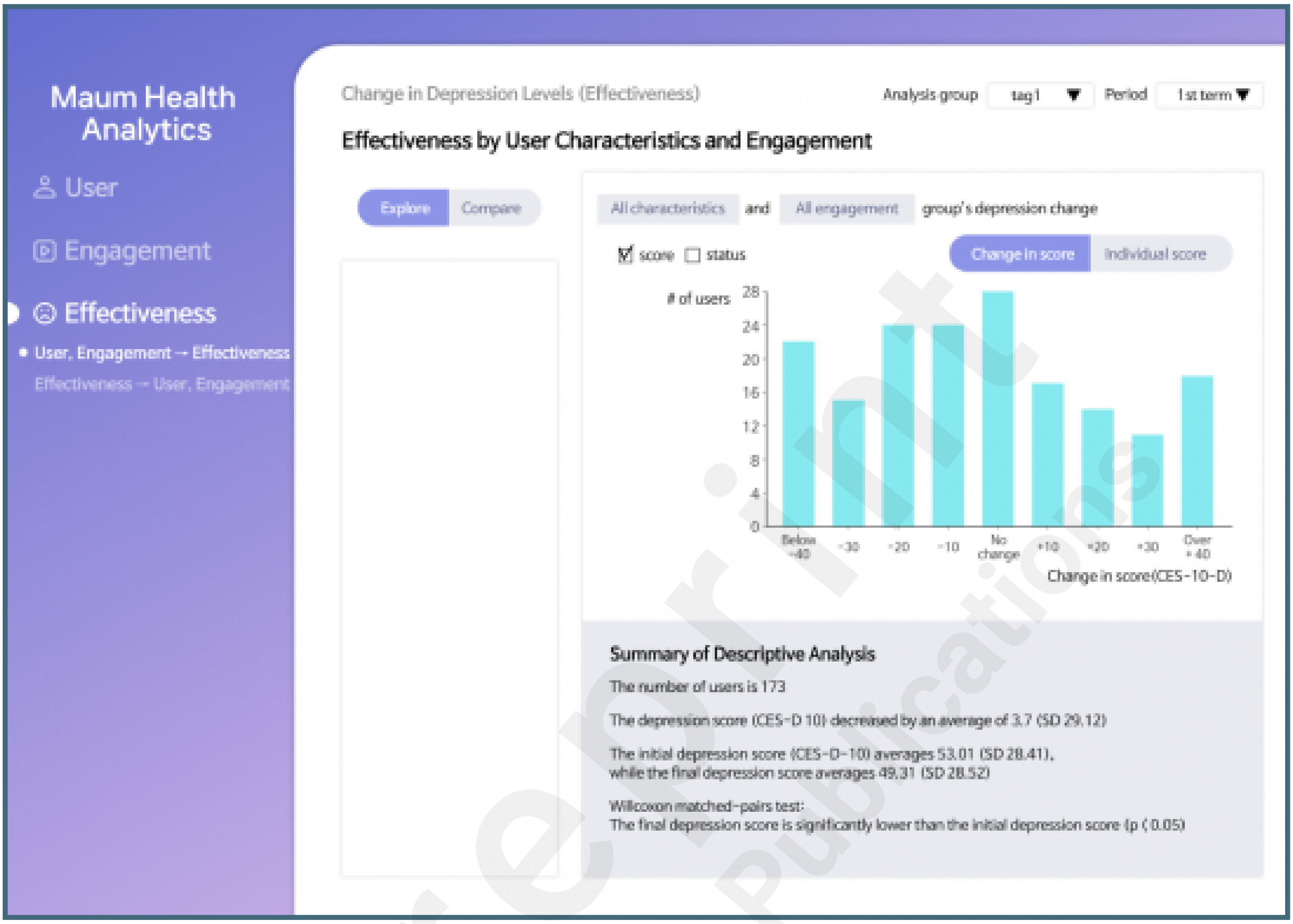
"User" page that supports tasks associated with user characteristics.



"Engagement" page that supports tasks associated with user engagement with DHIs.



"Effectiveness" page that supports tasks associated with effectiveness of DHIs on the target symptom.



Multimedia Appendixes

Description of analysis tasks supported by maum health analytics.
URL: http://asset.jmir.pub/assets/4bb4bad6fa6e2dec3fcd2dfa807fb42a.docx